The University of Maine
DigitalCommons@UMaine

Electronic Theses and Dissertations | Fogler Library

Fall 12-2021

Use of the New England Aquarium to Evaluate Environmental DNA Metabarcoding of Gulf of Maine Vertebrates and Invertebrates

Samantha Silverbrand
University of Maine, samantha.silverbrand@maine.edu

Follow this and additional works at: https://digitalcommons.library.umaine.edu/etd

Part of the Marine Biology Commons

Recommended Citation
Silverbrand, Samantha, "Use of the New England Aquarium to Evaluate Environmental DNA Metabarcoding of Gulf of Maine Vertebrates and Invertebrates" (2021). *Electronic Theses and Dissertations*. 3538.
https://digitalcommons.library.umaine.edu/etd/3538

This Open-Access Thesis is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.

USE OF THE NEW ENGLAND AQUARIUM TO EVALUATE ENVIRONMENTAL DNA METABARCODING OF GULF OF MAINE VERTEBRATES AND INVERTEBRATES

By

Samantha Silverbrand

B.S. University of Maine, 2019

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

(in Marine Biology)

The Graduate School

The University of Maine

December 2021

Advisory Committee:

Michael Kinnison, Professor of Evolutionary Applications

Kristina Cammen, Assistant Professor of Marine Mammal Science

Erin Grey, Assistant Professor of Aquatic Genetics

USE OF THE NEW ENGLAND AQUARIUM TO EVALUATE ENVIRONMENTAL DNA METABARCODING OF GULF OF MAINE VERTEBRATES AND INVERTEBRATES

By Samantha Silverbrand

Thesis Advisor: Dr. Michael Kinnison

An Abstract of the Thesis Presented
in Partial Fulfillment of the Requirements for the
Degree of Master of Science
(in Marine Biology)
December 2021

Environmental DNA (eDNA) metabarcoding is a tool that has been used to characterize biodiversity in a range of diverse systems. However, blind application of eDNA metabarcoding primer sets to new regions and species pools can result in poor taxon coverage and unaccounted detection biases. For the Maine-eDNA EPSCoR program, one of the main focuses is to understand and characterize community assemblages in the Gulf of Maine (GoM) using eDNA to further inform conservation, monitoring, and sustainability. In this study, I selected a subset of the best performing vertebrate and invertebrate metabarcoding assys to test against GoM species present in the New England Aquarium, Boston MA, USA. Each metabarcoding primer set was applied to the same set of replicate water samples taken from each of multiple aquarium displays with distinct and censused GoM assemblages. Using these known positive communities of fish and invertebrates I assessed the relative taxonomic specificity and overlap of the different assays, whether sequence counts can be applied to estimate relative species dominance within a sampling region, and what level of sample replication is needed to reliably and repeatedly account for dominant taxa. This study found that combining multiple metabarcoding assays for vertebrates can resolve a majority of GoM vertebrates, with the 12S MiFish-U assay and the 16S MarVer3 assay working best in combination for this goal. Additionally, it was found that rank species sequence counts are often approximately indicative of relative biomass, suggesting that eDNA metabarcoding may reveal more about GoM communities than just species occupancy. Finally, while

there were always taxa missed by the vertebrate metabarcoding primer sets, rarefaction analysis suggested that as few as one or two samples were sufficient to detect most or all of the species that were ultimately detectable. For the invertebrate markers tested in this study, the18S set was unusable due to possible laboratory or sequencing errors. The COI assay used in this study provides promising results for broad invertebrate taxonomic coverage, even down to species level detections for GoM taxa. However, this wide taxonomic coverage came with a tradeoff of missing many known species within larger groups. Hence, while the COI invertebrate primer set might ultimately be a useful part of a metabarcoding toolset for resolving GoM invertebrates, it might often be best combined with other primer sets for GoM biodiversity questions requiring more comprehensive coverage of particular subgroups of invertebrates.

ACKNOWLEDGEMENTS

I'd like to thank my advisor, Dr. Michael Kinnison, for inviting me into his lab, encouraging me and providing the immense support needed to complete a graduate degree. My other committee members, Dr. Kristina Cammen and Dr. Erin Grey, have also been immensely supportive throughout my time working on this thesis, despite some last minute changes at truly un-ideal times on my part. Thank you so much for your support, guidance and flexibility.

All of the laboratory work and much of the thesis would not have been possible without the guidance of Geneva York throughout the entirety of this process. I owe much of my expertise and knowledge in molecular laboratory skills, eDNA knowledge and troubleshooting to her. Without her help, the metabarcoding performed here would not exist.

I'd also like to thank all of the graduate and undergraduate students that have helped with various lab related tasks, from filtering through making gels and helping with PCRs, and some of my coding issues throughout this journey. In particular, I'd like to thank Julia Sunnarborg, Rachel Presley, Vendy Hazukova, Joe Mohan, and Grayson Huston for help in lab processes, as well as Bradley Erdman, Zach Wood and Steven Allers for help in R, QIIME and in learning command line and python on the fly. Huge thank you to Melissa Kimble and all of my other friends and my family during graduate school for providing insight on visuals and data processing, and generally listening to me rant or be anxious over this. Thank you as well to Noah Burby, Daniella Leal-Espinal and Jessie Hutchinson for your help in making and visualizing many of my gels and being the best undergrads I've ever worked with.

Finally, this work would not be possible without the help of the many professionals and staff I've worked with throughout my time at UMaine. In particular, individuals on the Maine-eDNA EPSCoR grant have provided me the resources and support necessary to complete this thesis. Most pointedly, Dr. Laura Jackson who developed the bio-informatics pipeline used in this study as well as for all of her generous help and patience when I've been in a somewhat time-crunched learning process. Along with

her, Dr. Andrew Rominger has given extensive time and expertise in bioinformatics and data processing that have been incredibly useful. Additionally, Jami Downing, Lauren Crofton-MacDonald, Mike Hambrock, Trish Costello, and the staff of the New England Aquarium. You've all been so kind and helpful, and you do amazing work every day. Thank you so much for everything.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

LIST OF TABLES ix

LIST OF FIGURES x

1. THE USE OF ENVIRONMENTAL DNA METABARCODING IN BIODIVERSITY STUDIES 1

2. EVALUATING ALTERNATE AND COMBINED ENVIRONMENTAL DNA METABARCODING PRIMER SETS FOR GULF OF MAINE VERTEBRATES 9

2.1. Introduction 9

2.2. Methods 12

2.2.1 Literature review and assay selection 12

2.2.2. eDNA water sampling 13

2.2.3 eDNA preparation 14

2.2.4. PCR amplification and sequencing 15

2.2.5 Bioinformatic pipelines 16

2.2.6 Data Analysis 17

2.3. Results 19

2.3.1. In-silico analysis 19

2.3.2. Sequencing results 20

2.3.3. Taxonomic resolution 21

2.3.4. Effects of assay compounding 23

2.3.5 Biomass and eDNA relationships 25

2.3.6. Effects of sample replication 27

2.3.7 Off-target identification 29

2.4. Discussion 30

2.4.1. Assessing primer performance for Gulf of Maine vertebrates 31

2.4.2. Comparison to in-silico expectations 34

2.4.3. Effects of sample level replication in detection of vertebrate taxa 36

2.4.4. eDNA and relative abundance: can it be equated? 38

2.4.5. Concerning/ addressing contamination or detection errors 39

2.5. Conclusions 40

3. EVALUATING ALTERNATE AND COMBINED ENVIRONMENTAL DNA METABARCODING PRIMER SETS FOR GULF OF MAINE INVERTEBRATES 42

3.1. Introduction 42

3.2. Methods 43

3.2.1 Literature review and assay selection 46

3.2.2. eDNA water sampling 46

3.2.3 eDNA preparation 47

3.2.4. PCR amplification and sequencing 48

3.2.5 Bioinformatic pipelines 49

3.2.6 Data Analysis 50

2.3. Results 51

3.3.1. Sequencing results 51

3.3.2. In-silico analysis 52

3.3.3. Taxonomic resolution 53

3.3.4. Total population versus eDNA sequencing counts 54

3.3.5. Effects of sample replication 55

3.3.6. Off-target identification 56

2.4. Discussion 59

3.4.1. In-silico evaluation of markers 59

3.4.2. COI taxonomic detection 61

3.4.3. Effects of sample level replication in detection on invertebrate detection 63

3.4.4. Relating abundance and eDNA sequencing counts 64

3.4.5. Off-target amplification 65

3.5. Conclusions 66

REFERENCES 68

APPENDIX 75

BIOGRAPHY OF THE AUTHOR 118

LIST OF TABLES

Table 2.1. Sequence data summary table for vertebrate assays 20

Table 2.2. Spearman's rank correlation coefficients for vertebrate assays 25

Table 3.1. Sequence data summary table for invertebrate assays 52

Table 3.2. Spearman's rank correlation coefficients for invertebrate assay 55

Table A.1. Aquarium census data 75

Table A.2. Primers, thermocycling conditions and size selection protocols 82

Table B.1. Taxonomic assignment table for MiFish primer 90

Table B.2. Taxonomic assignment table for Riaz primer 92

Table B.3. Taxonomic assignment table for MarVer primer 94

Table B.4. Blank sequence reads for MiFish primer 96

Table B.5. Blank sequence reads for Riaz primer 97

Table B.6. Removed taxa list for vertebrate primers 99

Table B.7. In-silico expectations for vertebrate primers 101

Table C.1. Taxonomic assignment table for COI assay 104

Table C.2. Taxonomic assignment table for 18S assay 106

Table C.3. Blank sequence reads for 18S assay 110

Table C.4. Removed taxa list for COI assay 111

Table C.5. Detection table and census information for invertebrates in this study 112

LIST OF FIGURES

Figure 1.1 eDNA publications through the last decade..6

Figure 2.1 Venn diagram of species detections across vertebrate assays 22

Figure 2.2 Venn diagram of genera detections across vertebrate assays 23

Figure 2.3 Stacked bar of species resolution with addition of assays 24

Figure 2.4 Stacked bar of genera level resolution with addition of assays 25

Figure 2.5 Rank correlations by tank for vertebrate assays...26

Figure 2.6 Rarefaction analyses by tank for vertebrate assays..28

Figure 3.1 Stacked bar of taxonomic resolution for COI assay 54

Figure 3.2 Rank correlations by tank for invertebrate assay...55

Figure 3.3 Rarefaction analyses by tank for COI assay 56

Figure 3.4 Pie charts showing off-target proportion of OTUs for COI assay 58

Figure B.1 Pearson correlations for MiFish assay 84

Figure B.2 Pearson correlations for MarVer assay 85

Figure B.3 Rarefaction on sequencing reads for vertebrate assays..86

Figure B.4 Rarefaction on species versus sample replicate for vertebrate assays 88

Figure C.1 Rarefaction on sequencing reads for COI assay 102

Figure C.2 Rarefaction on species versus sample replicate for COI assay 103

CHAPTER 1

THE USE OF ENVIRONMENTAL DNA METABARCODING IN BIODIVERSITY STUDIES

Environmental DNA (eDNA) is defined as both a material object and as a methodological tool. As a material object, eDNA is DNA that is present and can be collected in the environment, such as in water, soil, or air. This DNA originates as genetic material naturally shed by organisms into their environment. In the case of many animals, this genetic material most commonly derives from epithelial cells lining their skin, digestive tract, excretory system, or respiratory system, but can also derive from other tissues in association with predation, death, or reproduction (Thomsen, Kielgast, Iversen, Møller, et al., 2012; Thomsen & Willerslev, 2015). As a methodological tool, eDNA has been defined as the science of assaying the above defined genetic material in environmental samples with molecular genetic approaches for the purpose of answering questions in ecology, conservation biology, paleontology and more (Thomsen & Willerslev, 2015). This tool has a wide range of applications, including directed monitoring of invasive (Gentile Francesco Ficetola et al., 2008), threatened (Jerde et al., 2011; Thomsen, Kielgast, Iversen, Wiuf, et al., 2012), or economically important species (Salter et al., 2019), as well as characterization of diverse species assemblages (Andruszkiewicz et al., 2017; Closek et al., 2019; Gold et al., 2021).

The utilization of eDNA approaches in ecological and environmental fields holds great potential (McElroy et al., 2020), but realizing this potential depends on identifying and optimizing eDNA tools that are best suited to the particular study systems and questions of interest. Detection and quantification of eDNA is dependent on multiple factors that shape the inherent production, transport and loss dynamics of eDNA in nature (Deiner, Bik, et al., 2017), the power of particular sampling methods and survey designs to capture this eDNA (Dejean et al., 2011), the efficiency and sensitivity of molecular detection in the lab (Wilcox et al., 2013), and the bioinformatic resolution of sequence data (Stoeckle et al., 2020). The first two of these can be considered "field factors" and latter two as "laboratory factors". Both the field- and

lab-side of eDNA approaches are equally important, because as a process stream, biases introduced at any step from sampling to data analyses can substantially influence inferences. This thesis is primarily concerned with evaluating and optimizing lab-side factors associated with a particular eDNA approach referred to as “eDNA metabarcoding”. However, it may be useful for many readers to understand how field- and lab-side factors interact to affect eDNA inferences in general. As such, I provide a brief account of some of the major field-side factors affecting eDNA inferences, before delving into the molecular and bioinformatic factors most relevant to my thesis research.

Field Factors Affecting eDNA Detection and Quantification

Broadly speaking, the probability of collecting DNA is highest in the area where a species has been, while as you move further away from such locations detection probability declines as a result of transport, dilution, and loss of genetic material (Wood et al., 2020). The loss of eDNA from environmental media like water is associated with settlement and sequestering of eDNA to sediments and biofilms, along with breakdown of eDNA into fragments too small to be informative. The breakdown rate of extracellular DNA is dependent on multiple processes in an environment, including microbial action (Harrison et al., 2019) and physical/chemical deterioration (e.g., UV light) (Dejean et al., 2011; Thomsen, Kielgast, Iversen, Møller, et al., 2012). These processes affecting loss are countered by other processes that can return eDNA to suspension, including turbulence, and seasonal mixing (Deiner, Renshaw, et al., 2017; Thomsen & Willerslev, 2015). Processes governing production, transport, and loss of eDNA can differ substantially among environments. Specifically in marine environments, reduction in eDNA detection has been suggested to be significantly faster than that of freshwater or sediment specific environments (Dell’Anno & Corinaldesi, 2004), indicating low preservation and dispersion of eDNA in marine environments. However, eDNA has also been proven to be transported and preserved for long periods of time, from weeks in lentic systems (Gentile Francesco Ficetola et al., 2008; Schmelzle & Kinziger, 2016) to thousands of years in aquatic sediments (Dejean et al., 2011; Thomsen & Willerslev, 2015).

Although rapid loss and degradation of the DNA is a limitation of the use of eDNA in some monitoring contexts, it also has its benefits. Rapid loss and degradation of eDNA fragments improves prospects for drawing contemporaneous inferences. Many field monitoring contexts seek to gather a snapshot of the diversity at a specific sampling location and specific period in time (Morey et al., 2020; Thomsen & Willerslev, 2015). For example, detecting a particular organism or group of organisms at particular spatial and temporal resolutions can lend insights into biological invasion monitoring (Dejean et al., 2012; Gentile Francesco Ficetola et al., 2008; Geerts et al., 2018), characterization of critical habitat for threatened species (Bush et al., 2020; Thomsen, Kielgast, Iversen, Wiuf, et al., 2012), assessment of changing species ranges with climate change (Stoeckle et al., 2020), as well as identification of other specific local factors driving changes in community composition.

Laboratory Factors Affecting eDNA Detection and Quantification

Although the factors shaping the production, transport, and loss of eDNA in nature, as well as the biases and power of particular survey designs, clearly affect eDNA inferences, factors affecting the molecular detection and quantification of eDNA in the lab are just as important. A carefully conducted field survey accounting for eDNA production, loss and transport can be of very limited value if laboratory and bioinformatic resources introduce significant and unaccounted biases in detection and quantification. To understand the lab-side biases of eDNA it is useful to distinguish the two major molecular approaches to eDNA – species-specific amplification, and metagenomic sequencing.

Species-specific amplification methods using traditional PCR, quantitative PCR (qPCR), or digital droplet PCR, make use of primer sets, and sometimes fluorescent probes, that are typically designed to anneal and amplify DNA of a single species of interest, while not amplifying the DNA of off-target species. All of these common methods apply some form of PCR amplification to make the very low copy numbers of target eDNA typical of macro-organisms detectable or quantifiable. Metagenomic methods make use of next generation sequencing and bioinformatics approaches to characterize a typically much

broader pool of taxa contributing their eDNA to an environment. The predominant method of multiple species detection with eDNA is often referred to as eDNA metabarcoding, and involves simultaneous identification of multiple taxa from environmental or biological samples with the use of universal PCR primers (Riaz et al., 2011; Taberlet et al., 2012; Zhang et al., 2020). eDNA metabarcoding is more common for studies of macro-organism eDNA than approaches like shotgun sequencing, because it includes a PCR amplification step to increase the amount of eDNA of such organisms to high enough levels to not be swamped by the vastly more prevalent microbial DNA in most environments.

Because most eDNA methods are PCR dependent, they are subject to a common set of processes that can be broadly categorized as PCR bias (Alberdi et al., 2018; Elbrecht & Leese, 2015; Nichols et al., 2018). PCR bias derives from the fact that PCR is a recursive amplification process, and because of this, any inefficiencies or copying errors that are introduced in the process tend to compound cycle-to-cycle. Even small difference in amplification efficiency can result in orders of magnitude different amounts of DNA after dozens of amplification cycles. In some cases, these subtle, and sometimes not so subtle, PCR biases can be the difference in whether eDNA is ultimately detected or sequenced. In the case of species-specific amplification approaches, like qPCR, investigators will often dedicate significant time and resources to evaluating and optimizing assays to minimize such PCR bias for the target species of interest (Bruce et al., n.d.). However, it is currently not possible to simultaneously optimize metabarcoding methods for a large number of species. eDNA metabarcoding employs less-specific “universal” primers that are designed to amplify the DNA of many species that often differ in their sequence match to those primers, the size of the fragment between those primers, and the base composition of that intervening sequence, all of which can result in differences in PCR efficiencies. As such, some degree of amplification and sequencing bias is largely accepted as a tradeoff for the larger taxonomic coverage of eDNA metagenomic approaches (Gold et al., 2021; Nichols et al., 2018).

