NIWA
Taihoro Nukurangi

Marine invertebrate voucher specimen sequence typing

Final Report: MFE26167

Prepared for Ministry for the Environment

October 2023

Climate, Freshwater & Ocean Science

Prepared by:
Jaret Bilewitch
Amber Brooks
Sadie Mills

For any information regarding this report please contact:

Jaret Bilewitch
Molecular Biologist
Environmental Isotopes & Molecular Biology
+64-4-386 0502
jaret.bilewitch@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
Private Bag 14901
Kilbirnie
Wellington 6241

Phone +64 4 386 0300

NIWA CLIENT REPORT No: 2023280WN
Report date: October 2023
NIWA Project: MFE23305

Revision	Description	Date
Version 1.0	Final version sent to client	29 September 2023

Quality Assurance Statement		
	Reviewed by:	Rachael Peart
	Formatting checked by:	Jess Moffat
	Approved for release by:	Judy Sutherland

Contents

Tables

Figures

Executive summary

Biodiversity based assessments of aquatic environmental health rely on sampling, sorting and identification of macroinvertebrate communities. In Aotearoa/New Zealand, recent monitoring efforts have been adding environmental DNA (eDNA) sampling to their survey designs, as a complementary or alternative tool with reduced processing requirements and less need for expert taxonomic knowledge. However, the efficacy of eDNA-based community analyses often depends on the use of libraries of reference DNA sequences to infer the identity of genetic signatures detected in eDNA samples. It is common practice to assemble these libraries from public repositories such as GenBank or the Barcode of Life Database (BoLD), which may lack some New Zealand fauna and their available data may be incorrectly identified. Gaps and errors in reference libraries result in uncertainty and imprecision in eDNA taxonomic assignments, reducing their reliability in monitoring efforts. As such, bolstering reference datasets with regionally relevant taxa is an important requirement for effective eDNA applications in environmental health.

The marine environment of New Zealand contains over 10,000 known invertebrate species, making it difficult to assess community diversity at finer taxonomic scales (species and genera) as many taxa have not been DNA-sequenced at genetic barcode loci that are typically employed in eDNA surveys. To overcome this limitation on a regional basis, we used a curated voucher collection of benthic macroinvertebrate specimens from subtidal sites in Otago Harbour. These produced reference DNA sequences at two genetic barcodes, contributing to development of a reference library for use in future eDNA surveys.

A total of 96 specimens representing 72 distinct taxa were selected for processing, with 68 of these specimens producing viable DNA sequence data from COI and/or 28S barcode loci. Sequence data were produced for 56 taxa from six invertebrate phyla, with most originating from specimens of molluscs, arthropods and annelids. Twenty-four taxa were DNA-sequenced for the first time, expanding the breadth of our regionally focused eDNA reference dataset. The 28S sequencing displayed a higher success rate than COI sequencing, with only 43 viable DNA sequences being produced by COI markers. Very few specimens required repeat 28S amplification while most specimens required repeat COI amplification before a positive result was achieved. The nuclear 28S locus shows promise as a complementary marker to the commonly used COI locus but using it as a routine eDNA marker would first require building extensive reference sequence libraries.

Based on our findings and the current state of eDNA research, we recommend expanding and developing regionally focused reference sequence sets for use in eDNA surveys, ideally using cost-effective genomic options that provide sequence data for many markers. This could include DNA-sequencing the remaining 30 unsuccessful invertebrate taxa from the Otago Region plus 32 taxa from the Wellington Region, using other specimens; 19 and 21 of these taxa lack sequence data of any kind, respectively. Other further considerations could include comparisons of sequence data from expert-identified vouchers to pre-existing sequences available in public repositories, and benchmarking the performance of our custom reference library to those that are created from noncurated public repositories, to gain insight into the ramifications of taxonomic gaps and uncertainty in the application of eDNA surveys in the marine environment.

1 Introduction

In New Zealand, over 10,000 invertebrate species are known from the marine environment (Gordon et al. 2010), making it difficult to assess coastal environmental health using broad-scale community analyses that include measurements of the diversity of invertebrate communities as an environmental health indicator (*e.g.*, the Macroinvertebrate Community Index; MfE 2020). The difficulties of fine-scale taxonomic identification of macroinvertebrates can be partly alleviated using regionally focused collections that have benefitted from rigorous identification of representative taxa. Recent marine surveys in Te Upoke o te Ika/The Greater Wellington Region and the Ōtākou /Otago Region have developed reference collections of benthic macroinvertebrates encountered subtidally, using expert taxonomic identifications to establish a set of morphological standards against which future surveys may be compared, to assist non-specialist identifications (Cummings et al. 2020a, b; Forrest et al. 2022a, b).

Traditional approaches to invertebrate community surveys using bulk sampling, sorting and identification of physical samples (*e.g.,* Stark et al. 2001) are costly when applied in the marine environment. They require the use of snorkelling or SCUBA for sample collection plus lengthy processes for sorting and identifying many specimens that are distributed across a diverse range of invertebrate phyla. Environmental monitoring efforts in New Zealand have increasingly been exploring and implementing approaches that include environmental DNA (eDNA) sampling coupled with metabarcoding approaches, in part to reduce the costs of traditional labour-intensive survey methods (Zaiko & Steiner 2020), but also to complement traditional methods with alternative data sources (Leduc et al. 2019; Azevedo et al. 2020; Gold et al. 2021). The implementation of targeted (single species) eDNA detection is also becoming more common in focused biosecurity surveillance programmes or species-specific conservation initiatives, but it is of lesser relevance to community health indices. Diversity surveys of eDNA from marine environments typically rely on filtration of seawater samples and/or sampling of benthic sediments, followed by bulk DNA extraction, then amplicon metabarcoding of one or more genetic markers to DNA-sequence a cross-section of community diversity (e.g., all metazoans: Nguyen et al. 2020).