Next generation sequencing is necessarily coupled to bioinformatics approaches in order to interpret the sequence data that is generated. Although there are some taxon-free approaches based on sequence

diversity alone (Deagle et al., 2014), most eDNA work is more applied and thus seeks to link the pool of captured sequences to their originating species. Identifying sequences to their source species is dependent on the sequence data available in a diversity of reference databases. Some of these references databases are much more complete than others, in large part due to the amount of prior genetic research and diversity of a given taxonomic group. For example, reference sequence data are much more complete for the mitochondrial gene cytochrome oxidase I than many others, due to its wide use as a "barcode of life" (Elbrecht & Leese, 2017). Likewise, some reference resources have been more carefully curated than others to reduce problems of species IDs. And finally, the particular genomic regions targeted by some metagenomic methods, like eDNA metabarcoding, simply include more informative sequence variation to provide greater taxonomic resolution for some taxa than others (Collins et al., 2019; Deagle et al., 2014; Jackman et al., 2021). The ultimate result of all these factors affecting the quality of reference resources for bioinformatics, is that some species that contribute eDNA to field samples may only be resolved to higher taxonomic levels like genera or families, or even confused for other species. Again, this degree of taxonomic imprecision is largely accepted as a compromise with the ability to simultaneously detect a wide range of taxa. Because they do not directly involve sequencing, most species-specific eDNA tools do not encounter these bioinformatic issues beyond the initial stage of having adequate sequence data to design primers etc.

eDNA Metabarcoding and Maine-eDNA

eDNA Metabarcoding has grown in popularity as a methodology in eDNA research. Some of the first studies to apply metabarcoding to eDNA of macro-organisms included those studying plant and animal records from the Pleistocene (Willerslev et al., 2003) and to detecting invasive frogs in France (Gentile Francesco Ficetola et al., 2008). Since that time, eDNA metabarcoding studies have increased exponentially as shown by Figure 1.1. As mentioned above, eDNA metabarcoding requires use of

universal PCR primers (Riaz et al., 2011; Taberlet et al., 2012; Zhang et al., 2020). These primers typically target hypervariable gene regions and choice of the gene region typically revolves around copy number of DNA and taxonomic resolution for the taxonomic group of interest (Riaz et al., 2011; Thomsen & Willerslev, 2015). Most cells of organisms contain many replicate mitochondrial genomes, and thus mitochondrial gene regions are most often used to design metabarcoding primers, as on the 12S, 16S and COI gene regions for vertebrates for example (Zhang et al., 2020). The resulting primers usually amplify short regions, up to a few hundred base pairs, in order to accommodate the fact that eDNA material is often very degraded in nature (Deagle et al., 2009; Leray et al., 2013; Valentini et al., 2016).

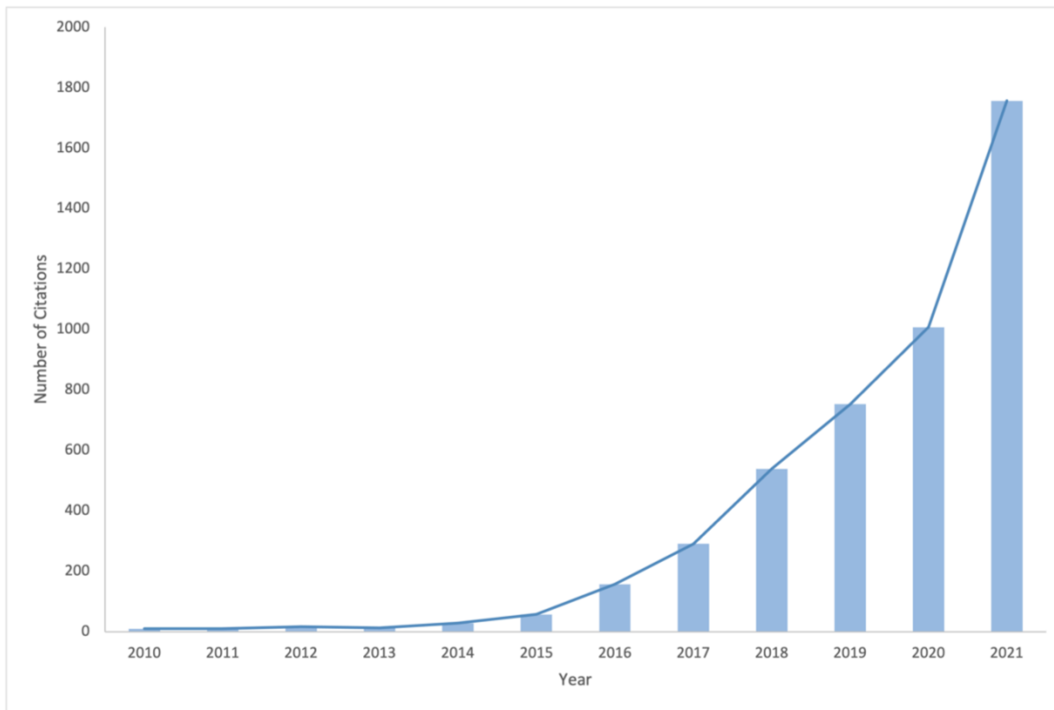

Figure 1.1 Publications from 2010 through 2021 that included the search terms "eDNA metabarcoding" and "environmental DNA metabarcoding" from GoogleScholar.

Because early studies of eDNA metabarcoding often emphasized development of a single optimal primer set for a regional species pool (Geerts et al., 2018; Gielings et al., 2021), there has been a proliferation of alternate eDNA metabarcoding primer sets and reference databases for some taxa. For

example, Zhang et al. 2020 recently reviewed 22 metabarcoding primers sets for teleost fishes. Unfortunately, use of such study-specific metabarcoding primer sets greatly limits the ability for data sharing and comparison across studies and regions. It has since been recognized that some metabarcoding primer sets perform very well across regions (García-Machado et al., 2021; Gold et al., 2021; Jackman et al., 2021) and that even greater regional coverage might be achieved via combining (multiplexing) of such high-performing primer sets, than by designing single locally-optimized sets of primers (Morey et al., 2020). Combining primer sets in this way of course requires its own vetting and evaluation to achieve some desired level of taxonomic coverage and resolution. An ideal metabarcoding multiplex would resolve all of the species in a given regional target assemblage with as few primer sets as possible to control the added cost and labor of sequencing library prep associated with each additional primer set. The goal of my thesis was to evaluate the individual and combined performance of alternate primers sets for vertebrate and invertebrate metabarcoding in the Gulf of Maine, as part of developing infrastructure for the Maine-eDNA EPSCoR RII Track 1 research program.

The Maine-eDNA EPSCoR program is a National Science Foundation funded grant program to develop eDNA-based ecological tools that provide inference around coastal GoM challenges such as climate change, emerging aquaculture industries, species range shifts, and sustainable fisheries. As part of this research effort, investigators seek to apply metabarcoding to understand the distributions and dynamics of many GoM taxa, from microbes to white sharks and from lakes and rivers to the ocean. However, eDNA metabarcoding has not been widely applied to vertebrates or invertebrates in this region. To develop this eDNA metabarcoding capacity, one would ideally evaluate alternate primer sets against real-world eDNA samples containing the full regional species pool. However, that is logistically infeasible given the sheer diversity of species and habitats that they occupy. As an alternative, I made use of an approach involving eDNA metabarcoding of large-scale mesocosms in the form of multiple display tanks of the New England Aquarium. These displays contained distinct censused assemblages of fish and invertebrates characteristic of different habitats in the GoM region.

Thesis Chapters

In following two chapters, I outline my research and findings evaluating alternate and combined eDNA metabarcoding primer sets for resolving vertebrate and invertebrate diversity of the New England Aquarium. Chapter 2 evaluates three eDNA metabarcoding primer sets for vertebrates, the MiFish-U primer set developed by Miya et al. (2015), the 12S-V5 primer set developed by Riaz et al. (2011), and the MarVer3 primer set developed by Valsecchi et al. (2020). In that study I show the use of multiple metabarcoding assays in creating higher taxonomic resolution and detection within the given mesocosm, as well as the potential for eDNA to be used as a proxy for relative biomass representation in certain systems.

In the subsequent chapter I evaluate the performance of two primer sets for invertebrates, the 18S primer set E572F and E1009R from Comeau et al. (2011) that is targeted for general eukaryotes, and the COI primer set BF1 and BR1 from Elbrecht et al. (2017) that targets aquatic invertebrates.. I had originally sought to assess three primer sets for invertebrates the same way I had for vertebrates, but one primer set performed too poorly to assess (18S), and I did not have the resources or time to add the third (Leray et al., 2013). Nonetheless, I was able to evaluate the Elbrecht et al. (2017) primer set and show the possible application of the COI gene region in providing high taxonomic resolution and detection for Gulf of Maine invertebrates.

CHAPTER 2

EVALUATING ALTERNATE AND COMBINED ENVIRONMENTAL DNA METABARCODING PRIMER SETS FOR GULF OF MAINE VERTEBRATES

2.1 Introduction

The monitoring of biodiversity in our local and worldwide water systems is essential for ecosystem conservation and sustainability (Kelly et al., 2014). However, overall environmental monitoring is overwhelmed with factors that constantly change the composition and function of ecosystems; from climate change and alien invasive species, to habitat degradation and fishing practices (Watts et al., 2019). Effective strategies for monitoring ecosystem change rely on the ability of diverse approaches to characterize taxonomic assemblages, detecting rare, cryptic, or typically elusive species, and tracking movements of key species within and among habitats (Rees et al., 2014; Sard et al., 2019; Zhang et al., 2020). When these monitoring efforts fall short, there can be catastrophic biological consequences, as well as economics and other social issues (Salter et al., 2019).

Traditional methods of monitoring and sampling for many mobile taxa include capture-based methods such as trawls, seines, and tagging, as well as well as visual and shiptime surveys (Andruszkiewicz et al., 2017; Kelly et al., 2014; Salter et al., 2019; Zhang et al., 2020). Although these methods have been used for decades, they are often time consuming, costly, and require specific expertise for identification, and they are often associated with high false negative rates of detection (Jerde et al., 2011; Tyre et al., 2003). Additionally, these methods have been proven in some cases to be harmful to the species or habitats they're applied to monitor (Andruszkiewicz et al., 2017; Zhang et al., 2020). The need for rapid biodiversity surveying tools that improve upon traditional surveying is apparent (Deiner, Bik, et al., 2017; Kelly et al., 2014). Implementation of genetic tools like environmental DNA for the monitoring of community diversity offers possible alternatives to mitigate many of these concerns.

The use of environmental DNA (eDNA) in sampling for biodiversity, species monitoring, and conservation has grown significantly in the last decade due to its broad applicability across project types and its sensitivity in detecting rare or difficult to capture taxa (Deiner, Bik, et al., 2017; Morey et al., 2020; Thomsen & Willerslev, 2015). Because of the constant shedding of DNA by organisms into their environment, genetic material can be collected and used for molecular analyses that aid in species-specific targeted studies or larger biodiversity monitoring. This use of eDNA provides a method to substantially increase the amount of information we can gather on a less invasive scale. Not only is it cost effective, less invasive, and less destructive than traditional survey methods (Schmelzle & Kinziger, 2016), it has been proven to improve detection of many species over traditional methods (Dejean et al., 2012; Sard et al., 2019; Valentini et al., 2016). Environmental DNA metabarcoding in particular has become an increasingly popular genetic method for the multi-species assessment of community composition (Fonseca et al., 2010; Zhang et al., 2020). eDNA metabarcoding is a metagenomic method that uses universal primers to amplify a taxonomically informative gene region for subsequent next generation sequencing (Taberlet et al., 2012; Valentini et al., 2016). This method can be used to characterize presence of a large number of taxa in a sampling site without having to design genetic assays to detect each individual species and thus provides a potentially powerful tool for management, conservation and supplement to traditional survey methods. Indeed, eDNA metabarcoding has been applied in many survey settings and proven a useful and powerful tool in monitoring local biodiversity (Gold et al., 2021).

There exists a wide selection of possible primer sets (sometimes called "assays") when choosing to use eDNA metabarcoding for community composition and monitoring studies. Typically, the most commonly targeted genes for vertebrate metabarcoding are the 12S and 16S mitochondrial ribosomal subunit genes, as well as the cytochrome oxidase I (COI) mitochondrial gene (Zhang et al., 2020). Mitochondrial genes provide high copy numbers per cell to increase environmental concentrations, and these particular gene targets have the properties of being short, hypervariable sequence regions that are

capable of resolving many finer level taxa (e.g. species or genera) within broader taxonomic groups (Riaz et al., 2011; Valsecchi et al., 2020). There are now dozens of previously published metabarcoding assays available for a range of taxonomic groups from metazoans, bacteria, vertebrates, and invertebrates. Not only are the options broad, but the specificity of each option differs greatly between each publication. Peer reviewed, open source marker sets often suggests barcodes should be partly redesigned depending on the study or specific biological question being asked (Riaz et al., 2011; Taberlet et al., 2012). However, development of study-specific primer sets is very inefficient from the standpoint of time and resources, especially if eDNA methods are to be applied by a wider community of scientists and stakeholders. Moreover, the use of idiosyncratic primer sets strongly limits the ability to compare and combine data across studies and regions. Not surprisingly, many studies instead opt to apply a previously published primer set, often with good results (M. Miya et al., 2015). More recently, there has also been a growing recognition that combining (multiplexing) multiple existing primer sets might prove equally or more effective than designing a single, novel assay (Kelly et al., 2014; Morey et al., 2020).

Often, investigators seeking to apply previously published primer sets to a new regional species pool make their choice based on experiences elsewhere and some form of in-silico evaluation. Choosing primers sets in this fashion may not fully inform which assays will work best under field conditions and what biases might exist to influence inferences. Along these lines, while there have been large scale primer set comparisons for vertebrates (Zhang et al., 2020), it is important to note that these comparisons have all been conducted within a particular species pools and may not be fully relevant to other regional pools. Ultimately, empirical testing against samples with known species composition, such as from large mesocosms, is perhaps the most direct means to evaluate alternate assays. Large mesocosms with known species diversity, including aquaria displays, have been used to examine the efficiency of eDNA metabarcoding primer sets in sampling biodiversity of vertebrate taxa for various regions (Evans et al., 2016; Kelly et al., 2014; Morey et al., 2020). The advantage of these mesocosm approaches is they

provide strong positive control for assessing species detection and quantification, including possible gaps in detection of some taxa and how combinations of assays might cover such gaps.

The Gulf of Maine is one of the most rapidly warming coastal bodies of water on the planet and is the subject of a large scale NSF EPSCoR program to build eDNA-based ecological inference addressing climate change and other issues, including fisheries restoration, aquaculture, harmful algal blooms, and species range shifts. The objective of this study was to assess the single and combined performance of three commonly used metabarcoding primer sets for resolution of Gulf of Maine vertebrates, using the GoM specific displays of the New England Aquarium (NEAQ) as species assemblage mesocosms. Specifically, we sought to answer the following four questions:

1. How do the three primer sets compare in their ability to characterize known NEAQ species assemblages of vertebrates?
2. How much does species detection improve when combining the joint detection capabilities of two or more of these primer sets?
3. How much do additional sample replicates improve the species representation of each assay?
4. To what degree do the different primer sets capture approximate relative abundance information for key NEAQ/GoM taxa?

2.2 Methods

2.2.1 Literature review and assay selection

To evaluate efficiency and suitability of metabarcoding assays for resolving GOM/ NEAQ taxa, the literature was consulted to identify the most appropriate and widely-used vertebrate marker sets. Primers were selected from the literature based on taxonomic coverage with GoM species pools, specificity for vertebrates, and reported success in other studies. Final selected vertebrate assays included the MiFish-U set (hereafter referred to as 'MiFish') from Miya et al. (2015), the 12S-V5 set (hereafter

referred to as 'Riaz') from Riaz et al. (2011), and the MarVer3 set (hereafter referred to as 'MarVer') from Valsecchi et al. (2016). The MiFish-U set is the most-widely applied 12s primer set for marine and freshwater fishes in North America. A Google Scholar search of the terms "fish metabarcoding" and "vertebrate metabarcoding" suggests a minimum of 10 studies have applied this primer set for fishes. Additionally, it has been shown that the MiFish-U primer set is efficient in identifying closely related species (Miya et al. 2015). In a large review of fish metabarcoding primer sets, Zhang et al. (2020) found that the Riaz et al. (2011) set performed comparably to the MiFish-U set in total taxa resolved, and in a paper from Stoeckle et al. (2020) this primer set was shown to have potentially less primer mismatches to bony fishes than the MiFish-U set. Finally, the MarVer3 16S primer set was selected because it targets a different gene region than the two 12S assays, its performance in identifying a broader pool of vertebrate species including marine fishes and cetaceans (Valsecchi et al. 2016), and evidence that the 16S gene regions can outperform 12S and other genes for identification of vertebrates in some previously published studies (Morey et al. 2020).

2.2.2 eDNA water sampling

To evaluate the ability of these assays in identifying eDNA of fish and other vertebrates representative of the Gulf of Maine, water samples were collected from seven NEAQ exhibits that represent GOM species. The specific displays, with their abbreviation, volume and number of inventoried taxa were: the Stellwagen Boulder Reef community (BOUL, 15786 L, 7 vertebrate taxa), the Eastport, ME exhibit (E, 2120 L, 4 vertebrate taxa), Tidepool touch tank (TP, 9729 L, 3 vertebrate taxa), the Isle of Shoals exhibit (IS, 2120 L, 4 vertebrate taxa), the Stellwagen Sandy Bottom community (SS, 9464 L, 8 taxa), the Front Harbor Seal tank (HS, 181700 L, 1 taxon), and the Boston Harbor Islands/ Shorebirds exhibit (SB, 2180 L, 9 taxa). Across all seven displays, there were 30 inventoried vertebrate species present for detection. A list of all taxa and their respective populations at the time of sampling can be found in Appendix Table A1. Although all the sampled tanks in the NEAQ draw their source water from a Boston Harbor inlet, that water is treated by carbon filters and UV light before distribution to the tanks.

Water is not directly exchanged among the aquarium displays sampled in this study, however all are maintained by the same staff personnel, and tank openings and sumps are all located in the same hallway and are on recirculating systems. Hence, we expected the samples collected from each display would mostly amplify taxa present in each individual exhibit, with more minor amplification of species from the harbor or other displays.

Samples were collected using previously sterilized 1L Nalgene bottles. Prior to the sampling event, all Nalgene bottles were sterilized using a 10% bleach solution, and then rinsed thoroughly with tap water and finally DI water. Bottles were then UV sterilized for one hour before being sealed and placed into sterile bags to await use during sampling. Prior to being packaged into sterilized bags and coolers, one bottle per tank was filled with 1L deionized water to act as a “cooler blank” during field collection. During the sampling, water samples were collected from the surface of the display tanks by directly dipping the bottle into the display. At each tank, five 1L water sample replicates were taken, and the 1L deionized water blank was opened for approximately 30 seconds and then resealed. Samples were immediately stored on ice and transported back to the University of Maine, where they remained refrigerated until filtered within 48 hours of collection.

2.2.3 eDNA preparation

Sample filtration was carried out in eDNA-specific laboratory spaces that were sterilized with 10% bleach and UV lights prior to filtration. All filtration equipment was also bleached for at least 10 minutes, thoroughly rinsed and UV sterilized before filtering. To control for contamination during the filtration process, 1L lab blanks were filtered at the beginning and the end of each filtering event (n=3). Water samples were filtered through 47mm diameter glass fiber filters (0.7uM, Whatman) using a vacuum pump. Each 1L replicate was filtered entirely on its own individual filter, resulting in 6 GFF filters per tank (five replicates and one blank).

After filtration, filters were stored at -20°C until extraction, which occurred 10 days after collection. Extraction of eDNA filters followed the DNeasy Blood and Tissue kit protocol (Qiagen, CA).

Extraction of all samples was performed simultaneously, with the addition of an extraction blank to control for possible contamination during the extraction process (blank glass fiber filter that underwent the same extraction process as eDNA filters). During extraction, one sample was lost (HS03) due to an unexpected interruption. A resulting total of 34 environmental samples and 11 blanks were extracted. Extracted DNA was then stored at -20°C until used in sequencing for each assay, respectively.

2.2.4 PCR amplification and sequencing

Construction of libraries after DNA extraction took place at the University of Maine. Extracted DNA from each tank was first amplified using published PCR protocols, with some minor modifications for specificity and to attach the specific primer and overhang adapters (Appendix Table A2). In addition to all environmental samples (n=34), all negative controls were also included for PCR amplification, as well as a no-template control (NTC) containing nuclease free water in place of DNA template. PCR reactions for initial amplification occurred at a 20uL volume. All DNA samples were amplified using the KAPA HiFi HotStart ReadyMix (Roche Molecular Systems, Inc.). Successful amplification of environmental samples and confirmation of negative amplification for blanks and NTCs was visualized after initial PCR via gel electrophoresis. Any positive amplification for sample blanks or NTCs was carried through sequencing. Amplification products were size selected using the Zymo Select-a-Size DNA Clean and Concentrator MagBead Kit (Zymo Research Corporation, 2021) following manufacturer's protocol (for bead concentrations see Table A2) before the second indexing PCR was performed to attach Illumina indices and adapter sequences. Samples were amplified for indexing PCRs as follows: 12.5uL KAPA HiFi HotStart ReadyMix, 9uL nuclease free water, 1.25uL of Nextera Indexing primers (Illumina, Inc.) , and 1uL template DNA. PCR amplification for indexing of all samples followed an initial denaturation at 95°C for 10 minutes, followed by 8 cycles of 95°C for 10 seconds, 55°C for 30 seconds, 72°C for 30 seconds with a final elongation at 72°C for 5 minutes. Following indexing PCRs, samples were visualized for positive amplification and attachment of dual indices via gel electrophoresis,

and then cleaned again using the same size selection kit and protocol (Zymo Research Corporation, 2021) (for specific concentrations, see Table A2).

Following cleanup, samples were quantified on a QuBit 4 Flurometer (ThermoFisher Scientific) using a Qubit dsDNA HS Assay Kit and were then pooled to equimolar concentrations. Final libraries constructed for each assay were then quantified using QuBit and qPCR with the KAPA Library Quantification Kit (Roche Sequencing, Inc.), and amplicon sizes were verified using a high-sensitivity dsDNA assay chip run on a 2100 Bioanalyzer. Libraries were then paired-end sequenced on an Illumina MiSeq platform (Illumina Inc.) at the University of Rhode Island's Genomics and Sequencing Center using the MiSeq Reagent Kit v3 (600 cycle) to accommodate long read lengths and significant sequencing overlap. The amplicon fragment length for the MiFish-U, Riaz, and MarVer3 assays are 180bp, 112bp, and 245bp respectively. To improve low diversity library runs, 10-20% PhiX was included in each sequencing run.

2.2.5 Bioinformatic Pipelines

Demultiplexed paired-end sequence reads were quality filtering using a QIIME2 data pipeline. Chimeric sequences were removed and low-quality base pairs were trimmed using a Phred score cutoff of 20. Once trimming was completed, paired-end reads were aligned, merged and sorted into representative ASVs using DADA2. ASVs were in turn taxonomically assigned sample-by-sample using a blast-consensus method and sequence reference databases created for 12S markers, and a reference Silva database for the 16S marker.

The 12S metabarcoding sequences were run against a proprietary regional database containing a mixture of full mitochondrial genomes and 12S gene sequences (Thomas, D.W., 2021) due to previous publications showing that taxonomically constrained databases perform better (Gold et al., 2020; Stoeckle et al., 2020). The 16S metabarcoding sequences were run first against the Silva reference database (version 138.1) (Quast et al., 2013), due to broad availability of 16S sequencing data that covers the species of interest in this study, using the vsearch method in QIIME2. Due to poor assignment against the

Silva database, the 16S assay was then run against the same proprietary database as the two 12S sets, due to the availability of additional gene sequence coverage, using a blast-consensus method. Final taxonomic assignment for all unique sequences was determined via correspondence of identities from the given reference databases, blast comparison to the National Center for Biotechnology Information (NCBI)'s nucleotide (nt) database, and the known species list from the aquarium. Taxonomic assignment for all sequences based on the original reference database assignment versus the final taxonomic assignment given is seen in Appendix Tables B1, B2 and B3. Once all ASVs were assigned, they were collated into operational taxonomic units (OTUs) for further analysis in R.

Number of reads in OTUs for any filtration, extraction, or field blanks were then compared to the number of reads in samples with the corresponding OTU. For those OTUs that had reads in blanks, the maximum number of reads from the blanks was subtracted from all environmental samples for each assay (Appendix Tables B4 and B5). Subsequently, all reads from non-vertebrate taxa, non-aquatic vertebrates, or a species not known on the census list were filtered for most subsequent analyses of "off-target" detections (Appendix Table B6).

2.2.6 Data Analysis

To better understand how our empirical primer evaluation compares to the more common in-silico primer selection approaches used by many investigators, we conducted an in-silico evaluation for the known NEAQ taxa. To do so, taxonomy lists obtained from the aquarium for each of the displays sampled were compared against the reference databases used in this study to determine coverage of represented taxa. Additionally, available gene sequences for any known species were downloaded from NCBI and aligned against each primer assay using the online web software Benchling (Benchling, 2021). Efficient primer compatibility to the reference sequences was determined by having less than 3 base pair mismatches on any given primer. All subsequent analyses were conducted using the software R (version 4.1.1).

Overlap in vertebrate taxa detected with different primer sets was evaluated using venn diagrams to display differences in taxonomic resolution using the package ggVennDiagram, as well as through the use of permutational ANOVA. PERMANOVA was run to compare samples and their taxonomic resolution both among assays within display tanks and among tanks within assays, with tanks set as the permutation, using the package adonis2. To understand the effects of compounding multiple assays for use in species detection we began with the assay having the largest individual species list, and then quantified the increase in taxonomic coverage afforded by each subsequent assay in order of their additional contributions. Because our goal was to assess the capacity of the different primer sets to resolve a known species assembly, we constrained these analyses to only taxa present in the NEAQ census.

To understand sample level reproducibility for taxonomic resolution, rarefaction power analyses were performed on all assays based on the independent replicate samples collected within each tank. For this analysis, only sequences relating to known census taxa for a given tank were included. Rarefaction effects were determined via bootstrapping individual samples to obtain species lists based on 1-5 samples using 1000 sample randomizations in a for loop in baseR. We in turn enumerated the proportion of randomizations in which the full set of genetically detectable taxa (based on empirical total across all samples) were detected. This approach was chosen instead of traditional rarefaction analysis based on number or proportion of total taxa detected, to take advantage of our known census data and to account for the fact that our results show that not all census taxa are empirically detectable with our primer sets. Additional traditional rarefaction on sequencing read depth and replication in revealing species number were also completed in the same manner (see Appendix B).

To address the question of eDNA amounts being a predictor for fish biomass in a given sampling area, we considered all sequences that were directly related to the vertebrate taxa known to be present in each tank. For the purposes of this relationship, the harbor seal tank (HS) was removed from the analysis, as it only has one vertebrate species to detect. Fish mass was estimated using the Bayesian length-weight conversion calculation (fishbase.de). Average lengths of each fish species, as well as coefficients for the

Bayesian length-weight formula were obtained from the reference database Fishbase (fishbase.de). Average mass for harbor seals was provided by NOAA Fisheries (Fisheries, N., nd), while those for bird species were obtained from Cornell's Ornithology Lab (Cornell Lab of Ornithology, 2019). Biomass for all species was then calculated using the average weight of the organism multiplied by the number of organisms in a given tank. For sequence read counts, all values were log10 transformed.

Both Pearson and Spearman's rank correlations were calculated for the MiFish and MarVer assays to estimate the relationship between fish biomass and sequence count using the stats package in R. Correlations were calculated individually on a tank-by-tank basis. Finally, Fisher's method for combining p-values was performed as a meta-analysis to combine tank-specific tests for each primer set.

2.3. Results

2.3.1 In-silico analysis

Of all the species censused in the NEAQ, all but two had representative sequences on either the NCBI nt database or Silva138. The two that did not have any available full mitochondrial, 12S, or 16S genes for reference were the piping plover (*Charadrius melodus*) and the chain cat shark (*Scyliorhinus retifer*). Their genera, however, are represented in those databases as well as within the proprietary database used in this study.

For the 12S proprietary database used in this study, it was found that all but piping plover (*Charadrius melodus*), chain cat shark (*Scyliorhinus retifer*) and ocean eelpout (*Zoarces americanus*) were represented, therefore leading to an expected resolution of 27 of 30 species when using the proprietary database alone for both 12S sets used here. However, the genera *Charadrius* and *Zoarces* were represented, leading to a possible resolution of 24 genera out of 25 represented in the aquarium. Additionally, in silico primer binding testing found that the MiFish set had the capability to effectively bind to 18 of the 30 species, while the Riaz set was considered to effectively bind 27 species and 1 possible genus level detection (represented by the genus *Zoarces*) for the represented reference sequences of the census species (Appendix Table B7).

For the MarVer set, it was found that only 8 of 30 species were represented in the Silva database used, and 18 genera out of the 25 possible. When comparing against the proprietary database also used for the 12S sets, it was found that When the MarVer set was tested for in-silico primer binding against the species in this study, it was found that it would be successful for all species except those that did not have any representative 16S sequences, which were the following: *Fundulus majalis*, *Myoxocephalus aenaeus*, *Myoxocephalus octodecemspinosus*, and *Ulvaria subbifurcata*. This results in a possible of 24 of the 30 species being resolved by the MarVer set based on in-silico primer binding, and one possible genus level detection (represented by the genus *Scyliorhinus*).

2.3.2 Sequencing results

Between the three sequencing runs, there were anywhere from 33,931,998 - 45,792,818 total reads generated on the Illumina MiSeq (Table 2.1). After quality filtration, there were between 7,506,159 and 8,703,949 generated sequences per run (Table 2.1). Although no NTCs were amplified during PCR, some lab, filtration, and extraction blanks were amplified for the MiFish and Riaz primer sets (n=6). These blanks had between 342,329 and 721,019 number of reads accounting for between 3.93-8.96% of total filtered reads (Tables B4 and B5).

Table 2.1 Summary table of sequence processing read counts during the bioinformatic pipeline and data cleaning steps. Non-relevant taxa refers to any non-vertebrate taxa, as well as un-censused or contaminant vertebrate taxa.

	MiFish	**Riaz**	**MarVer**
Total Reads	45,792,818	33,931,998	36,600,726
Pair-merged sequences	8,807,055	8,173,055	7,899,864
Sequences post quality filtration	8,703,949	8,045,041	7,506,159
Sequences from Blanks	342,329	721,019	-
Total ASVs	1035	226	1017
Sequences of non-relevant taxa	1,354,887	6,414,223	2,302,918

Table 2.1 cont.

Sequences after removal of non-relevant taxa	7,349,062	1,630,818	5,203,241

2.3.3 Taxonomic resolution

Using combined species assignments from the bioinformatics pipeline and manual secondary assignments using NCBI's nt database, a total of 36, 37 and 73 total unique taxa were detected using the MiFish, Riaz and MarVer primer sets. Of the 30 vertebrate taxa censused in the aquaria sampled (Appendix Table A1), 20 were detected using the MiFish primer set, 19 using the Riaz primer set, and 22 with the MarVer3 set (Tables B1, B2, B3). In combination the assays resolved all but 2 of the aquarium taxa. The only missing detections across all assays were for the radiated shanny (*Ulvaria subbifurcata*) and piping plover (*Charadrius melodus*). Among all three metabarcoding primer sets, 13 species were jointly-resolved (Figure 2.1), which is roughly the null expectation assuming each assay resolved approximately ⅔ of the species pool. Likewise, pairwise comparisons of metabarcoding primer sets resulted in 1-3 overlapping species IDs (Figure 2.1). Between all three metabarcoding sets, 24 of the 25 possible genera were resolved, resulting in between 1-24 overlapping genera level detections between assays (Figure 2.2).

Across all markers, the majority of taxa were resolved to the species level, with only 1-3 taxa providing ambiguous species assignments necessitating a genus-level determination. The MiFish primer set had 19 identifications to species (of which 2 were unique to this set) and one to genus level (Table B1). Unique identifications for this primer set included the Longhorn sculpin (*Myoxocephalus octodecemspinosus*), and the grubby sculpin (*Myoxocephalus aenaeus*). The Riaz primers had 17 identifications to species (1 unique to this set) and 2 identifications to genus (Table B2). Unique to this primer set was successful identification down to species for the common tern (*Sterna hirundo*). The identification of the common tern along with with the genus *Calidris*, and the species *Charadrius semipalmatus*, shows the Riaz marker set has the ability to resolve at least some marine birds. The

MarVer primer set had 19 identifications to species and 3 identifications to genus (Table B3). Unique to this marker was the identification of the genus Scyliorhinus, representing the chain cat shark (*Scyliorhinus retifer*). The identification of this genus, along with successful species identification of the little skate, indicates this primer set is able to able to resolve some chondrichthyes.

A permutational analysis of variance (PERMANOVA) based on assay and tank effects was overall significant (R2 = 0.345 and p=0.001). Both different assays (R2=0.090 and p=0.001) and different tanks (R2 = 0.324 and p=0.001) were associated with significantly different community profiles. However, the estimated tank effect size was much larger than the assay effect (based on approx. R2), suggesting that although assays do provide somewhat different representations of communities, this is secondary to overall power of all assays for distinguishing different GoM vertebrate communities.

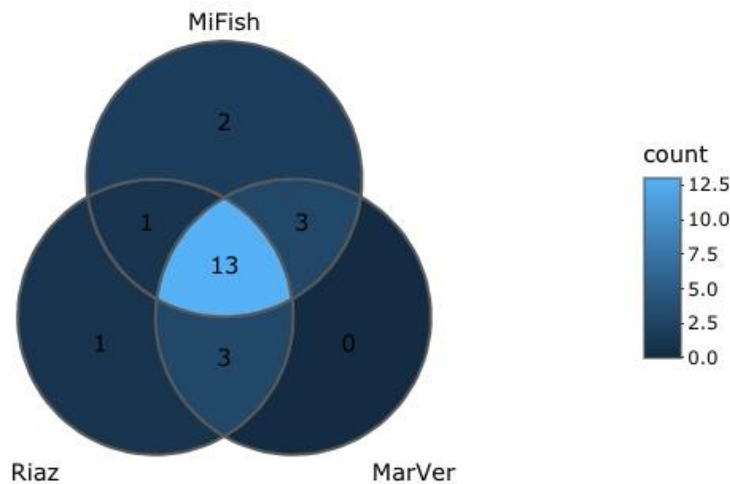

Figure 2.1 Venn diagram of species resolution and overlap of all three assays in this study. Lists of species detected in this study can be found in Tables B1, B2 and B3.