In order to translate DNA sequences into meaningful community diversity data, the sequence data produced by metabarcoding must be cross-referenced to genetic databases. These reference databases must be taxonomically reliable and contain genetic information for the same taxa found in the eDNA sampling area in question. If not, false positives for other species (including non-natives) may result when taxonomic assignment to a 'nearest neighbour' occurs: a sequence cannot be matched to its true source (because it hasn't been sequenced before) and instead is matched to the next most related (but different) species (Claver et al. 2021). Alternatively, assignation of sequences to a higher-level taxonomy (family, genus) may occur when reference libraries are incomplete or of uncertain accuracy (Stat et al. 2017; Locatelli et al. 2020), which reduces the sensitivity and precision of eDNA approaches by reducing taxonomic resolution. The ideal solution to low taxonomic coverage in reference libraries is to supplement them with reliable sequence data from specimens that have a high level of confidence in their identification, such as material from taxonomic studies or regional voucher sets that have benefitted from expert attention (Dopheide et al. 2022).

Previously a reference genetic database for marine invertebrates from Te Whanganui a Tara/Wellington was initiated (Bilewitch et al. 2022) to advance species-level eDNA surveys in areas with similar benthic communities. The current study did the same for marine invertebrates from Otago Harbour. A curated molecular sequence database was compiled using expertly identified

voucher specimens from Otago, to assist with future validation and implementation of marine eDNA surveys.

The objectives of this study were to:

1. Examine an ORC voucher collection and select up to 140 taxa for genetic analysis,
2. PCR-amplify and DNA-sequence selected voucher subsamples using two barcode markers,
3. Submit curated DNA sequence data to the Barcode of Life Database (BoLD),
4. Examine the efficacy of the resulting dataset for use in future eDNA surveys.

This work supports and provides context to the future use of eDNA sampling in the Otago Region and will form an important case study for the rest of New Zealand. It will also make a valuable contribution to broader endeavours facilitating the application of eDNA to monitor trends in ocean biodiversity, which are being undertaken by NIWA.

2 Methods

2.1 Specimen prioritisation, selection, and sampling

Specimens selected for DNA sequencing were chosen from among 290 voucher specimens from the Otago Region, which are archived in the NIWA Invertebrate Collection (see Forrest et al. 2022a, b for collection sites and sampling methods). These included 85 unique taxa confirmed by specialist taxonomists. From the 290 vouchers, 132 were prioritised for this study, representing the breadth of taxonomic diversity among the expert-identified specimen collection. Thirty-six of these specimens were excluded due to their small size and uniqueness (they were the only individual of that species for the sample site), thus vouchers would be completely consumed by the DNA extraction process and be unavailable as morphological references for future study. The remaining 96 specimens were subsampled for DNA extraction and sequencing, representing 72 unique taxa.

Each selected sample was examined under a dissecting microscope and an approximately 1-5mm^3 tissue was subsampled using bleach-sterilised dissecting tools. Tissue subsamples were placed on heat blocks at 56°C for 2 hours to evaporate excess ethanol and Rose Bengal stain prior to DNA extraction.

2.2 DNA extraction, amplification, and sequencing

DNA extractions from selected invertebrate samples used a DNeasy Blood & Tissue kit (Qiagen), following the manufacturer's recommended protocol except for the following modifications:

1. All samples were routinely digested in proteinase K overnight in a 56°C rotary incubator at 80 rpm.
2. All samples were eluted with two 40 µl volumes of Buffer AE, to maximize DNA concentrations.
3. Small specimens (<5 mm) used half volumes of all extraction reagents, except two 40 µl elutions were used.

Amplification of genetic loci used primers previously identified in the literature that would produce amplicon sizes suitable for eDNA workflows, based on length restrictions of DNA sequencing platforms typically used for metabarcoding (*e.g.,* Illumina MiSeq) and increased detection probabilities for shorter eDNA fragments that are subject to natural environmental degradation. Although the commonly used COI marker was an obvious choice for a primary barcode marker, there was no clear consensus in the recent literature on a second marker for metazoan eDNA applications. Based on our previous investigations of alternative markers (Bilewitch et al. 2022), we chose 28S rDNA as a secondary barcode capable of species-level discrimination and identification.

Loci were PCR-amplified from genomic DNA extracts using mlCOIintF (GGW ACW GGW TGA ACW GTW TAY CCY CC) and jgHCO2198 (TAN ACY TCN GGR TGN CCR AAR AAY CA) to amplify ca. 330-350bp of the COI 'barcode' locus (Leray et al. 2013), and 28SF_8 (GGG AAA GAA GAC CCT GTT GAG) and 28SR_11 (GCT TGG CBG CCA CAA GCC AGT TA) to amplify ca. 400-500bp of the nuclear 28S rDNA locus (Machida & Knowlton 2012). PCR reactions were conducted in 25µl total reaction volumes containing 1X MyTaq Mix (Bioline), 600nM of each primer, 12.5µg BSA, and 2-8µl of DNA extract. Thermal profiles used an initial denaturation of 95°C for 3 min, followed by 40 cycles of 95°C for 15 sec, annealing for 25°C sec, and 72°C for 20 sec, with a final extension of 72°C for 2 min. Reactions used 49°C and 58°C as annealing temperatures for COI and 28S amplifications, respectively. PCR

products were visualised on 1.5% agarose gels and were treated with ExoSAP-IT (Applied Biosystems) prior to submission for bi-directional Sanger DNA sequencing at a commercial facility (Macrogen Inc.).