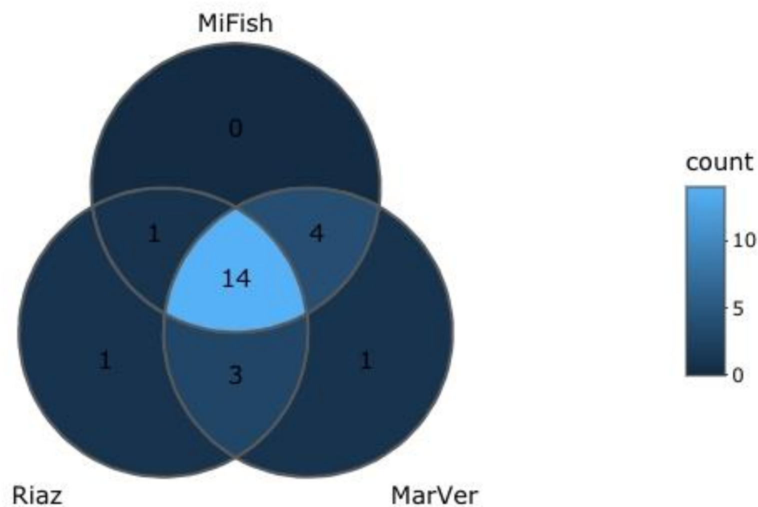

Figure 2.2 Venn diagram of genera level resolution and overlap of all three assays in the study. Lists of genera detected in this study can be found in Tables B1, B2 and B3.

2.3.4 Effects of assay compounding

When looking at vertebrates from the New England Aquarium, both the MarVer and MiFish sets performed equally when used on their own, each resolving 63.3% of species (19 out of 30), but differed in the particular species resolved. Specifically, the MarVer set resolves the species *Charadrius semipalmatus*, *Helicolenus dactylopterus* and *Leucoraja erinacea* while the MiFish instead resolves the species *Fundulus majalis*, *Myoxocephalus aenaeus* and *Myoxocephalus octodecemspinosus.* Combing the MiFish and MarVer sets increased species coverage by an additional 10% (3 species) for a total of 73.0% species coverage (22/30; Figure 2.3). Addition of the Riaz primer set as a third compounded assay only added one more species (*Sterna hirundo*) for a total resolution of 76.6% (23/30). Combining the MiFish and Riaz primers produced the same total species coverage with no species added by including MarVer primers.

Of the 25 genera present in the aquaria sampled, both the MiFish and Riaz sets resolved 76% (19/25) of the taxa, while the MarVer set resolved 88% (22/25) (Figure 2.4). When looking at compounding assays for genus level resolution, the addition of the Riaz marker to the MarVer primers added 2 genera for a total of 96% (24/25) coverage (Figure 2.4). The addition of the MiFish set as an

alternative to the Riaz primers was only marginally worse 92% (23/25). Adding the MiFish primer set to the MarVer and Riaz sets did not resolve any further taxa. If looking at the differences between the MiFish and Riaz sets for genera resolution, the former resolves the fish genera *Hippoglossus*, *Morone*, *Myoxocephalus*, while the latter instead resolves birds in the genera *Calidris*, *Charadrius* and *Sterna* as well as the skate genus *Leucoraja*.

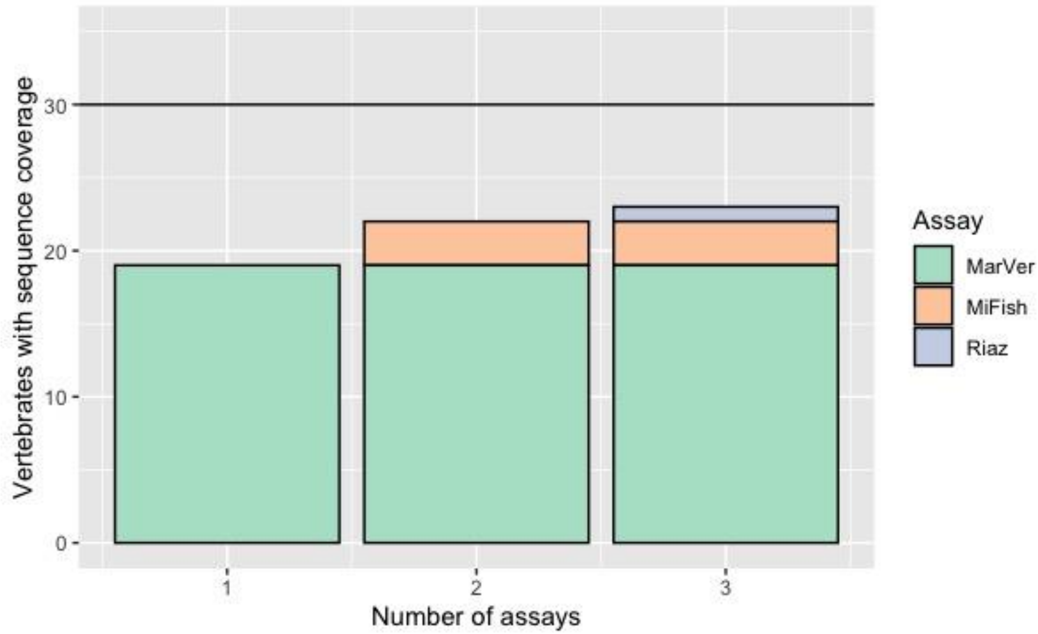

Figure 2.3 Vertebrate species accumulation with the compounding of metabarcoding assays. The solid line indicates the maximum number of vertebrate species present in the NEAQ aquaria sampled (n=30).

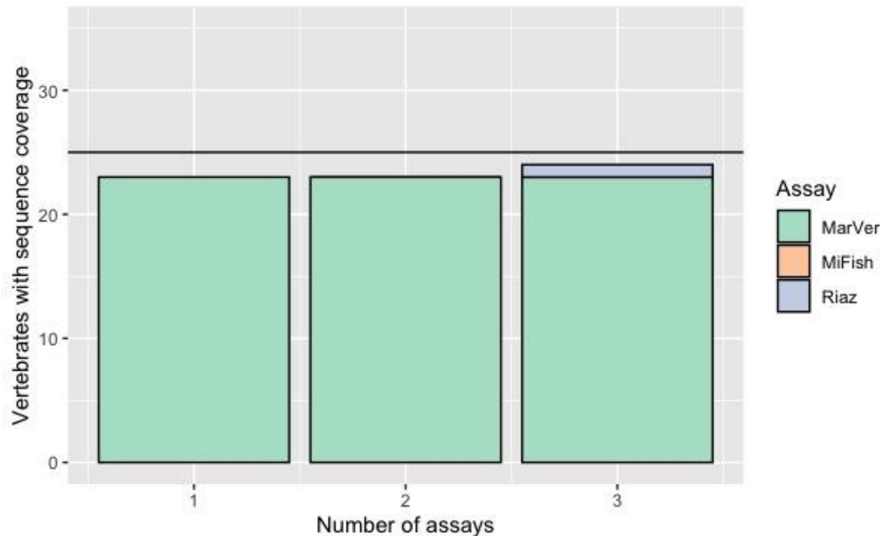

Figure 2.4 Vertebrate genera accumulation with the compounding of metabarcoding assays. The solid line indicates the maximum number of genera that could be detected in the NEAQ aquaria sampled in this study (n=25).

2.3.5 Biomass and eDNA relationships

For the MiFish 12S set, neither the linear nor rank relationship between species biomass in a tank and sequence read count was statistically significant (Figure 2.5, Table 2.2), but showed modest positive trends for 3 out of 4 display tanks with a minimum of 3 taxa. When a meta-analysis was conducted on all rank significance values across tanks, the relationships between biomass and read counts was not statistically significant (χ^2=8.523555, p= 0.7429959, df=12).

Table 2.2 Spearman's rank correlation coefficients and p-values for all tanks and all assays

	MiFish			**Riaz**			**MarVer**		
Tank	Rho	p-value	S	Rho	p-value	S	Rho	p-value	S
Boulder	0.6	0.247	14	*NA*	*NA*	*NA*	0.829	0.058	6
Eastport	1	1	2.22*10^-16	*NA*	*NA*	*NA*	1	0.333	0
Isle of Shoals	-0.5	1	6	*NA*	*NA*	*NA*	0.5	1	2
Shorebirds	1	0.333	0	-0.5	1	6	1	0.083	0
Sandy Shoals	0.657	0.175	12	-1	1	2	0.657	0.175	12
Touch Tank	-1	1	2	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*

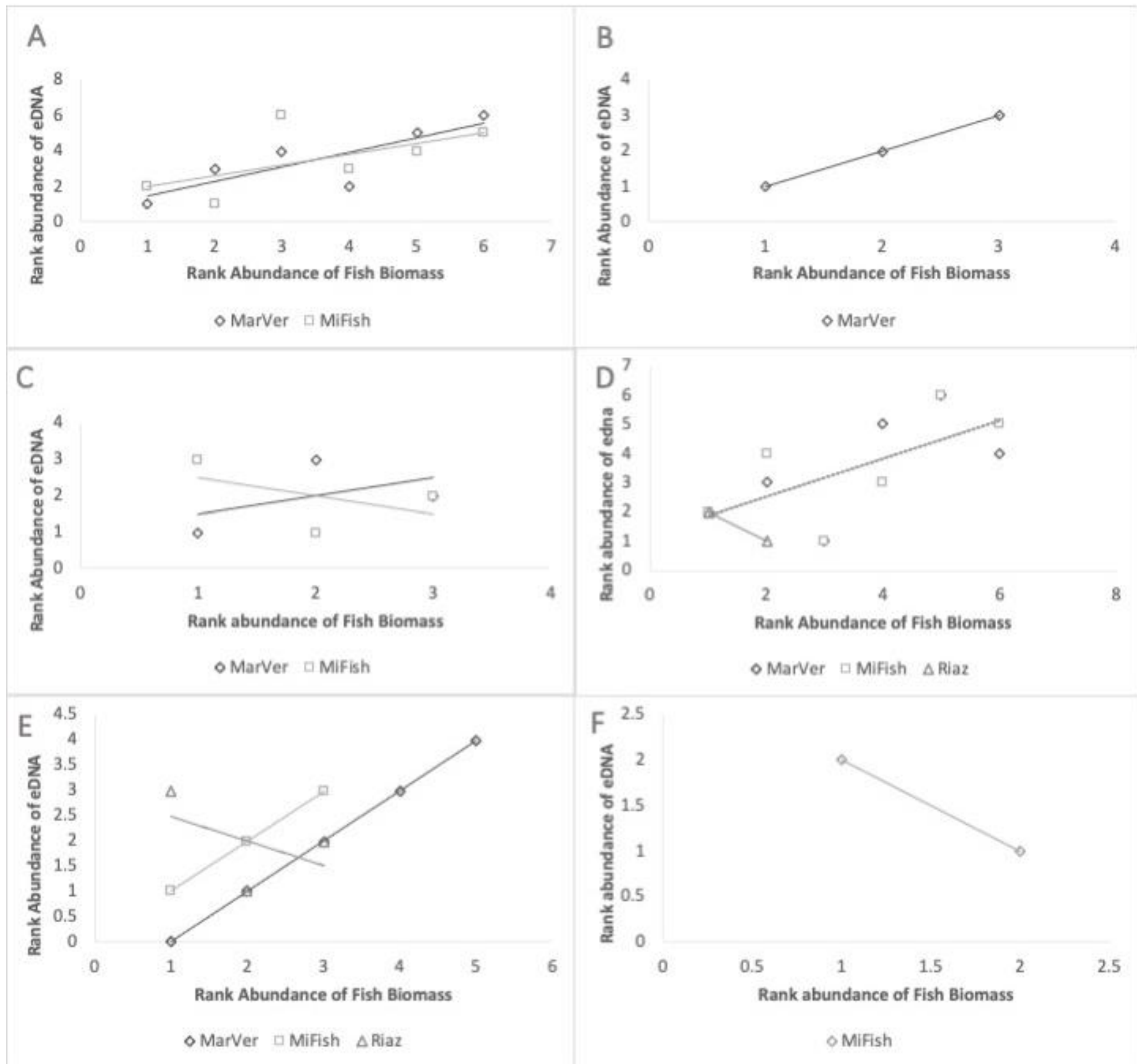

Figure 2.5 Spearman's rank correlation plots for each tank in the study for both the MiFish, Riaz and MarVer assays; a.) Stellwagen Boulder Reef b.) Eastport, ME c.) Isle of Shoals d.) Stellwagen Sandy Bottom e.) Boston Harbor Islands/ Shorebirds f.) Touch tank

For the MarVer 12S set, none of the linear or rank relationships between fish biomass in a tank and eDNA read count were individually significant, albeit 5 out of 5 tanks with 3 or more taxa showed positive trends in rank values (Figure 2.5, Table 2.2). When using the Fisher's method for meta-analysis

of p-values from the Spearman's correlation calculations, the relationship between biomass and eDNA count was marginally significant (χ^2=16.33 df=10 p=0.09).

For the Riaz primer set, the relationship between fish biomass and eDNA sequence count could only be estimated for two out of the 6 tanks that were sampled (SS and SB) and only for two species per tank. The statistical relationship between biomass and eDNA count could not be statistically tested with only two points per tank, but was negative in each instance (Table 2.2).

2.3.6 Effects of replication

Rarefaction power analyses on the three primer sets revealed that our sample replication most often readily provided enough power to resolve the full set of genetically-resolvable species with five or fewer samples. Across all three sets, when doing rarefaction power analyses on the probability of identifying the full set of detectable fish in a given tank most assays and tanks achieved full survey power between one and three samples (Figure 2.6). In some tank assay combinations, full detection power was never fully reached, such as for the Eastport and Touch tanks when sequencing with the Riaz primer, the Sandy Bank tank for the MarVer primer set, and the Isle of Shoals tank for the MiFish primer set.

Rarefaction analysis on sequence count and species saturation revealed that all replicates within individual tanks reach saturation after relatively few sequences compared to overall sequencing depth (Appendix Figure B3). However, rarefaction analysis to understand number of species detected with additional sampling replicates revealed that saturation of number of species did not occur with increased sampling replication across all three assays and all tanks (Appendix Figure B4).

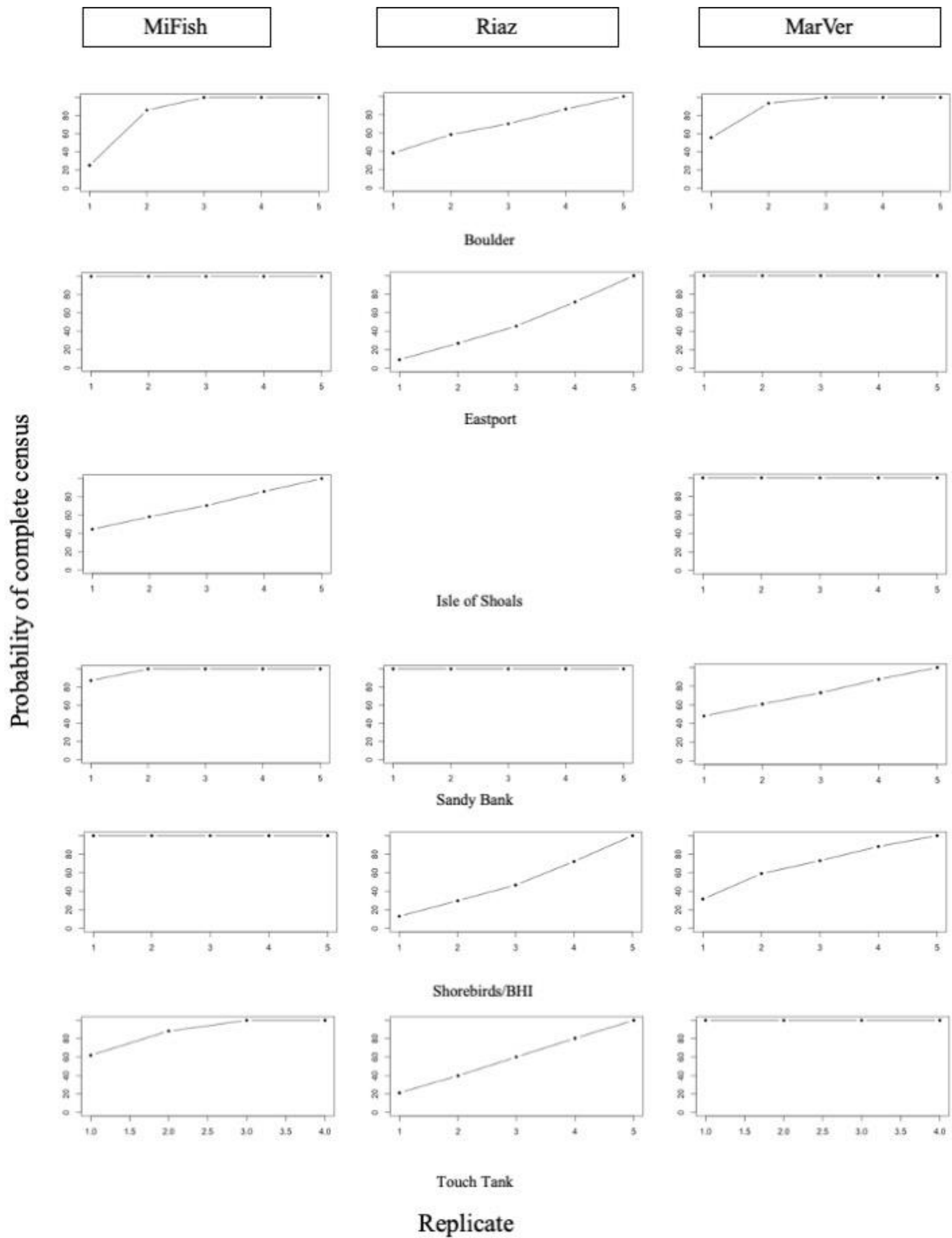

Figure 2.6 Rarefaction power curves for all three primer sets in each tank that was sampled in the study; y axis represents the probability of detecting all of the species present on the NEAQ census, while the x axis represents the number of sample replicates.

2.3.7 Off- target identification

Approximately 15.6% , 79.7% and 30.7% of the total sequence reads were assigned to off-target (non-aquarium) taxa for the MiFish, Riaz and MarVer primer sets respectively (Table 2.1). These off-target sequence reads were attributed to eDNA from unaccounted taxa in the actual displays, eDNA from water intakes, eDNA from feed sources (Table B6), human and other terrestrial vertebrate contamination, sampling equipment contamination, or in lab contamination. The first three of these sources represent actual eDNA in the sampling environment, and are not true contamination per se. The same can be said of human and terrestrial vertebrate eDNA to the extent it was present in the displays, rather than entered during the subsequent sample processing. However, we encountered some DNA contamination tied to equipment and processing, which was accountable by sequencing negative field and lab controls (Table B4 and B5). Removal of the maximum number of reads found in any of our blanks from any associated aquarium sample eliminated between 54-74% of the taxa identified in sequencing, but none of the actual census taxa.

The majority of the off-target identifications for the MiFish assay came from bacterial amplification, while common vertebrate contaminants and miscalls, including human, mouse, pig, and cow were also identified. White sucker (*Catostomus commersonii*), brook trout (*Salvelinus alpinus*), Atlantic salmon (*Salmo salar*), alewife (*Alosa pseudoharengus*), black crappie (*Pomoxis nigromaculatus*), common minnow (*Phoxinus phoxinus*), and fall fish (*Semotilus corporalis*) were also found to be present within the blanks of the MiFish assay. These detections occurred in the extraction, lab and 2 aquarium blanks and represent probable in-lab contamination from other fish research at UMaine. It is also possible that some of the Atlantic salmon contamination is accounted for by commercial fish feeds. Other fish DNA in aquarium feeds likely accounted for some remaining off-target species reads, including detections of capelin and Atlantic silversides (Appendix Table B4).

For the Riaz primer set, the majority of off-target amplification came from human DNA and bacterial amplification, while common vertebrate species including mouse, cat, dog, and pig accounted

for another large portion of the sequencing reads. Common Boston Harbor species were also detected in some tank replicates but not in any extraction or lab blanks, including Atlantic cod and herring. It is possible that these detections can be attributed to the unknown contents of contaminant fish feed, or commercial dry fish feed as well. Finally, in lab contamination was also detected due to the presence of common minnow (*Phoxinus phoxinus*) in some of the samples (Appendix Table B5).

Although the MarVer3 primer set did not have any form of blanks amplify positively for eDNA, it still showed evidence of off-target contamination from laboratory, feed and in-aquarium sources. Specifically, this included many species of algae, the same common vertebrate contaminants (mouse, human, pig), as well probably species from feed used in the aquarium (Appendix Table B6). In addition to contaminant taxa, the MarVer primer set had reads from a variety of invertebrates that are known to be present in the aquarium tanks sampled, including orange-footed sea cucumber (*Cucumaria frondosa),* green sea urchin *(Stronglyocentrotus droebachiensis),* purple sea urchin *(Arabacia punctulata),* common northern whelk *(Buccinum undatum),* purple sun star (*Solaster endeca),* Northern sea star (*Asterias rubens),* spiny sun star (*Crossaster papposus),* and the genera *Hippasteria* and *Henricia*.

2.4. Discussion

In this study, I surveyed seven different tanks from the New England Aquarium, and used three different vertebrate primer sets to test the efficiency of metabarcoding for eDNA based biodiversity studies for Gulf of Maine vertebrates. By using mesocosms with censused assemblages, variables that are uncontrollable in most field studies, such as species composition and abundance, volume of water, and source contamination, were accounted in assessing assay performance. Overall, I found that the three widely employed vertebrate assays resolved comparable numbers of species and genera, but differed in some taxonomic biases. In combination, these assays were able to resolve a decisive majority of the represented species, and nearly all of the represented genera, in the NEAQ aquaria. Positive among-taxa rank order biomass to eDNA trends were apparent in the majority of the aquarium tank assemblages with

three or more taxa for the MiFish and MarVer assays suggesting potential capacity to quantify relative biomass dominance in particular assemblages. Rarefaction modelling of the sampling replicates from each display showed that our sampling design was mostly sufficient to detect all possible taxa that amplified with each given assay. For some tanks, rare taxa possibly caused lack of saturation with sampling replicates. In the following sections I take up each of these findings with respect to my original study goals and put them into context for the field and Maine-eDNA research program.

2.4.1 Assessing primer performance for Gulf of Maine vertebrates

Taxonomic specificity can be an important factor for certain studies using eDNA metabarcoding as a biodiversity tool, especially when focused on one taxonomic group such as fish (Zhang et al., 2020). For the purposes of this study, somewhat broad taxonomic coverage was important when understanding how chosen primers could be applied at a whole ecosystem level, which would include marine vertebrates outside of traditional fish-specific surveys. Due to factors such as primer bias, sequencing depth, and DNA shedding rates, the successful recovery of target taxa can vary greatly across marker sets (Kelly et al., 2014). We found that in a controlled mesocosm study of 30 Gulf of Maine species, detection success across genetic primer sets for vertebrate species varied modestly for total taxa, with all assays resolving roughly 2/3 of the total species pool. However, the 16S MarVer primer set used in this study consistently performed as well or better than both the MiFish and Riaz sets (both located on the 12S gene). This result is consistent with a study by Morey et al. (2020) that compared 12S and 16S markers in aquarium detections (Ripley's Aquarium, Toronto, Canada). Of the taxa with reference sequences in our study, the 16S assay was able to detect 76% of those present, while the 12S sets were only able to detect 63% each. The MarVer assay identified 19 taxa to the species level, and had 3 genera level detections, cumulatively surpassing both other sets in its ability to detect taxa present in the tanks sampled. This is in contrast to other studies that have previously reported 12S identifications between 80- 100% of known present taxa

(Evans et al., 2016; Kelly et al., 2014; M. Miya et al., 2015). The success rate of identification in this study is, however, greater than that of a recent aquarium mesocosm study in which the 12S markers performed poorly in comparison to these previous studies, uncovering only 13% of species with available reference sequences (Morey et al., 2020).

The markers used in this study resolve comparable amounts of taxa to both species and genus (Figures 2.1 and 2.2). Most identifications for all three markers were made down to species level, and those identified to genus seem to be unique to the specific primer set, indicating its applicability for certain taxonomic groups (Valsecchi et al., 2020). More specifically, the MarVer assay was able to resolve the genus *Scyliorhinus* in addition to successful identification of the little skate, *Leucoraja erinacea*. The detection of the two elasmobranchs present in the aquaria sampled suggests that this marker may be successful for use on other elasmobranchs in the Gulf of Maine. This result is supported by stated capabilities of this primer set to detect an array of elasmobranchs in its original development study (Valsecchi et al., 2020). In addition to the detection of these elasmobranchs, this primer set was successful in identifying the genus Calidris, as well as having identification to species for the semipalmated plover, *Charadrius semipalmatus*. This is suggestive of this primer set's ability to detect possible bird species of interest in GoM.

When comparing the two 12S sets for taxonomic resolution, the MiFish marker is the superior when identifying to the species level, but there is overall greater richness when using the Riaz marker for genera level detections. The Riaz assay was able to distinguish an entire vertebrate group, marine birds (including those in the genera *Calidris*, *Charadrius* and *Sterna*), that the MiFish assay was unsuccessful in resolving. As a tradeoff, the MiFish assay had a greater richness in the actinopterygian fish specific taxa it amplified. This greater ability to resolve marine fishes has been shown in other comparative studies (Zhang et al., 2020) and likely derives from the original taxonomic intent of the two primer sets. Specifically, the Riaz assay was designed to include some terrestrial vertebrates (Sarcopterygii), while the MiFish assay was designed for marine fishes (Actinopterygii), of which this study is mainly composed

and focused on. Depending on the study at hand, taxonomic generalization and breadth of taxa may be useful, such as when trying to characterize whole communities including fish, marine mammals, and other marine vertebrates relevant to an ecosystem. However, the potential for large amounts of non-target species can greatly affect desired results, especially when using broad range primers. It was shown in this study, as well as those previously using the Riaz assay that it strongly amplifies human DNA (Kelly et al., 2019; Stoeckle et al., 2020; Zhang et al., 2020). This could greatly reduce the amount of target amplification occurring if it is preferentially binding to human DNA sequences, which are very difficult to restrict from many eDNA field sites or workspaces. To increase fish and general marine vertebrate identification, the primer assay could be used with a human-blocking oligo in future work, which have been proven highly effective in reducing overall amplification of human sequences (Zhang et al., 2020).

Based on the results in this study, we adjudge a combination of primer sets can provide substantive improved species or genus level resolution, as well as taxonomic coverage. The greatest increase of species or genera is of course obtained by adding a second assay to the first, with the benefit of adding a third assay being much more limited in terms of return on investment. For species level detections, the maximum number of taxa resolved occurred when using all 3 of the genetic markers in this study (Figure 2.3). Although the respective primer sets identified to similar species richness levels (19 for both the MarVer and MiFish, 17 for the Riaz), addition of the MiFish set to the MarVer (or vice versa) provided an increase in 3 unique taxa. With the addition of the Riaz marker, another species is added that is unique to this marker only. However, when looking for genera level detections the combination of the MarVer and Riaz assays was sufficient to maximize genera coverage in the study, the MiFish assay not resolving anything unique from that pairing. That said, the choice of which two assays to combine is perhaps not as clear as might be implied by this tabulation. The total number of taxa in this aquarium study is well below that of most GoM habitats. As such, subtle differences of 1 or 2 unique taxa from one combination of primer sets versus another in this aquarium study should not be extrapolated to larger

natural communities. It may thus be more important to select primer sets based on their complementary taxonomic biases or ability to resolve unique GoM communities.

Along these lines, when looking at how assays compare in PERMANOVA, it is shown that while there is a statistically significant assay effect, it accounts for much less taxonomic variation than different tank communities. This adheres to and supports the fact that all of the primer sets overlap in the majority of the taxa they resolve and are thus apt to discern different vertebrate communities in a relatively parallel fashion. This is not to say that there are not potential benefits to compounding assays, but doing so is not likely to be a limiting factor for statistically discerning community profiles from very different habitats. On the other hand, combining assays could be much more useful for characterizing more subtle habitat gradients. Yet, it is also important to note that in attempts to use multiple markers for detection, studies have often run into bioinformatic challenges that have not been faced with using multiple metabarcoding assays such as error removal and marker specific parameterization (Morey et al., 2020). Therefore it is also important to consider the use of bioinformatics pipelines and reference databases in conjunction with assay choice when designing and implementing a metabarcoding study.

2.4.2 Comparison to in-silico expectations

Since there is such a wide choice of eDNA metabarcoding primers in the published literature, investigators often find themselves performing in-silico analyses to determine potential primer biases for the particular study community of focus (Zhang et al., 2020). Primer amplification and specificity can vary greatly for any given taxonomic group or regional assemblage (Clarke et al., 2014). As discussed previously, high taxonomic specificity typically comes with a tradeoff of reduced taxonomic breadth. Additionally, identifications of target taxa rely on accurate and comprehensive reference libraries (Stoeckle et al., 2020). For the purposes of this study, in-silico analysis of primer sets against all known target vertebrate species was performed by aligning primer sets against the target genes in Benchling, as

well as checking the reference databases used in this study for all species coverage from the known NEAQ list. As noted in previous studies, in-vitro outcomes for metabarcoding primers often fall short of in-silico predictions for amplification (Zhang et al., 2020). In our study, in-silico expectations over-predicted the amount of vertebrate amplification and detection by 27% for the Riaz primer assay and 5% for the MarVer assay. In contrast, it under-predicted the amount of vertebrate amplification that would be seen from the MiFish assay by 17%. Specifically, the Riaz assay did not successfully identify previously predicted amplifications from the in-silico analysis for *Hippoglossus hippoglossus*, *Morone saxatilis*, *Phoca vitulina*, *Scyliorhinus retifer* or anything within the genus *Myoxocephalus*. Additionally, it only successfully amplified birds within the genus *Calidris* to the genus level, when previous in-silico expectations were identifications to species. The MarVer assay did not identify the species *Fundulus majalis* or *Sterna hirundo*, and could only identify to genus the species present within *Myoxocephalus*. In contrast, the MiFish assay successfully identified *Centropristis striata*, *F. heteroclitis* and *F. majalis, Hydrolagus colliei* and *Microgadus tomcod* when previous in-silico analysis predicted suboptimal primer matches (> 3 mismatches in a primer).

As noted previously, different primers sets were designed with different taxonomic goals, ranging from fish-specific amplification (Miya et al., 2015) to terrestrial vertebrates (Riaz et al., 2011), to some combination of the two (Valsecchi et al., 2020). However, the range of taxa amplified by given metabarcoding primer set can often be broadened or narrowed by adjusting the stringency of the amplification conditions. In particular, choice of annealing temperature can significantly affect the specificity and breadth of taxa one identifies with any given primer set (Clarke et al., 2014). In this study, PCR reaction conditions mostly matched those of original studies, with some modifications to improve specificity against non-target taxa shown to amplify in previous studies using these assays. Those changes equated to switching from a constant 65°C annealing temp (Miya et al., 2015) to a touchdown PCR peaking at 69.5°C for the MiFish assay, and an increase from an original 57°C touchdown PCR (Valsecchi et al., 2020) to a constant 60°C annealing temp for the MarVer assay. These relatively

stringent conditions are likely part of the reason why we resolved somewhat lower total taxa than some prior studies (e.g. M. Miya et al., 2015).

Certainly, another key piece of any successful metabarcoding assay is the reference database and bioinformatics pipeline it is paired with. As noted in the studies by Stoeckle et al. (2020) and Gold et al (2020), comprehensive sequence libraries can maximize our ability to gain information from eDNA metabarcoding data, noting the use of curated reference libraries in increasing success. We implemented a mixed method for identifying the reference sequences generated by our bioinformatics pipeline, by using a curated reference database for relevant Northeastern USA organisms as well as large, open-source databases such as NCBI's nt database and the Silva 138 database for eukaryotes. This proved a successful verification process for assignment in the bioinformatics pipeline, as well as useful in covering almost the entire range of targets in this study. However, it is recommended based on the results of this and previous studies that reference databases based on known locally sequenced tissue samples are curated for taxa of interest. This may greatly reduce limitations in studies from lack of relevant species in the databases available, uncertainties associated with regional hybridization caused from introgression of mitochondrial genomes or overall primer bias leading to erratic detections (Deiner, Renshaw, et al., 2017; Stoeckle et al., 2020).

2.4.3 Effects of replication in detection of vertebrate taxa

Environmental DNA is often heterogeneously distributed in the environment, due a diversity of factors related to the ecology of organisms and eDNA itself, such habitat preferences, flow patterns, turbulence, and variable eDNA shedding rates (Deiner, Renshaw, et al., 2017; Morey et al., 2020). These processes contribute to variability in the eDNA that is ultimately captured in a given water sample and are complemented by additional subsampling heterogeneity associated with lab processing (e.g., pipetting). As such there is value in considering what sampling volumes or sampling replication are needed to

achieve a desired degree of taxonomic coverage or reproducibility (Bessey et al., 2020). There is as yet no set standard volume for eDNA samples, and such requirements may vary somewhat depending on the inferential goals of a given study. However, in many previous eDNA metabarcoding studies, both in mesocosms and in field environments, a sample volume of 1L has proven effective (Andruszkiewicz et al., 2017; Gold et al., 2021; Lacoursière-Roussel et al., 2016; Morey et al., 2020; Zhang et al., 2020). By comparison, previously studies vary widely in the number and type of sample replicates collected per site. Common study designs include anywhere from one sample per site (Lacoursière-Roussel et al., 2016), to splitting large single volume samples into multiple subsample replicates (Kelly et al., 2014; M. Miya et al., 2015), to collection of many independent samples from the same site (Andruszkiewicz et al., 2017; Gold et al., 2021; Morey et al., 2020). Our design for this study included samples collected in a 1L volume with five replicates per aquarium display, to permit evaluation of sample power and reproducibility.

The results of our rarefaction analysis revealed that our sample replication was typically more than powerful enough to resolve all of the vertebrate taxa in a display (Figure 2.6). Across the majority of tanks in all three primer sets, full power occurred often as soon as the first sampling replicate was taken, if not by the third. Additionally, read rarefaction analysis also indicates that sequencing depth was sufficient to detect the species that would amplify in this study (Appendix Figure B3). It should of course be noted that aquarium display tanks likely represent habitat of very high eDNA concentrations, and that detection might be appreciably lower in larger open systems. Nonetheless, we would suggest that sample replication in metabarcoding studies should continue, but that our sample replication design was generally sufficient to recover what dominant taxa were possible with our particular assays.

2.4.4 eDNA and relative abundance: can it be equated?