Resulting DNA sequences were trimmed and assembled using Geneious Prime v2020.1.1 and chromatograms were visually inspected for quality. The fidelity of sequence data was checked using BLASTn searches (blast.ncbi.nlm.nih.gov/blast.cgi), resulting hits were examined for evidence of contamination (best matches to human, bacterial, protistan or fungal taxa), and sequences matching non-target phyla were excluded. For each of the three marker datasets, alignments of each of the Mollusca, Arthropoda and Annelida (the most abundant taxa) were used to calculate average pairwise identity values, for comparisons of marker variability.

3 Results and Discussion

3.1 Success rates and taxonomic coverage

In total, 96 specimens were extracted, representing 72 unique macroinvertebrate taxa (Table 3-1). All 96 successfully amplified for the 28S marker but 35 of these failed to produce high quality sequences, or amplified contaminant DNA, resulting in 50 out of 72 unique taxa sequenced at this locus. For the COI marker, 78 of the 96 specimens successfully amplified but 30 produced low-quality sequences that were discarded. A further five produced contaminant sequences (one human, one bacterium, one arachnid and two algae) and were discarded. Among the 43 specimens that successfully produced COI sequences, 38 unique taxa were represented. Overall quality sequences were produced for 56 unique taxa using either 28S or COI regions. As with our previous efforts (Bilewitch et al. 2022), the small size of many invertebrate vouchers plus a need to maintain morphological integrity of vouchers for future use presented challenges for extracting sufficient DNA for amplification and sequencing (Figure 3-1). We also suspect that Rose Bengal staining previously added to voucher specimens to assist with morphological identification was inhibiting DNA extraction or amplification, as seen in previous studies (Fonseca & Fehlauer-Ale 2012; Watanabe et al. 2016).

Figure 3-1: Example of ORC invertebrate voucher specimen used for DNA sequencing. This specimen of *Lasaea parengaensis* was successfully sequenced at both barcode markers, despite it being less than 2mm in diameter. Photo: Michal Ferries, NIWA.

Nine samples produced contaminant sequences for 28S and were discarded, resulting in 61 specimens with successfully sequenced 28S markers. Contamination of 28S PCRs were attributed to amplification of fungal DNA, which produced low quality sequences with high numbers of ambiguous positions. Such contamination risks, although easily identifiable, are difficult to prevent since universal metazoan primers used for broad scale eDNA diversity assessments do not necessarily exclude non-metazoan eukaryotes (fungi, protozoans) (Sepulveda et al. 2020). The likelihood of amplification of these non-specific background contaminants is seemingly increased when the availability of target organism DNA is low or non-existent, allowing off-target amplicons with lower primer annealing and amplification efficiency to proliferate during PCR cycling.

Of 96 samples extracted for genomic DNA, 68 (71%) produced viable, non-contaminant sequence data at one of the two loci (Appendix A), which were submitted to BoLD (JBMFB001-072). This included 56 unique taxa or 78% of the taxa available (Table 3-1).

Table 3-1: Summary totals of samples and unique taxa that were extracted and successfully sequenced for each locus. Sequenced = successful at either (or both) loci.

Region	Extracted	Sequenced	COI	28S
Samples	96	68	43	61
Taxa	72	56	38	50

The taxonomic distribution of sequencing results is shown in Table 3-2 and taxa that produced no viable sequence data from either of the attempted markers (n=30) are listed in Table 3-3. The proportion of unsuccessful taxa was evenly spread over the different taxa for COI but not for 28S. For the COI marker, approximately half of the Annelid, Arthropod and Mollusc taxa were unsuccessful. 28S had a higher success rate across all three main phyla. COI produced one more cnidarian sequence than 28S. Both markers produced one nemertean sequence but no nematode sequences (Table 3-2).

Table 3-2: Sequencing results for included invertebrate phyla. For each phylum, the number of samples available for DNA extraction are given, along with the subset of these that produced DNA sequence data for each of the two targeted barcode loci.

Phyla	Extracted	COI	28S
Annelida	46	18	25
Arthropoda	22	11	17
Cnidaria	4	2	1
Mollusca	21	11	17
Nemertea	2	1	1
Nematoda	1	0	0

Table 3-3: List of Otago vouchered taxa lacking sequence data. Taxa are listed for which no sequence data was produced at either of the two targeted loci. Taxon names represent expert identifications. *Previous work in the Wellington Region (Bilewitch et al., (2022) produced sequence data for this species. **Previously reported as not sequenced by Bilewitch et al. (2022).

Annelida:	
Armandia maculata	Orbiniidae
Boccardia sp.	*Owenia petersenae*
*Boccardia syrtis***	Paraonidae
Disconatis accolus	*Pettiboneia* sp.
Exogone sp.	*Platynereis* sp.
Goniada sp.	*Scoloplos cylindrifer*
Maldanidae	Sphaerodoridae
Microspio maori	*Spio readi*
Nephtyidae	Syllidae
Nereididae	
Arthropoda:	
Colurostylis lemurum	*Paracalliope* sp.
Eusiridae	*Paracorophium* sp.
Isocladus sp.	*Proharpinia* sp.
Lysianassidae	*Protorchestia* sp.
Mysida sp.	
Mollusca	
*Macomona liliana**	
Nematoda:	
Unidentified nematode	

Fifty-six taxa among 68 specimens that produced sequence data at one or more loci in this study were compared to pre-existing sequence data in GenBank (at any locus – not necessarily just those used here), to produce a list of 24 macroinvertebrate taxa (21 excluding those sequenced in Bilewitch et al. 2022) that have herein been sequenced for the first time (Table 3-4). The production of new marker data for these taxa represents a significant advancement on the current state of genetic resources for subtidal areas in the Otago region. At the outset of this project, 43 of the 85 unique taxa provided had no prior sequence data available in GenBank (and by extension in the BoLD, which cross-links entries to NCBI-GenBank). This work reduced the number of unsequenced taxa represented in the Otago voucher collection from 43 to 19.

We have also generated complementary sequence data for 32 taxa that already have existing sequence data available. Although it was outside the scope of this study, a comparison to examine concordance of our vouchered sequence data to those already in GenBank or BoLD would provide insight into the prevalence of misidentifications – both within public repositories and among expert-identified collections. There are also an additional 11 taxa that possess pre-existing GenBank sequences, which were not sequenced here; these require closer inspection prior to inclusion in reference datasets, to determine if they originate from reliably identified specimens (*sensu* Locatelli et al. 2020).

Table 3-4: List of newly sequenced macroinvertebrate taxa. List of identified ranks that were sequenced in the current study, for which no previous sequence data is found in GenBank. *Sequence results were also produced for these taxa in Bilewitch et al. (2022), **Previously reported as not sequenced by Bilewitch et al. (2022).

Annelida:	
Aglaophamus macroura	*Naineris* sp.
Aonides trifida	*Nicon aestuariensis**
Capitella cf. *capitata*	*Orbinia papillosa*
Glycera sp.	*Prionospio aucklandica**
Leodamas sp.	*Protocirrineris nuchalis*
Macroclymenella stewartensis	*Scolecolepides benhami*
Microphthalmus riseri	
Arthropoda:	
Exosphaeroma planulum	*Paramoera chevreuxi*
*Halicarcinus whitei**	*Torridoharpinia hurleyi***
Josephosella awa	*Urothoe* sp.
Mollusca:	
Arthritica sp.	*Neoguraleus* sp.
Lasaea parengaensis	*Nucula nitidula*
Legrandina turneri	

3.2 Barcode marker-specific considerations

A comparison of sequencing results for the two barcode markers trialled here (Table 3-2) indicated varied per-phyla success rates for COI and 28S. Sequencing of 28S produced more successes than COI for annelids, arthropods and molluscs, whereas both markers produced a single sequence for Nemertea and both failed to produce data for a single nematode. COI produced one more Cnidaria sequence than 28S, which may be due to stochastic errors in PCR amplification of low-concentration DNA. Previous barcoding for invertebrates from Wellington and Porirua harbours showed a slightly different pattern, with 28S displaying a higher success rate for molluscs and comparable rates to COI for annelids, arthropods and four other less common phyla (Bilewitch et al. 2022). A comparison of a larger sample size of specimens that also included an examination of the consistency of replicate PCRs would be necessary to distinguish between marker-specific differences and random error effects, but future efforts should include both markers rather than rationalising the selection of one over the other.

Alignments of sequence data for each of the three most abundant phyla were used to compare the amount of variation observed between the barcode markers (Table 3-5). COI showed substantially higher average variation (lowest percent identity) for annelids and molluscs, and only slightly lower variation in arthropods. This same pattern was observed in the GWRC samples, with 28S showing higher average variation only in arthropods (Bilewitch et al. 2022). These inconsistencies in marker variability are likely due to differences in the range of taxa included for each phylum. However, both markers displayed sufficient variation to make them suitable for species-level discrimination (COI= 54-70%, 28S= 68-80%; Bilewitch et al. 2022 showed COI= 64-70%, 28S= 62-82%:). This indicates that while COI is well-suited to its role as the 'standard' barcode marker for eDNA applications, incorporating a complementary marker such as 28S into standard eDNA metabarcoding practice may yield more precise and accurate identifications.

Table 3-5: Comparison of variation of marker sequences. For each phylum, the average percent-identity (similarity) is given, along with sample size (in parentheses).

Phylum	COI	28S
Annelida	54 (18)	80 (25)
Arthropoda	70 (11)	68 (17)
Mollusca	69 (11)	79 (17)

For aquatic invertebrates, eDNA metabarcoding surveys have typically used a COI marker alone (Elbrecht & Leese 2017; Nguyen et al. 2020) or in combination with 18S rDNA (Leduc et al. 2019; Leite et al. 2021) to characterise freshwater and marine communities. Such a combination of mitochondrial and nuclear barcodes is preferable since each genome can track speciation events independently (Degnan 1993; Moore 1995) and thus they can differ in their ability to distinguish and identify species. However, 18S rDNA metabarcoding is generally regarded as too invariant to provide species-level classifications for most metazoans (Drummond et al. 2015)., and the efficacy of alternative nuclear ribosomal markers has not been explored for barcoding. For example, the 28S marker has not been applied to metabarcoding since early trials (Machida & Knowlton 2012), although it has been effective at discriminating species in some cryptic marine invertebrate taxa (Kessel et al. 2022) and shows promise for three common and diverse phyla based on the current study and Bilewitch et al. (2022). Examples where other markers are used for eDNA metabarcoding of metazoans are uncommon, but some studies have targeted subsets of the invertebrate community using mitochondrial 16S primers for crustaceans (Berry et al. 2017) and cephalopods (Peters et al. 2015). Thus, although it is depauperate compared to more common barcoding markers such as COI and 18S, the 28S rDNA gene is worthy of further development as a 'universal' species-level marker for metazoans – particularly for marine invertebrates.