Although documenting species occupancy if very useful, many ecological and natural resource management questions revolve around relative species abundances or biomass (Lacoursière-Roussel et al., 2016). For this reason, we wanted to investigate if the metabarcoding assays used in this study can capture relative taxa abundance in our sampled mesocosms. Many prior studies of eDNA have provided evidence of some capacity to discern locations of more or less abundance for specific taxa (Bylemans et al., 2018; Morey et al., 2020; Stoeckle et al., 2017). However, that abundance relationship is predicted to be somewhat more straightforward to resolve than discerning the relative biomass of different taxa within a given habitat because different species are expected to vary widely in the amounts of eDNA they shed into a given environment due to their different habits, body constitution, and number of mitochondria per cell. In the current study we did not find statistically significant rank correlations of estimated biomass and eDNA sequence reads on a tank-by-tank basis, but that statistical limitation was almost certainly a result of the small number of taxa per tank (1 to 9) limiting statistical power. More revealing was that the majority of tank-by-assay rank correlations (7 out of 12) were very clearly positive, often strikingly so, and this pattern became notably consistent in all cases with more than 3 taxa per tank. Given the very coarse nature of our biomass estimates (based on literature values), this is even more noteworthy. So while it may be premature to use metabarcoding sequencing reads to provide precise estimates of species abundance or biomass, our finding suggest that metabarcoding may be sufficient to ascertain coarse patterns of local relative biomass representation in a species pool. This is supported by other mesocosm and field studies that have found approximate relationships between biomass and sequencing across some subset of species in metabarcoding studies (Morey et al., 2020; Stoeckle et al., 2017).

2.4.5 Concerning/addressing contamination or detection errors

Contamination and non-target amplification are some of the downfalls and concerns of using eDNA metabarcoding for assessing biodiversity. The taxonomically permissive primer sets of eDNA metabarcoding can result in overwhelming amplification of abundance non-target sequences, swamping the sequences of the taxonomic group being targeted (Collins et al., 2019; Zhang et al., 2020). In past studies, non-target amplification has been cited as relatively low, but as high as 95% due to a variety of possible lab, extraction, and sequencing contamination or errors (Morey et al., 2020). The results in this study found that all three markers identify a significant proportion of non-vertebrate organisms, even when coupled with rigorous laboratory preparation and sequencing methods. We found that overall the MiFish primer set had the least amount of off-target identification (15.57% of representative sequences), despite picking up the most in-lab contaminants out of all three assays. The low rate of off-target amplification likely relates to the higher design specificity of the marker set for marine fish taxa. The greater level of in-lab contaminant sequences may reflect that this was the first eDNA metabarcoding assay to be processed in this lab facility and the assay processing included an extra handling step to reduce microbial sequences that are known to be co-amplified. Both of these circumstances may have resulted in an increased opportunity for aerial DNA contamination. Indeed, all of the fish species detected in the lab blanks corresponded to other species recently processed in this lab facility.

The MarVer assay was shown to have a higher rate of off-target amplification than the MiFish primer set, however some of these non-vertebrate amplifications could be seen as a benefit in some cases in that they often represented invertebrate taxa present in the aquaria studied. Interestingly, the original publication for the MarVer primer set highlights the value of the assay for resolving vertebrate biodiversity without hinderance from invertebrate sequences (Valsecchi et al., 2020). The fact that the assay did amplify some invertebrate sequences, despite more stringent amplification conditions, should be evaluated in field eDNA samples for the Gulf of Maine to ensure that this not purely a mesocosm effect.

The poorest performing assay for non-target amplification was the Riaz primer set, where off-target sequences comprised 79.73% of total reads. Given the aquarium study system it is expected that some off-target amplification from bacteria and human DNA will occur. In this case human DNA constituted upwards of 54% of the off-target reads making human sequences a significant competitor for other vertebrate target sequences. In retrospect this is not surprising given this assay was designed to target sarcopterygian vertebrates, which includes humans. It has been suggested in other studies that a human-blocking primer be added to increase specificity of the marker (Zhang et al., 2020).

Inevitably, most eDNA studies face some level of contamination despite stringent laboratory protocols and bioinformatic filtering to control contaminants (M. Miya et al., 2015; Stoeckle et al., 2020; Zhang et al., 2020). This contamination risk can be reduced to a degree through very strict laboratory and gear decontamination procedures, but often not completely removed. This potential for contamination in turn emphasizes the key importance of including negative controls throughout the eDNA processing stream from sampling through sequencing.

2.5 Conclusions

The Gulf of Maine is one of the fastest warming coastal marine systems in the world (Gulf of Maine Research Institute, 2021) and has experienced a long history of fisheries collapses (Pershing et al., 2016), threatened species determinations, emergence of harmful algal blooms (Anderson et al., 2005), and species range shifts due to invasions (Harris & Tyrrell, 2001). Relatively new industries like shellfish and macroalgae aquaculture, and wind power development present new competing uses. All of these challenges and competing demands put a high priority on effective species and biodiversity monitoring to understand ecological outcomes in the region. The Maine-eDNA EPSCoR program seeks to develop key eDNA capacity to meet this need. Although the program has extensive experience with taxon-specific eDNA approaches like qPCR quantification, eDNA metabarcoding is a newer approach for the region that

required initial development and refinement, especially for vertebrate taxa. Here, I evaluated the capacity of three existing eDNA metabarcoding assays to resolve marine vertebrate communities within the Gulf of Maine. From this work I would suggest that a combination of at least two primer sets would be the most useful when trying to identifying a broad group of marine vertebrates in the Gulf of Maine, specifically the newer MarVer primer set (Valsecchi et al., 2020) and the more established MiFish set (M. Miya et al., 2015). I also found evidence to support that metabarcoding with these primer sets may be coarsely indicative of interspecific biomass relationships in many contexts. That said, there is still ample room to further refine vertebrate metabarcoding approaches for the GoM region, including further development of regional reference databases and refining sampling methods to better capture species assemblages. Finally, it should be recognized that while the use of aquarium systems offers a high degree of control for evaluating relative eDNA assay performance, such systems are not fully representative of natural habitats and species assemblages. As such, the next critical step in optimizing metabarcoding assays for the GoM region should involve analysis of field samples to obtain field-specific performance parameters than can be used in informing robust survey designs.

CHAPTER 3

EVALUATING ALTERNATE AND COMBINED ENVIRONMENTAL DNA METABARCODING PRIMER SETS FOR GULF OF MAINE INVERTEBRATES

3.1 Introduction

Assessing current biodiversity baselines of our aquatic systems is essential to understanding species diversity, economic values of biodiversity, and how anthropogenic pressures will influence changes in community structure and, consequently, whole ecosystem services (Pearman et al., 2016). Unfortunately, most current approaches to gathering biodiversity data depend heavily on labor intensive, specialized, and sometime destructive, survey tools that are generally difficult to deploy on large scales and also very dependent on a limited pool of taxonomic expertise (Meyer et al., 2021; Watts et al., 2019). Therefore, there is great need for tools to better gather ecosystem biodiversity data at large scales and with less reliance on specialized field infrastructure and taxonomic expertise. New molecular approaches, such as environmental DNA metabarcoding, offer promising potential to achieve these greater scales of data gathering, but depend on careful initial development and evaluation before being widely deployed.

One of the main groups of organisms often used to indicate ecosystem health and anthropogenic effects have been macroinvertebrates. Not only are invertebrates critically important for ecosystem nutrient cycling, microbial community structure, and serve as a key link in the food web (Gielings et al., 2021), they are particularly sensitive to stressors, and therefore key biological indicators of ecosystem change (Elbrecht & Leese, 2017; Keeley et al., 2018). Invertebrates are often seen as bioindicators, or indicators of environmental change that may act as a warning indicator to severe environmental change (Gerlach et al., 2013; Yu et al., 2012). Additionally, invertebrate groups are among the most impactful to humans, providing benefits to not only our ecology but our economy through harvest and tourism (Pearman et al., 2016; Wangensteen et al., 2018).

The current monitoring techniques for invertebrate diversity face similar challenges to those of vertebrates, but often on much larger scales due to the orders of magnitude greater diversity of invertebrate taxa. In addition to the vast array of methods often required to sample invertebrates from markedly different habitats, processing associated samples is especially time consuming and constrained by the very limited number of professionals with sufficient expertise to identify specimens based on morphology (Leray et al., 2013; Watts et al., 2019). Arguably more so than vertebrate taxa, invertebrate taxa are difficult to characterize due to the presence of many cryptic species complexes that are recalcitrant to rapid morphological differentiation (Keeley et al., 2018; Wangensteen et al., 2018). For example, many studies of benthic macroinvertebrate communities avoid species determinations to species and genus levels, often resolving to lower levels of families or orders due to the extreme labor intensity or required to morphologically differentiate taxa for which keys are often limiting and dichotomous determinations are not possible from features visible to the naked eye (Elbrecht & Leese, 2017; Wangensteen et al., 2018). On top of this, taxonomic expertise is declining worldwide, and skills in light microscopy are no longer common in biological training (Gielings et al., 2021; Keeley et al., 2018). These limitations not only affect the quality of invertebrate biodiversity data that is gathered but impede scaling up such studies to the scope needed for assessing larger ecosystems.

Advances in molecular DNA based approaches have been shown to provide alternatives to classical methods of evaluation of biodiversity in invertebrate communities. Particularly, environmental DNA (eDNA) approaches can be applied to characterize invertebrate communities using relatively simple water or sediment sampling methods (Klymus et al., 2017; Watts et al., 2019). By recovering DNA shed by invertebrates into their environment, eDNA constitutes a potentially powerful tool at both broad and narrow taxonomic levels (Thomsen & Willerslev, 2015). Such approaches not only have the potential to document species occupancy, they can in some cases provide insights into relative abundances or biomass (Keeley et al., 2018; Lejzerowicz et al., 2015).The predominant method for monitoring biodiversity of broad taxonomic groups using eDNA is referred to as eDNA metabarcoding, a method that involves

simultaneous identification of multiple taxa from environmental or biological samples with the use of universal PCR primers and next-generation sequencing (Taberlet et al., 2012). Not only can eDNA metabarcoding provide multi-species identification for common taxa, it can often reveal taxa that are unexpected, transient, cryptic, or in low abundance in an area of sampling (Fonseca et al., 2010; Lejzerowicz et al., 2015). In comparison to traditional organismal survey and identification methods, eDNA is typically adaptable to more species and habitats, less invasive and destructive, more cost effective for sample collection, and has a more rapid turnaround for sample processing (Dejean et al., 2012; Taberlet et al., 2012). Environmental DNA metabarcoding in particular has become an efficient and essential method for the assessment of taxonomic assemblages (Fonseca et al., 2010; Lejzerowicz et al., 2015; Zhang et al., 2020).

The inclusion of invertebrates in biodiversity monitoring using eDNA metabarcoding has lagged significantly behind other taxa (Fonseca et al., 2010; Gerlach et al., 2013; Klymus et al., 2017; Watts et al., 2019) such as vertebrates. However, molecular resources do exist. Currently, the 18S nuclear and COI mitochondrial gene regions are the most broadly used targets for invertebrate environmental DNA due to their sequence variability providing suitable resolution for taxonomic identification and beta diversity estimation (Wangensteen et al., 2018). These gene regions are also better represented in reference databases than many others (Lejzerowicz et al., 2015; Leray et al., 2013), although this coverage is still often spotty and biased with respect to different taxonomic groups and their associated habitats. For example, published metabarcoding assays and reference databases for invertebrate detection were often designed for monitoring of terrestrial invertebrates, rather than aquatic taxa (Watts et al., 2019; Yu et al., 2012). This is at odds with the fact that alterations to aquatic invertebrate assemblages have often been monitored for the express purpose of assessing aquatic habitat impairment (Klemm et al., 2003).

Development of invertebrate eDNA tools also faces a need to optimize the methods for water-based detection. Many of the applications of invertebrate metabarcoding in the literature focus on either discerning representation in bulk samples of whole or partial organisms (Keeley et al., 2018; Lejzerowicz

et al., 2015; Pearman et al., 2016; Wangensteen et al., 2018), concentrated out of larger soil, sediment, or water volumes, or identifying gut contents from predator or herbivore species (Deagle et al., 2009; Leray et al., 2013). These relatively concentrated DNA sources present very different detection conditions than for characterization of cellular or subcellular invertebrate eDNA circulating in the aquatic environment. Finally, most studies focusing on marine or freshwater environments have used 454 pyrosequencing (Comeau et al., 2011; Fonseca et al., 2010; Porazinska et al., 2009), a now slightly outdated sequencing technology compared to the Illumina sequencing now used in most eDNA metabarcoding studies. Improved information on the efficiency and depth of invertebrate marker sets and their capacity to identify marine taxonomic groups is sorely needed.

The main objective of this study was to evaluate alternative primer pairs from the literature that could potentially be applied to identifying invertebrate diversity and abundance in the Gulf of Maine region. This region is currently the focus of a large NSF EPSCoR Track 1 research program on environmental DNA in coastal systems that seeks to develop eDNA-based ecological inference around pressing challenges and opportunities such as climate change, sustainable fisheries, emerging aquaculture industries, harmful algal blooms and species range shifts. Meeting the goals of this research program entails identifying, developing and optimizing eDNA tools to understand diverse coastal habitat assemblages, including the marine invertebrates that serve key roles in the function of these ecosystems. This study makes use of a suite of censused displays of the New England Aquarium (NEAQ) to quantitatively evaluate two common invertebrate metabarcoding primer sets. The questions to be answered in the study are as follows:

a) To what degree do these alternate primer sets capture the taxonomic representation of distinct GOM species assemblages of invertebrates, alone or in combination?
b) To what degree do the different primer sets capture approximate relative abundance or dominance of the key GOM taxa.

3.2 Methods

3.2.1 Literature review and assay selection

In order to determine the suitability of metabarcoding as a method to study invertebrate diversity in the GOM/ NEAQ, the literature was consulted to identify previously published primer sets for invertebrates. Due to the high usage of 18S and COI in sequencing invertebrate taxa found in the literature, these two genes were identified as regions with the best currently available metabarcoding resources. The final primer assays that were selected for use in this study include the 18S primer set (V4 region) E572F and E1009R, from Comeau et al. (2011), hereafter referred to as "18S", and the COI primer set BF1 and BR1 from Elbrecht et al. (2017), hereafter referred to as "COI". Due to the lack of broad range comparative studies for invertebrate metabarcoding, and none based on a multi-species mesocosm design, primer sets were selected based on their reported ability in amplification, specificity for the taxa reported in the NEAQ, and ability to be implemented without special modifications. The 18S set has previously shown high detection ability in eukaryotes (Comeau et al. 2011), while the COI set from Elbrecht has been widely applied to other marine invertebrate diversity studies (Meyer et al. 2020).

Initially, this study proposed to include another widely used COI marker developed by Leray et al. (2013). However, after some initial assessment, it was not included in the subsequent amplification of aquarium samples due to time and resource limitation.

3.2.2 eDNA water sampling

Water samples were collected from 6 GOM-specific exhibits at the NEAQ. These include the Stellwagen Boulder reef community (BOUL, 9 taxa), the Eastport enclosure (E, 18 taxa), the Tidepool Touch Tank (TP, 24 taxa), the Isle of Shoals exhibit (IS, 4 taxa), the Stellwagen Sandy Reef community (SS, 3 taxa) and the Boston Harbor Islands Shorebirds enclosure (SB, 6 taxa). In total, there were 38

unique invertebrate taxa across the 7 GOM specific tanks on the day of sampling. To see a complete list of taxa and their populations in each exhibit on the day of sampling, refer to Appendix Table A1. Tanks that were sampled at the NEAQ all draw their source water from Boston Harbor, which is drawn in and filtered through charcoal and UV light before circulating to tanks. Although water is not directly exchanged between the tanks sampled, aquarium staff maintain these tanks as a group and they are all accessed from the same corridor. Due to this knowledge, we expected that water collected from each display would mostly amplify invertebrates known to be within each respective tank, with additional minor amplification of species from Boston harbor or other tanks in the aquarium.

In preparation for the collection of samples, 1L Nalgene bottles were sterilized using a 10% bleach solution for at least 10 minutes, and then rinsed thoroughly with tap water followed by another rinse in DI water. Bottles were then UV sterilized for 1 hour and then sealed, labeled, and placed into sterile bags according to display name. Prior to being bagged, one of the sterilized bottles for each tank was filled with 1 L of deionized water to act as a "cooler blank" to control for contamination in the field. During the sampling of each tank, bottles were unsealed one at a time and dipped into the back of the display or the sump of the tank, or in the case of the field blank were opened and exposed to the air for a full 30 seconds. For each display, 5 replicates were taken, totaling to 5L of sample water and 1L of control water. After collection, samples were transferred back into coolers and packed with ice. Samples were then transported to the University of Maine where they were filtered within 48 hours of collection.

3.2.3 eDNA preparation

Filtering of samples took place at the University of Maine in eDNA specific laboratories that were sterilized with 10% bleach solution and UV light sterilization beforehand. Equipment used in the filtration process was also bleached, rinsed and UV sterilized before use. For additional laboratory contamination control, 1L deionized water "lab blanks" were filtered prior to the start of each filtering

session. Samples and blanks were filtered using filtration cups and a vacuum pump, and were filtered through 47mm diameter Whatman glass fiber filters (0.7uM, Whatman). Each tank replicate was filtered on its own filter, and the entire 1 L sample was filtered through. Filters from samples were then stored in 1.5mL tubes at -20C until extraction.

Extraction of eDNA samples occurred 10 days after filtration, using the DNeasy Blood and Tissue kit following manufacturer's protocols (Qiagen, Germany). During extraction, one sample was lost due to an unexpected interruption (HS03). Samples were eluted to 100uL and then stored at -20C until preparation for sequencing.

3.2.4 PCR amplification and sequencing

Library preparation for next-generation sequencing of samples took place in University of Maine laboratories. Initial amplification of samples to attach primers and adapters was performed following the respective publication PCR protocols for each set, with occasional lab modifications to increase specificity and binding, at a volume of 20uL (Table A2). For all PCR amplification, the same high-fidelity KAPA HiFi HotStart ReadyMix (Roche Molecular Systems, Inc.) was used. Successful amplification was visualized for all samples via gel electrophoresis.