The industry 'standard' for the development and supplementation of eDNA reference databases is currently at a crossroads. Next-generation DNA sequencing (NGS) services have become more affordable as they have become more widely available – mostly due to market competition and scale – while costs associated with traditional Sanger approaches focussing DNA sequencing on a single marker from a single individual have remained static. NGS remains more costly, but it carries significant advantages such as producing sequences for multiple markers per individual, multiplexing multiple individuals (and markers) in a single sequencing reaction, and the production of 2-3 orders of magnitude more sequencing data per individual sample. Furthermore, the per-individual cost of NGS is usually reduced with increasing scale whereas traditional Sanger sequencing costs increase proportionally with sample size. These advantages and scalability have led researchers to favour genomic approaches to reference sequence development (Taberlet et al. 2018), particularly NGS methods that use low-genomic sequencing coverage to produce complete mitochondrial and ribosomal sequence sets (Dodsworth 2015; Trevisan et al. 2019). This 'genome-skimming' approach has been effectively used to develop reference libraries for plants (Dodsworth 2015) and animals (Hoban et al. 2022), including marine metazoans (Therkildsen & Palumbi 2017). The wealth of resulting genomic data obviates the need to choose between particular markers and thus acts as a future-proofing approach for instances where eDNA metabarcoding may shift towards multiple, different markers for different taxa in the near future (*e.g.*, Alexander et al. 2020). Commercial NGS services for genome skimming currently cost approximately $230/sample – nearly 20x the cost of Sanger services for a single marker from a single individual. However, skimming produces an average of 5 gigabytes of data per sample, whereas a Sanger sequencing reaction typically produces up to 1 kilobyte of data (5 million-times less), making the former a cost-effective prospect for generating large amounts of reference data for taxa which are under-represented in eDNA reference libraries.

4 Conclusions and Recommendations

In summary, we have produced 104 new DNA sequences at two loci for common macroinvertebrates in the harbours of the Otago Region. Of these, 24 taxa had previously not been sequenced before, three of these had only ever been sequenced during the previous GWRC barcoding project (Bilewitch et al. 2022). Annelids, molluscs, and arthropods contained most of the sampled taxa and produced most of the DNA sequences. Annelids had a lower success rate than the other two primary phyla, likely because of the small size of most of these specimens. Both COI and 28S markers produced sequence data but the 28S produced more sequencing success for annelids, arthropods and molluscs than COI; other per-phylum results were similar between markers. However, while the COI marker has large amounts of reference data in public repositories, the 28S marker has limited data availability and its implementation in eDNA monitoring would require further development of reference sequences for a wide range of taxa (especially marine invertebrates). Such marker-specific limitations could be avoided in the development of reference libraries through the use of a genome skimming approach, which can produce reference sequence data for all mitochondrial and nuclear ribosomal markers simultaneously, along with other novel nuclear gene markers.

Based on these findings, we propose the following recommendations for future consideration and advancement:

- Thirty-two taxa sequenced here have pre-existing data in GenBank. Wherever these data cover the same markers, they should be compared to provide an indication of the scale of identification errors in either the voucher collection used here or public sequence repositories such as GenBank and BoLD.
- Further development of reference sequence libraries should employ genome skimming to generate large scale, marker-independent datasets that buffer against future changes to common practice in eDNA biomonitoring. When applied at suitable scales, skimming represents a more cost-effective approach compared to traditional marker-specific approaches.
- Future sequencing efforts should focus on the remaining 30 taxa collected in the Otago Region that produced no data in this study and the remaining 31 GWRC taxa that produced no data in last year's study. 11 of the 30 ORC taxa not successfully sequenced have pre-existing data in GenBank, which can be used for comparison and confirmation of sequence fidelity.
- The COI universal barcode region works well for most tested taxa but ideally it should be coupled with a second nuclear marker (potentially 28S) to provide independent corroboration of species-level identifications.
- A database of the sequences produced here and in Bilewitch et al. (2022) should be tested as a taxonomic assignment tool for eDNA sequence data generated from the Otago and Greater Wellington Regions, in comparison to methods using BoLD or Genbank data alone.
- Expansion of effort to include other regional collections of expertly identified invertebrate specimens in New Zealand. The NIWA Invertebrate Collections contain several such collections that could be prioritised for this purpose.

5 Acknowledgements

The authors thank Sarah Hailes, Amelia Connell and Kate Neill (NIWA) for providing information and assembling the ORC voucher collection upon which this study was based, and Sam Thomas (Otago Regional Council) for providing permission to use the collection. We also thank Lisa Smith (NIWA) for her assistance and advice on laboratory workflows. We acknowledge the Ministry for the Environment for funding this project and thank Carolyn Mander & Michael Bates (MfE) for their support and involvement.

6 References

Alexander, J.B., Bunce, M., White, N., Wilkinson, S.P., Adam, A.A.S., Berry, T., Stat, M., Thomas, L., Newman, S.J., Dugal, L., Richards, Z.T. (2020) Development of a multi-assay approach for monitoring coral diversity using eDNA metabarcoding. *Coral Reefs,* 39(1): 159-171. 10.1007/s00338-019-01875-9

Azevedo, J., Antunes, J.T., Machado, A.M., Vasconcelos, V., Leao, P.N., Froufe, E. (2020) Monitoring of biofouling communities in a Portuguese port using a combined morphological and metabarcoding approach. *Scientific Reports,* 10(1): 13461. 10.1038/s41598-020-70307-4

Berry, T.E., Osterrieder, S.K., Murray, D.C., Coghlan, M.L., Richardson, A.J., Grealy, A.K., Stat, M., Bejder, L., Bunce, M. (2017) DNA metabarcoding for diet analysis and biodiversity: A case study using the endangered Australian sea lion (*Neophoca cinerea*). *Ecology & Evolution,* 7(14): 5435-5453. 10.1002/ece3.3123

Bilewitch, J., Kessel, G., Leduc, D., Mills, S., Sutherland, J. (2022) Marine invertebrate voucher specimen sequence typing. Prepared for Ministry for the Environment, *NIWA Client Report* 2022313WN: 21p.

Claver, C., Canals, O., Rodriguez-Ezpeleta, N. (2021) Assessing accuracy and completeness of GenBank for eDNA metabarcoding: towards a reliable marine fish reference database. ARPHA Conference Abstracts, 4. 10.3897/aca.4.e64671

Cummings, V., Halliday, J., Olsen, G., Hale, R., Greenfield, B., Hailes, S., Hewitt, J. (2022a) Te Whanganui-a-Tara (Wellington Harbour) subtidal monitoring: Results from the 2020 survey. Prepared for Greater Wellington Regional Council, *NIWA Client Report* 2021309WN: 70p.

Cummings, V., Halliday, J., Olsen, G., Hale, R., Greenfield, B., Hailes, S., Hewitt, J. (2022b) Te Awarua-o-Porirua Harbour subtidal sediment quality monitoring: Results from the 2020 survey. Prepared for Greater Wellington Regional Council, *NIWA Client Report* 2021309WN: 69p.

Degnan, S.M. (1993) The perils of single gene trees - mitochondrial versus single-copy nuclear DNA variation in white-eyes (Aves: Zosteropidae). *Molecular Ecology,* 2: 219-225.

Dodsworth, S. (2015) Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science,* 20(9): 525-527. 10.1016/j.tplants.2015.06.012

Dopheide, A., Brav-Cubitt, T., Podolyan, A., Leschen, R., Ward, D., Buckley, T., Dhami, M.K. (2022) Fast-tracking bespoke DNA reference database generation from museum collections for biomonitoring and conservation. *Molecular Ecology Resources,* 00:1-12. 10.1111/1755-0998.13733

Drummond, A.J., Newcomb, R.D., Buckley, T.R., Xie, D., Dopheide, A., Potter, B.C., Heled, J., Ross, H.A., Tooman, L., Grosser, S., Park, D., Demetras, N.J., Stevens, M.I., Russell, J.C., Anderson, S.H., Carter, A., Nelson, N. (2015) Evaluating a multigene environmental DNA approach for biodiversity assessment. *Gigascience*, 4: 46. 10.1186/s13742-015-0086-1

Elbrecht, V., Leese, F. (2017) Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. *Frontiers in Environmental Science,* 5. 10.3389/fenvs.2017.00011

Fonseca, G., Fehlauer-Ale, K.H. (2012) Three in one: fixing marine nematodes for ecological, molecular, and morphological studies. *Limnology and Oceanography: Methods,* 10(7): 516-523. 10.4319/lom.2012.10.516

Forrest, B.M., Roberts, K.L., Stevens, L.M. (2022a) Fine Scale Intertidal Monitoring of Tautuku Estuary. Salt Ecology Report 092, prepared for Otago Regional Council, June 2022. 27p.

Forrest, B.M., Roberts, K.L., Stevens, L.M. (2022b) Fine Scale Intertidal Monitoring of Pleasant River (Te Hakapupu) Estuary. Salt Ecology Report 093, prepared for Otago Regional Council, June 2022. 29p.

Gold, Z., Sprague, J., Kushner, D.J., Zerecero Marin, E., Barber, P.H. (2021) eDNA metabarcoding as a biomonitoring tool for marine protected areas. *PLoS One,* 16(2): e0238557. 10.1371/journal.pone.0238557

Gordon, D.P., Beaumont, J., MacDiarmid, A., Robertson, D.A., Ahyong, S.T. (2010) Marine biodiversity of Aotearoa New Zealand. PloS One, 5(8). 10.1371/journal.pone.0010905

Kessel, G.M., Alderslade, P., Bilewitch, J.P., Schnabel, K.E., Norman, J., Tekaharoa Potts, R., Gardner, J.P.A. (2022) Dead man's fingers point to new taxa: Two new genera of New Zealand soft corals (Anthozoa, Octocorallia) and a revision of *Alcyonium aurantiacum* Quoy & Gaimard, 1833. *European Journal of Taxonomy,* 837: 1-85. 10.5852/ejt.2022.837.1923

Leduc, N., Lacoursière-Roussel, A., Howland, K.L., Archambault, P., Sevellec, M., Normandeau, E., Dispas, A., Winkler, G., McKindsey, C.W., Simard, N., Bernatchez, L. (2019) Comparing eDNA metabarcoding and species collection for documenting Arctic metazoan biodiversity. *Environmental DNA,* 1(4): 342-358. 10.1002/edn3.35

Leite, B.R., Vieira, P.E., Troncoso, J.S., Costa, F.O. (2021) Comparing species detection success between molecular markers in DNA metabarcoding of coastal macroinvertebrates. *Metabarcoding and Metagenomics,* 5. 10.3897/mbmg.5.70063

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T., Machida, R.J. (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology,* 10: 34. 10.1186/1742-9994-10-34

Locatelli, N.S., McIntyre, P.B., Therkildsen, N.O., Baetscher, D.S. (2020) GenBank's reliability is uncertain for biodiversity researchers seeking species-level assignment for eDNA. *Proceedings of the National Academy of Sciences USA,* 117(51): 32211-32212. 10.1073/pnas.2007421117

Machida, R.J., Knowlton, N. (2012) PCR primers for metazoan nuclear 18S and 28S ribosomal DNA sequences. *PloS One,* 7(9): e46180. 10.1371/journal.pone.0046180

Ministry for the Environment (2020) National Policy Statement for Freshwater Management 2020. Published by Minister for the Environment. 70p.