For cleanup of unwanted PCR products, primer dimers and excess primers, the Zymo Select-a-Size DNA Clean and Concentrator MagBead Kit was used following the manufacturer's protocol (for bead concentrations used in each cleanup, see Table A2). The secondary PCR amplification to attach Illumina indices and adapters used 12.5uL KAPA HiFi HotStart ReadyMix, 9uL nuclease free water, 1.25uL of Nextera Indexing primers (Illumina, Inc.), and 1uL template DNA. The PCR amplification for indexing primers included an initial denaturation at 95°C for 10 minutes, followed by 8 cycles of 95°C for 10 seconds, 55°C for 30 seconds, 72°C for 30 seconds with a final elongation at 72°C for 5 minutes. After successful indexing, which was visualized via gel electrophoresis for confirmation of attachment and

positive amplification, samples were again purified using the same size-selection bead kit and protocol (see Table A2).

Following cleanup, samples were quantified using a Qubit 4 Fluorometer (ThermoFisher Scientific) using the Qubit dsDNA HS Assay kit and were then pooled to equimolar concentrations in a final library. This final library was then quantified using the same QuBit and kit, as well as with qPCR quantification using the KAPA Library Quantification Kit (Roche Sequencing, Inc.). Amplicon sizes were verified using a high-sensitivity dsDNA assay chip run on a 2100 Bioanalyzer. Each library was then frozen and shipped overnight to the University of Rhode Island's Genomics and Sequencing Center where they were sequenced on an Illumina MiSeq platform using a MiSeq Reagent Kit v3 (600 cycle) to accommodate long sequencing reads. Sequencing was performed on the MiSeq platform due to its ability in providing long paired-end reads, given the amplicon lengths for the 18S and COI assays are approximately 509 bp and 217 bp respectively. For each sequencing run, libraries were run with PhiX standard at 10-20% spike in to improve low diversity.

3.2.5 Bioinformatic Pipelines

Paired-end sequencing results were demultiplexed on the Illumina MiSeq, and then initial data quality and filtration occurred in a QIIME2 data pipeline (Script S1). Low quality base pairs were trimmed using a threshold of 20 on the Phred score, and then sequences were aligned, merged and sorted into representative sequences. Chimeric sequences were removed, and then taxonomic assignments were made sample-by-sample using either a classifier for the 18S sequences, or a consensus-blast approach for the COI data, both occurring in a QIIME pipeline.

The 18S assay was run against a curated Silva database created from the SILVA138 classifier used in this study. The COI assay was run against a Barcode of Life (BOLD) database created in combination from publications by Robeson et al. (2020) and O'Rourke et al. (2020) using a consensus-

blast method in QIIME2. Taxonomic assignment was used via BLAST method rather than vsearch due to its ability to locally align to a portion of a sequence that overall may be very divergent. Thresholds were set to 90% match. After this initial assignment, it was determined that a more specific database might be helpful in gaining more resolution, and thus a curated reference library using the known species list from the aquarium was created using the program BC Databaser (Keller et al., 2020). With this process a total of 36 of the 38 census species in the aquarium had represented sequences in the newly formed database. This reference database was then used in the same way as the former, using a consensus-blast for assignment using the same parameters.

Final taxonomic assignments for all unique sequences for both assays was determined first using the reference database assignments, then against both the NCBI nt database and the BOLD database, and finally against comparison to known species lists from the aquarium. Taxonomic assignment from the bioinformatics pipeline versus final taxonomic assignment can be found in Appendix Tables C1 and C2. Once all ASVs were assigned, they were compressed into OTUs for further analysis in R.

For the OTUs that had sequences in the blanks, the maximum number of reads from any of the samples was then subtracted from the reads in all associated environmental samples with that OTU (Table C3). Following appropriate blank amplification subtractions, any remaining non-target taxa (i.e. those not censused within the aquarium) were removed from the datasets for ancillary analysis of potential off-target and contaminant eDNA sources (Appendix Table C4).

3.2.6 Data Analysis

To assess the possible efficiency of the selected markers with respect to the reference databases and bioinformatic pipelines being used within the study, coverage within reference databases for each of the study organisms was checked against the known taxon list.

Efficiency of the primers selected for the study were evaluated based on the sequencing results and taxonomic resolution that was capable for each given set after sequence processing. In addition, sequence quality of the runs and overall reference database availability for those species on the known taxon list based on the in-silico evaluations were considered. In evaluating the relationship between total population within a tank and eDNA sequence per organism, the eDNA sequence count was first averaged across all tank replicates for each individual organism. Then, the log of both eDNA average sequence count and the total population within the tank was taken. Spearman's rank correlations were then estimated to understand the relationship between total population and eDNA. Relationships between total population and average eDNA count were visualized using bar charts that took the log transformed values of each of the values in order to scale them appropriately.

Rarefaction power analysis in the study was performed at the sampling level, and therefore effects of environmental sample replication in recovering taxa within the aquarium was assessed by sampling all sequences related to organisms known to be within a study within each of the given tanks using a bootstrap method. Tanks and replicates at each level were subsampled 1000 times to give sufficient depth to rarefaction. We in turn enumerated the proportion of randomizations in which the full set of genetically detectable taxa were detected. Additional traditional rarefaction on sequencing read depth and replication in revealing species number were also completed in the same manner (see Appendix C). Finally, sample off-target amplification was visualized using pie-charts breaking down the spread of contaminants to the class level within each tank. All analyses were performed in R version 4.1.1 (R Core Team, 2021).

3.3 Results

3.3.1 Sequencing results

A total of 28,624,052 and 34,521,516 reads were generated from the Illumina MiSeq runs of the COI and 18S processed samples (Table 3.1). Following quality filtration of reads, a total of 25,897 and 5,384,143 reads remained between all samples and blanks for the 18S and COI markers (Table 3.1). Although no NTCs were amplified during PCR runs, some blanks from lab or field sources amplified (n=8), were sequenced, and accounted for 3.86% of the reads for the 18S data.

Table 3.1 Sequence data from the bioinformatics steps in this study. Non-relevant taxa refers to the taxa that were from outside sources from the aquarium, non-invertebrate sequences, or known contaminants.

Assay	**Total Reads (MiSeq)**	**Pair-merged sequences**	**Sequences post quality filtration**	**Sequences from Blanks**	**Total ASVs**	**Non-relevant taxa**	**Removal of non-relevant taxa**
18S	34,521,516	25,897	20,414	788	277	20,195	219
COI_B1	28,624,052	5,348,143	2,234,098	-	41,724	2,220,073	14,025

3.3.2 In-silico analysis

When searching for reference taxonomy in the SILVA SSRU r138.1 database, a total of 27 of the 38 species in this study has sequence coverage for the 18S gene, producing a maximum total possible resolution of 27 species, 23 genera and 9 classes of the known taxa list. Within the BOLD database used in part for the COI sequence alignment, all of the species in the study had representative sequences. After creation of the constrained reference database from BC Databaser, a total of 36 of the possible 38 species were represented in that database, covering 32 genera, and 9 classes. The two species excluded in this database were *Diodora aspera* and *Neptunea lyrata*. These two particular species were also not represented within the 18S database used, and had only 4 and 1 representative sequence/s within the BOLD database, respectively. Additionally, in silico primer binding testing found that neither the 18S or

the COI primer sets had the ability to bind to reference sequences from SILVA, BOLD, or NCBI with 3 or less base pair mismatches.

3.3.3 Taxonomic resolution

In total, the 18S assay had 3 correct identifications down to the class level out of 9 possible classes present in the aquarium. It did not have any finer level taxonomic classifications out of the sequencing data. Due to poor sequence quality, the 18S assay was dropped from any further analysis.

The COI assay detected only about a third of the censused invertebrate community of the NEAQ, but provided relatively fine-scale taxonomic resolution for the censused species OTUs that were detected, with the majority of target-taxa identifications happening at the species level. Taxonomic calls were somewhat different using the general BOLD database (O'Rourke et al., 2020) compared to the curated reference database produced from BC Databaser based on the known species list. The former reference database had 11 identifications to species (*Asterias forbesi*, *Asterias rubens*, *Buccinum undatum*, *Crossaster papposus*, *Cucumaria frondosa*, *Homarus americanus*, *Euspira heros*, *Mytilus edulis*, *Ophiopholis aculeata*, *Pagurus acadianus*, and *Strongylocentrotus droebachiensis*), 3 genus level identifications (*Metridium, Solaster, Henricia*) and 1 order level classification. Use of the constrained reference database added 2 more species level identifications (*Solaster endeca* and *Metridium senile*). Combined, the use of the two reference databases against the COI data resolved 13 of the 38 possible known species within the tanks, covering 13 of the 34 possible genera, and 6 of the 9 possible classes (Figure 3.1, Appendix Table C5). Across possible classes, the most represented in the sequencing results were those within the class Stelleroidea, with 6 of the 7 possible species within the class being detected.

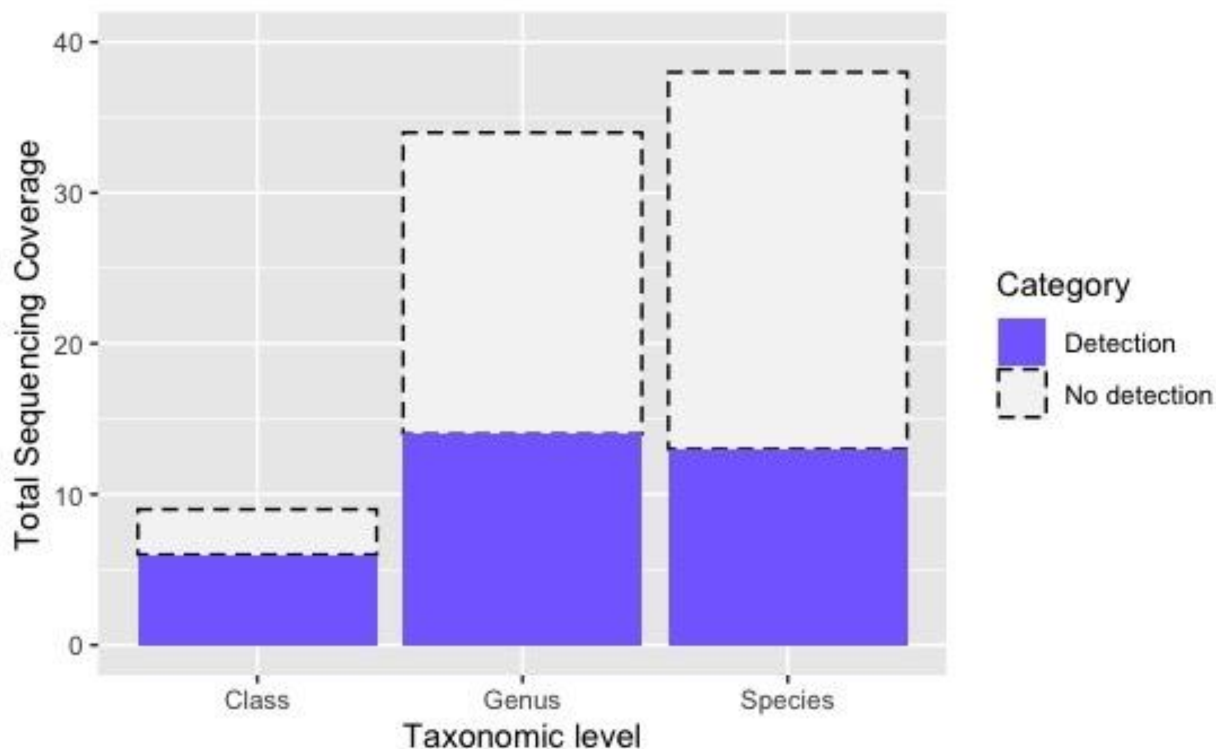

Figure 3.1 Detections plot for the class (n=9), genus (n=34) and species (n=38) level using the COI marker

3.3.4 Total population versus eDNA sequence count rank correlations

The relationship between total populations within tanks and their corresponding eDNA sequence count averages were evaluated using rank plots (Figure 3.2) and spearman's rank correlations (Table 3.2). For the tanks Shorebirds/Boston Harbor Islands and Isle of Shoals, only a single species was detected so intra-tank correlations could not be computed. For all tanks sampled where a relationship could be estimated, the correlation between tank population of an organism and their average eDNA count range from -0.2 to 1 but were not statistically significant (Table 3.2, Figure 3.2). When performing a meta-analysis using Fisher's method across all tanks to test for significance, the combined p-value was again not statistically significant (χ^2=3.88, df=8, p=0.868).

Table 3.2 Spearman's rank correlation coefficients, p-values for the COI primer set.

Tank	Rho	p-value	S
Boulder	0.257	0.658	26
Eastport	0.2	0.917	8
Sandy Shoals	1	0.333	0
Touch Tank	-0.2	0.714	42

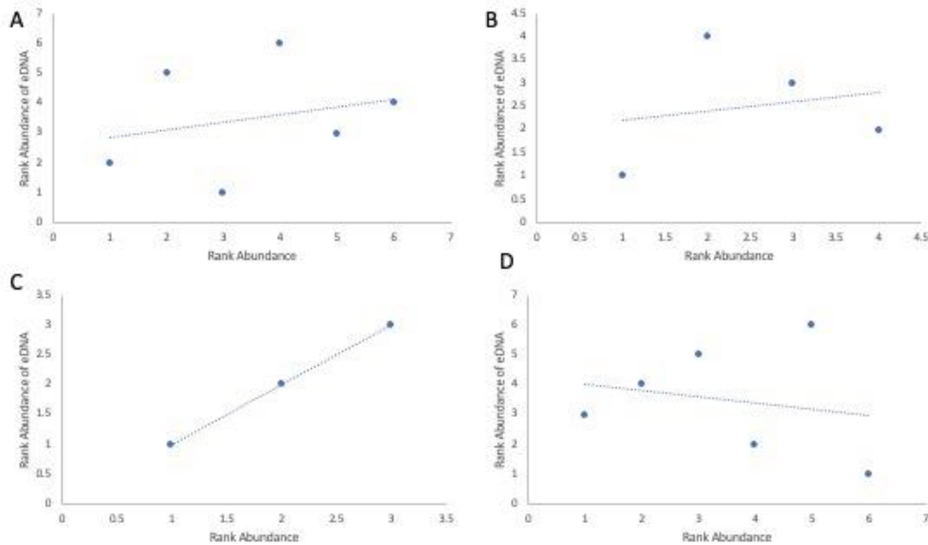

Figure 3.2. Relationship between total population counts and average eDNA sequence counts for all species detected by the COI marker in the study; a.) Stellwagen Boulder Reef b.) Eastport, ME c.) Stellwagen Sandy Banks d.) Touch tank

3.3.5 Effects of sample replication

Rarefaction power analyses on this primer set revealed that our sampling efforts were mostly sufficient to resolve all of the genetically-resolvable species present within a given tank with five or fewer samples. The probability of detecting all of the censused species within a given tank repeatedly showed that detection power was never fully reached after 5 samples, but was close to saturating at the fifth

sample (Figure 3.3). The exception to this is the Isle of Shoals tank, which reached total probability of detecting all species after two sampling replicates.

Rarefaction analysis on sequence count and species saturation revealed that all replicates within individual tanks reach saturation after much fewer sequences than the overall sequencing depth for each tank (Appendix Figure C1). Rarefaction of sampling replication versus number of species within each tank did not reveal saturation of species after 5 sampling replicates, and had some significant variability (Appendix Figure C2).

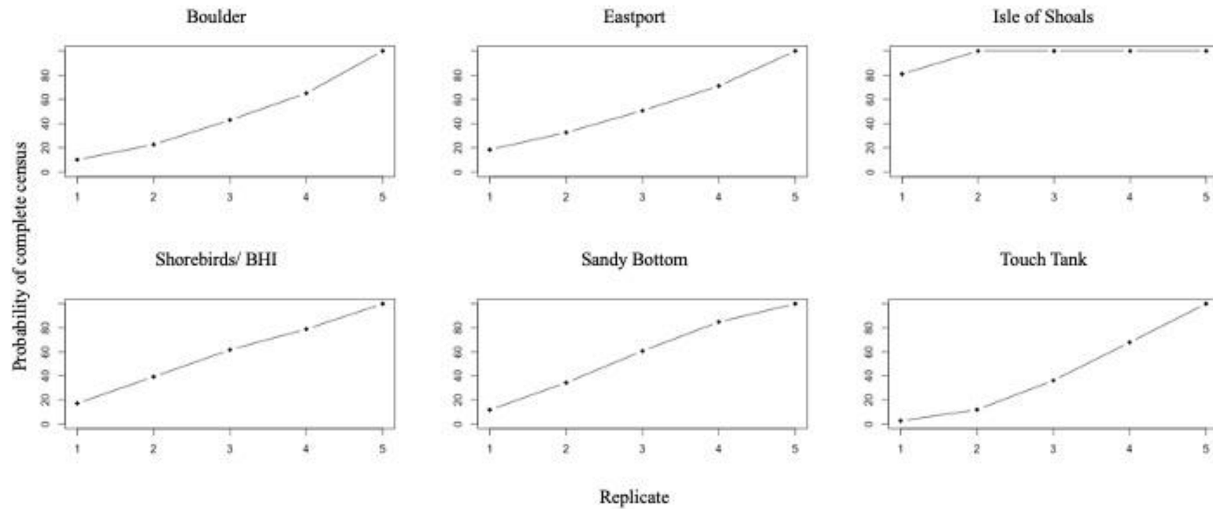

Figure 3.3. Rarefaction power curves for each tank based on probability of detecting all taxa in the tank; a.) Boulder Reef b.) Eastport, ME c.) Isle of Shoals d.) Stellwagen Sandy Banks e.) Touch tank

3.3.6 Off- target identification

For the COI assay, 99.4% of the sequences were assigned as non-target taxa. Within this off-target set, 91.3% were taxonomically unassigned sequences, which when randomly blasted in NCBI typically aligned to bacterial amplification. The other portions of the non-target taxa can largely be attributed to 1) eDNA of cryptic taxa likely present in displays but not accounted, 2) eDNA from taxa

employed in feeds used to support the aquarium communities, 3) eDNA transported from taxa in other aquarium displays, and 4) eDNA present in water sourced from Boston Harbor. Non-census invertebrate sequences included representative of the classes Amphipoda, Anthozoa, Brachiopoda, Cephalopoda, Demospongiae, Gastropoda, Hoplonemertea, Homoscieromorpha, Hydrozoa, Isopoda, Malacostraca, Ophiuroidea, Palaeonemertea, Pantopoda, Pilidophora, Polychaeta, Pycnogodia, Scyphozoa and Staurozoa (Figure 3.4). The greatest representation was from common isopods, amphipods, shrimp, snails, sponges, sea worms and jellyfish with ranges including Boston Harbor or that are employed in aquarium feeds.