Moore, W.S. (1995) Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution,* 49: 718-726.

Nguyen, B.N., Shen, E.W., Seemann, J., Correa, A.M.S., O'Donnell, J.L., Altieri, A.H., Knowlton, N., Crandall, K.A., Egan, S.P., McMillan, W.O., Leray, M. (2020) Environmental DNA survey captures patterns of fish and invertebrate diversity across a tropical seascape. *Scientific Reports,* 10(1): 6729. 10.1038/s41598-020-63565-9

Peters, K.J., Ophelkeller, K., Bott, N.J., Deagle, B.E., Jarman, S.N., Goldsworthy, S.D. (2015) Fine-scale diet of the Australian sea lion (*Neophoca cinerea*) using DNA-based analysis of faeces. *Marine Ecology,* 36(3): 347-367. 10.1111/maec.12145

Sepulveda, A.J., Hutchins, P.R., Forstchen, M., McKeefry, M.N., Swigris, A.M. (2020) The Elephant in the Lab (and Field): Contamination in Aquatic Environmental DNA Studies. *Frontiers in Ecology and Evolution,* 8. 10.3389/fevo.2020.609973

Stark, J.D., Boothroyd, I.K.G., Harding, J.S., Maxted, J.R., Scarsbrook, M.R. (2001) Protocols for sampling macroinvertebrates in wadeable streams. New Zealand Macroinvertebrate Working Group Report No. 1. Prepared for the Ministry for the Environment: 57p.

Stat, M., Huggett, M.J., Bernasconi, R., DiBattista, J.D., Berry, T.E., Newman, S.J., Harvey, E.S., Bunce, M. (2017) Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports,* 7(1): 12240. 10.1038/s41598-017-12501-5

Taberlet, P., Bonin, A., Zinger, L., Coissac, E. (2018) Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press. 10.1093/oso/9780198767220.001.0001

Therkildsen, N.O., Palumbi, S.R. (2017) Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. *Molecular Ecology Resources,* 17(2): 194-208. 10.1111/1755-0998.12593

Trevisan, B., Alcantara, D.M.C., Machado, D.J., Marques, F.P.L., Lahr, D.J.G. (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. 10.7717/peerj.7543

Watanabe, H.K., Senokuchi, R., Shimanaga, M., Yamamoto, H. (2016) Comparison of the efficiency of three methods of DNA extraction for deep-sea benthic copepods. *JAMSTEC Report of Research and Development, 23*(0): 52-59. 10.5918/jamstecr.23.52

Zaiko, A., Steiner, K. (2020) Environmental DNA screening of *Sabella spallanzanii* and *Styela clava* from water samples in the Tutukaka Marina. Cawthron Institute Report 3468: 18p

Appendix A

Table A-1: DNA barcoding results for 114 voucher specimens/taxa. For both barcode loci (COI & 28S), the number of ambiguously determined DNA sequence positions followed by the total DNA sequence length are given. Green indicates sequencing results with ≤2 ambiguities = good sequence quality; yellow indicates 3-9 ambiguities = moderate sequence quality; red indicates ≥10 ambiguities = poor sequence quality.

ORC number	COI	28S	Phylum	Full Taxon name
ORC-1	29/154	0/430	Mollusca	*Arthritica* sp. 5
ORC-3	4/313	3/518	Arthropoda	*Josephosella awa*
ORC-4	2/692	0/684	Arthropoda	*Exosphaeroma planulum*
ORC-5	4/339	2/543	Arthropoda	*Paracorophium excavatum*
ORC-10		5/500	Arthropoda	Aoridae
ORC-12	4/342	0/420	Mollusca	*Amphibola crenata*
ORC-14		0/442	Mollusca	Dotidae
ORC-15	0/315		Annelida	*Perinereis vallata*
ORC-17	55/313	5/425	Annelida	Naididae
ORC-22		0/436	Annelida	*Nicon aestuariensis*
ORC-23	2/344	1/415	Mollusca	*Potamopyrgus estuarinus*
ORC-25	0/314	0/434	Mollusca	*Austrovenus stutchburyi*
ORC-26		0/432	Mollusca	*Paphies australis*
ORC-27	0/283	0/408	Cnidaria	*Edwardsia* sp.
ORC-28	0/313	0/419	Mollusca	*Legrandina turneri*
ORC-29		13/401	Arthropoda	*Paracalliope* sp.
ORC-30		1/414	Mollusca	*Nucula nitidula*
ORC-31		0/521	Arthropoda	Tanaidacea
ORC-32		104/362	Mollusca	*Lasaea parengaensis*
ORC-33		0/412	Annelida	*Paradoneis lyra*
ORC-36	0/316	0/528	Arthropoda	*Torridoharpinia hurleyi*
ORC-38		37/547	Arthropoda	*Proharpinia* sp.
ORC-45	6/351	0/503	Arthropoda	*Urothoe* sp. 1
ORC-47	2/345	0/522	Arthropoda	*Torridoharpinia hurleyi*
ORC-48			Annelida	Nereididae
ORC-49	1/343	0/443	Annelida	*Capitella* cf. *capitata*
ORC-50			Annelida	Naididae
ORC-51	0/314	1/413	Mollusca	*Zeacumantus subcarinatus*
ORC-54			Annelida	*Platynereis*
ORC-56		7/513	Arthropoda	*Paracalliope* sp.

ORC number	COI	28S	Phylum	Full Taxon name
ORC-58		1/500	Arthropoda	*Parawaldeckia kidderi*
ORC-59	57/240		Annelida	*Scoloplos cylindrifer*
ORC-62		4/520	Arthropoda	*Paramoera chevreuxi*
ORC-66	5/317	0/381	Annelida	*Macroclymenella stewartensis*
ORC-67	6/289	0/442	Mollusca	Dotidae
ORC-68	6/683	0/336	Annelida	*Protocirrineris nuchalis*
ORC-77		0/477	Arthropoda	*Parawaldeckia kidderi*
ORC-79			Arthropoda	*Colurostylis lemurum*
ORC-80	19/314		Annelida	?*Platynereis*
ORC-81	23/180		Annelida	*Pettiboneia* sp [unknown, new?]
ORC-83	6/308	1/379	Mollusca	*Cominella glandiformis*
ORC-92	6/308	0/412	Annelida	*Prionospio aucklandica*
ORC-94	7/314	0/443	Annelida	*Capitella* cf. *capitata*
ORC-96	13/313	0/437	Annelida	*Scolecolepides benhami*
ORC-97	10/390	0/490	Arthropoda	*Paracalliope novizealandiae*
ORC-98	0/313	0/409	Annelida	*Heteromastus filiformis*
ORC-100	0/313	0/332	Annelida	*Paradoneis lyra*
ORC-101	38/243	0/178	Annelida	*Boccardia proboscidea*
ORC-102		187/749	Annelida	*Boccardia* sp. unknown
ORC-103			Nemertea	Nemertea
ORC-104			Cnidaria	*Edwardsia* sp.
ORC-105			Nematoda	Nematode
ORC-107			Annelida	*Platynereis*
ORC-108	0/313		Annelida	*Microphthalmus riseri*
ORC-111			Arthropoda	*Paracorophium* sp.
ORC-113			Annelida	Sabellidae, indeterminable
ORC-116		4/473	Mollusca	*Lasaea parengaensis*
ORC-125		16/401	Cnidaria/Anthozoa	Cnidaria/Anthozoa
ORC-126			Annelida	*Scoloplos cylindrifer*
ORC-129	4/345	0/493	Arthropoda	*Hemiplax hirtipes*
ORC-133	4/395	0/441	Annelida	*Scolecolepides benhami*
ORC-150	2/321	1/512	Arthropoda	*Austrohelice crassa*
ORC-163	22/305		Mollusca	*Austrovenus stutchburyi*
ORC-164	0/313	0/411	Annelida	*Glycera* sp.
ORC-166		0/391	Annelida	*Aglaophamus macroura*

ORC number	COI	28S	Phylum	Full Taxon name
ORC-167	0/313	0/520	Arthropoda	Tanaidacea
ORC-170			Annelida	*Exogone* sp.
ORC-177			Annelida	*Microspio maori*
ORC-179		0/459	Annelida	Oligochaeta
ORC-180	2/318		Annelida	*Naineris*
ORC-192	0/313	0/411	Annelida	*Aricidea* sp.
ORC-194	5/225	0/410	Annelida	Sabellidae, indeterminable
ORC-204			Annelida	Syllidae
ORC-209		0/422	Mollusca	*Cominella glandiformis*
ORC-217	2/286	2/450	Annelida	?*Leodamas* sp.
ORC-218	0/310	0/281	Annelida	*Orbinia papillosa*
ORC-227	0/313	0/352	Annelida	?*Thelepus* sp.
ORC-228	0/282		Cnidaria	*Edwardsia* sp.
ORC-229	35/316		Mollusca	*Macomona liliana*
ORC-235	0/259	0/422	Annelida	*Barantolla lepte*
ORC-239	0/313	0/396	Annelida	?*Thelepus* sp.
ORC-246	0/313	1/224	Nemertea	Nemertea
ORC-248		0/413	Annelida	*Aonides trifida*
ORC-250	2/321	0/465	Arthropoda	*Halicarcinus whitei*
ORC-255	0/310		Annelida	*Naineris* sp. naineris-A
ORC-256	2/315	0/445	Mollusca	*Micrelenchus huttonii*
ORC-258	72/313		Annelida	*Capitella* cf. *capitata*
ORC-260	13/342	0/523	Mollusca	*Notoacmea scapha*
ORC-264	25/430	50/492	Annelida	*Armandia maculata*
ORC-267	0/314	0/420	Mollusca	*Cominella glandiformis*
ORC-277	0/313	27/404	Mollusca	*Diloma subrostrata*
ORC-280		2/409	Annelida	*Aonides trifida*
ORC-281	0/313	6/408	Mollusca	*Neoguraleus* sp.
ORC-285	3/345	67/387	Arthropoda	*Austrominius modestus*
ORC-286			Annelida	*Owenia petersenae*
ORC-288			Annelida	Sphaerodoridae