There was also evidence of substantial eDNA transport among displays. Most prominently was contamination of the common plumose anemone (*Metridium senile*) in multiple tanks that do not have known presence of the species, but border the Isle of Shoals tank, Stellwagen Sandy Bank tank, and the Touch tank that do house this species. Similarly there was evidence of eDNA transport among adjoining displays for the green sea urchin (*S. droebachiensis)* that was housed only in the Stellwagen Boulder Reef tanks.

Non-marine invertebrate eDNA represented only small portions of the off-target amplification pool, and included sequences from Arachnida and Insecta (Figure 3.4). Interestingly, the COI primer set also successfully identified two of the known vertebrates in the tanks sampled, harbor seal (*Phoca vitulina*) and winter flounder (*Pseudopleuronectes americanus*) (Figure 3.4).

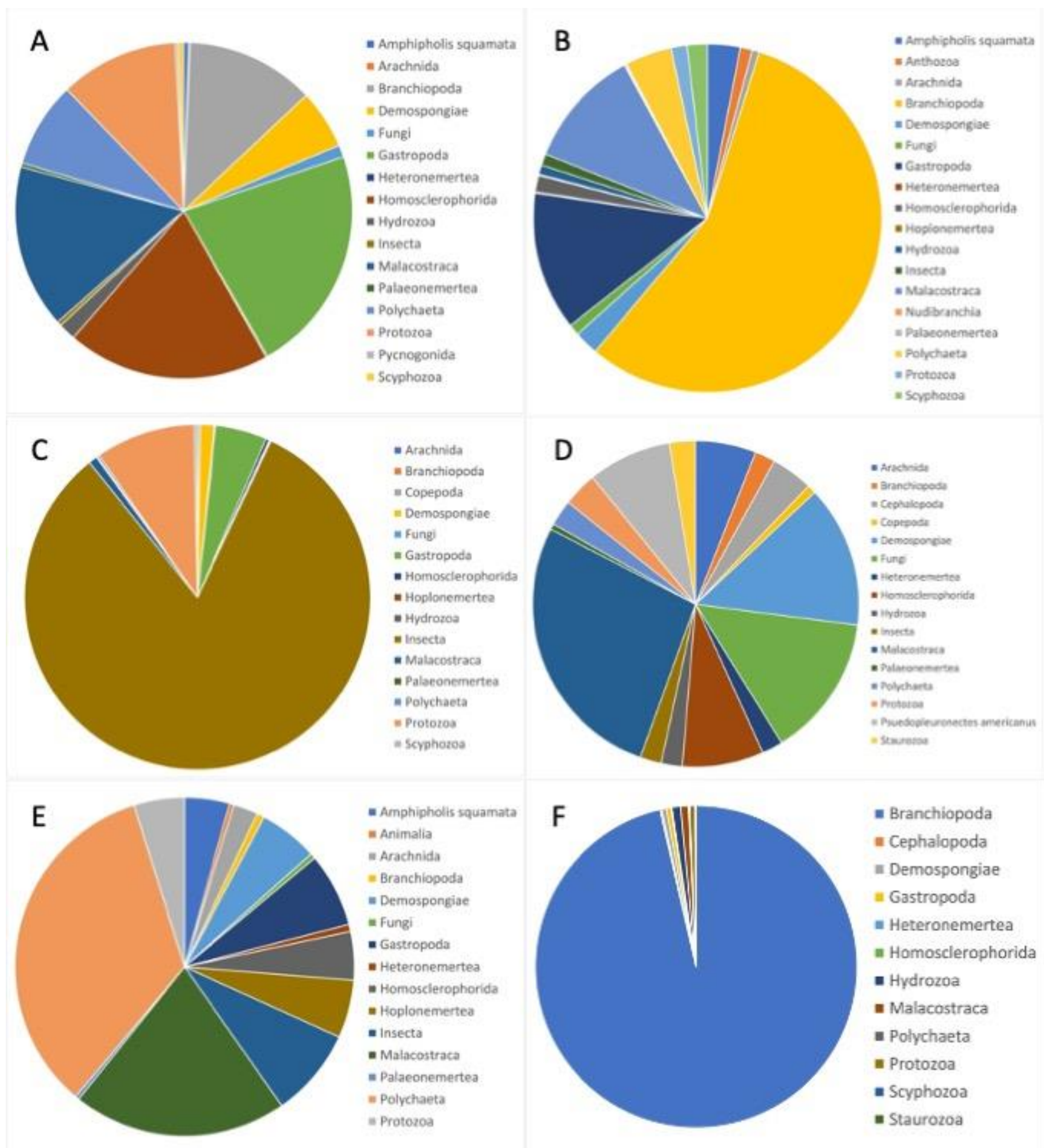

Figure 3.4. Proportions of off-target OTUs by taxa; a.) Boulder Reef b.) Eastport, ME c.) Isle of Shoals d.) Stellwagen Sandy Banks e.) Shorebirds/Boston Harbor Islands f.) Touch tank

3.4 Discussion

In this chapter, I surveyed 6 different tanks from the New England Aquarium initially using two different metabarcoding primer sets with the intent of comparison between the two for assessing marine

invertebrate community assemblages. After failure to produce sufficient sequencing results with the 18S set, remaining analyses focused on performance of the COI primers in assessing the community assemblages within the NEAQ tanks sampled. The COI primer set detected approximately one third of the census species in the aquarium displays, but also a much larger set of taxa representing six invertebrate classes with eDNA deriving from diverse sources within and beyond the aquarium. Most of the census species that were identified were resolved to a species level, but metabarcoding sequence copy number for these species was not clearly related to their known abundances. I now consider these findings in the context of how these assays might function to survey invertebrate diversity of Gulf of Maine habitats.

3.4.1 In silico evaluation of markers

The original intent of this chapter was to compare the performance of metabarcoding primers for species resolution against a set of known assemblages in mesocosms. After data processing, it was revealed that the 18S marker used in this study (Comeau et al., 2011) had insufficient sequencing results to move forward with analysis. Although it had some taxonomic assignment to the class level, overall sequencing depth was not sufficient for further empirical analyses. However, in silico analysis suggests there are other possible reasons that this assay might not be optimal for marine invertebrates. Other studies have found that 18S markers may have unsuitable reference databases for some taxa despite the fact that the gene covers broad taxonomic depth (Elbrecht & Leese, 2017). This was indeed supported by in-silico evaluations of available reference databases for 18S sequences, with only 71% species coverage available within the SILVA database for those species censused in the NEAQ. However, this limitation might be at least partly overcome by future improvements to SILVA or through development of more specialized regional databases.

In contrast to using ribosomal subunit markers such as 18S for metabarcoding, the use of COI in metabarcoding has been cited as providing the possibility for a broad range of highly specific taxonomic

identification due to the high interspecific variation in the gene region (Collins et al., 2019; Leray et al., 2013; Yu et al., 2012). It has been shown in previous studies that this gene region is successful in detecting many marine invertebrate taxa (Gielings et al., 2021; Steyaert et al., 2020). Not only does the COI gene region provide high variability and coverage for many organisms, it also has one of the most expansive and curated reference databases in the world (Yu et al., 2012). Nonetheless, that does not go to say that these reference databases do not still have substantial gaps, as was shown in a study using the COI marker for littoral bottom invertebrate community assessment (Wangensteen et al., 2018).

On the other hand, this gene region has been noted to be a potentially inefficient metabarcoding marker, due to some of the same reasons that make it useful: a high variability and lack of taxonomically conserved regions (Collins et al., 2019; Deagle et al., 2014). Such primer sets may hold significant bias for identification of particular taxonomic groups (Clarke et al., 2014). Even without in-vitro testing, our in-silico analyses of reference database coverage still unveiled taxonomic gaps. Specifically, two of the species found within this study (*Neptunea lyrata and Diodora aspera*) had very low sequence coverage within the respective taxonomic databases. Additionally, in-silico primer binding testing for this particular assay revealed insufficient binding of these primers to any of the species represented within the aquarium. This could be attributed to high amounts of degenerate bases and the program used to do the primer binding in this study, however it can also indicate this primer's inefficiency at binding to species of interest. Additional functional gaps were identified during actual application of the assay to known aquarium taxa, but these gaps may not all relate to the resolution power of the primer set itself.

3.4.2 COI taxonomic detection

Taxonomic identification and specificity is widely variable when it comes to invertebrate metabarcoding studies, with studies often seeking only coarse taxonomic resolution to phyla (Fonseca et al., 2010), order or family (Yu et al., 2012), with the rarity of identification to species level. This has been

attributed to multiple factors, including the aforementioned issues with variability in primer binding sites (Deagle et al., 2014). However, there are also a multitude of recent eDNA metabarcoding studies that cite the use of COI markers as efficient in the recovery of invertebrate taxa in comparison to traditional survey methods such as manual morphological identification (Elbrecht & Leese, 2017; Meyer et al., 2021). The capacity of DNA barcoding to provide species level resolution should be emphasized in environmental monitoring given the potential to overcome the need for gross generalizations concerning environmental sensitivities that are inherent when many related taxa and lumped due to taxonomic imprecision of standard methods (Bush et al., 2019).

Overall, 34% of the possible census species were captured from our sequencing, with the class Stelleroidea being predominantly amplified across all tanks (*Asterias rubens*, *A. forbesi*, *Crossaster papposus*, *Ophiopholis aculeata*, *Solaster endeca*). COI also detected a sea cucumber to species (*Cucumaria frondosa*). These outcomes suggest the potential utility of the primer set for identifying some marine echinoderms, despite that this group is only very distantly related to the freshwater arthropods for which the assay was designed (Elbrecht & Leese, 2017). There was also successful identification to species level for one of the anthozoans and for two vertebrate taxa, despite the fact that the ancestors of these lineages diverged from other animals long before the origins of arthropods. This ability to resolve species so distant from the original design group is somewhat surprising but supports the versatility of the marker in identifying a broad diversity of marine invertebrate taxa. This finding is supported by findings from two previous studies looking at marine invertebrate characterization in other regions. Specifically, both Leray et al. (2013) and Steyaert et al. (2020) cite the superiority of using COI in taxonomic identification of a broad range of marine invertebrates.

In contrast, organisms in the class Crustacea in this study had relatively poor detection for the COI primers, only recovering 2 of the possible 8 species. Phylogenetically, this seems odd given the designed capabilities of COI for identifying organisms within the phylum Arthropoda. However, previous marine oriented eDNA studies focused on arthropod taxa within the class Crustacea have noted the

relatively low recovery of eDNA from this group (Crane et al., 2021; Forsström & Vasemägi, 2016; Geerts et al., 2018). Principally, this low eDNA recovery has been determined to be linked with life stage due to possible hinderance of eDNA shedding associated with the presence of an exoskeleton (Crane et al., 2021; Geerts et al., 2018). Additionally, there was relatively little detection of organisms within the classes Bivalvia and Gastropoda, which might also be attributed to the presence of an outer acellular shell. However, we observed substantial detection of American lobster eDNA and it has been found that molluscan eDNA is recoverable in some applied field studies (Klymus et al., 2017; Steyaert et al., 2020). As such, low detection of these classes of organisms might be attributed to a larger combination of factors, including possible primer bias for the COI fragment, which has been shown to alter amplification efficiency of particular taxonomic groups (Clarke et al., 2014; Elbrecht & Leese, 2015). Nonetheless, these taxonomic gaps in our relatively controlled study suggest that investigators seeking to use COI for particular crustaceans, bivalves, or gastropods should evaluate detection under field conditions before conducting wider surveys.

Despite low overall species and genus level detection for COI in this study, it must be noted that the use of metabarcoding as a whole, and the use of COI as a genetic marker cannot be disregarded for understanding invertebrate community assemblages, biodiversity, and ecosystem monitoring. Previous literature has extensive support of the use of metabarcoding in broadening ecosystem monitoring for aquatic invertebrates (Leray et al., 2013; Steyaert et al., 2020; Wangensteen et al., 2018; Yu et al., 2012), even referring to the method as "Biomonitoring 2.0" method for ecosystem assessment (Baird & Hajibabaei, 2012). DNA barcoding has opened a breadth of tools that continue to grow in capacity, even allowing for detection of organisms previously undetectable by traditional monitoring efforts or taxonomic experts (Bush et al., 2019). Hence, it may be more profitable to focus less on the limitations of a single metabarcoding primer set and more on approaches that might overcome many of these limitations. For example, a redesign or change in the region of the COI primer set could significantly improve regional species coverage, particularly for key focal taxa, leading to fewer gaps than the current

primer set. For example, there may be benefit in evaluating the region previously designed by Folmer et al. (1994) that has since been adapted and successfully used to measure biodiversity for marine invertebrates in field settings (Steyaert et al., 2020). Additionally, the use of multiple complementary primer sets could greatly expand taxonomic recovery (Alberdi et al., 2018; Gielings et al., 2021; Meyer et al., 2021). Indeed, evaluating this potential was a goal of the current study that was not achievable due to performance of the 18S sequencing.

3.4.3 Effects of sample level replication on invertebrate detection

Sufficient sampling effort is important for reliable and repeatable detection of any survey methods. It is well known that replication sampling not only reduces the number of false negatives in surveys, but also increase detection rates of taxa across the board (Gentile F. Ficetola et al., 2015). In most studies, some amount of replication is employed to provide for adequate recovery of a target assemblage (Bush et al., 2019). Although eDNA sample replication can occur at multiple levels from water collection through sequencing (Alberdi et al., 2018; Gielings et al., 2021), sequencing of replicate field water samples provides the most inclusive assessment of the compounding suite of stochastic processes that interact to influence eDNA detection probabilities and were thus assessed in the current study.

Our results show that sampling efforts performed in this study were mostly sufficient to recover taxa present in the aquarium census. Sample rarefaction curves were created for all tanks to show the probability of detecting all census taxa using a bootstrapping method. Across the majority of the tanks sampled, power of detection for the total census appears to be close to being resolved, with most tanks reaching 100% probability of detection of census taxa after five sample replicates (Figure 3.3). The exception to this is in the Isle of Shoals tank, which showed saturation of detection of all census taxa after three samples taken. Additionally, read rarefaction analysis indicates that sequencing depth in this study

was sufficient to identify genetically detectable species (Appendix Figure C1). These results indicate that our sampling efforts are sufficient to detect the taxa that will amplify in this study, but that sampling replication should continue in future studies to capture any rare taxa.

3.4.4 Relating abundance and eDNA sequencing counts

One of the most common questions in current eDNA metabarcoding approaches it is whether the approach can be used to approximate organismal abundance in an ecosystem (Barnes & Turner, 2016). However, it is important to recognize that there are two distinct versions of this question: 1) the degree to which sample-to-sample variation in sequence reads of a given OTU represents abundance variation of that particular species, and 2) the degree that relative sequence read abundances for different taxa in a sample reflects their relative dominance in an assemblage. The latter is challenged by more factors than the former given the tremendous variation among taxa in their eDNA shedding propensities (Crane et al., 2021) but may often be of interest where community snapshots are used to infer ecosystem function, such as in the case of indices of biotic integrity (Klemm et al., 2003).

The results of our study do not support any overt relationship between the census abundances of species and their relative eDNA sequence representation from water samples. This could be due to a variety of factors, including limitations in the study design such as low eDNA recovery for the focal species compared to other sources, and the use of species abundance rather than biomass. The invertebrate species in this study exhibit an extreme diversity of body plans, sizes, cellular composition, and microhabitat use, all of which may make simple abundance a potentially poor predictor of eDNA production rates. For example, it has been previously noted that invertebrates such as crustaceans have relatively low eDNA shedding rates due to the presences of an exoskeleton (Crane et al., 2021). Use of biomass and constraining comparisons to taxonomic groups having more similar body plans and ecology could improve prospects for using eDNA to infer relative community representation. Greater evidence of

a relationship between biomass and sequence copy number for fish in my prior chapter support this contention. The nature of the COI primer set itself may be a factor in whether sequence copy numbers are informative of relatively representation in communities. Primer sets that target such a broad spectrum of species may be more subject to a degree of amplification bias due to sequence variants that distort the affinity of primer binding (Elbrecht & Leese, 2015).

It is of course possible that with larger numbers of detected taxa it might have been possible to detect some positive relationships across taxa, but our findings are sufficient to at least caution against the assumption that such relationships can be assumed. Previous studies have also noted limits on the ability of eDNA metabarcoding to reliably produce relationships between populations or biomass of a species and their sequencing reads (Bush et al., 2019; Elbrecht & Leese, 2015, 2017; Steyaert et al., 2020). Current limitations to assess relative taxonomic dominance might someday be partly overcome through more thorough understanding of amplification biases. We also want to be clear that our findings should not be construed as evidence against the possibility that sequency copy number is potentially informative of spatial or temporal variation in abundance or biomass for any given invertebrate species. That prospect requires a different study design to assess than we employed here.

3.4.5 Off-target amplification

Some off-target amplification was expected for this study, as this primer set was originally designed for freshwater insect detection rather than marine invertebrates. However, the majority of OTU assignment went to non-target or unassigned bacterial sequences. This is a common problem for eDNA metabarcoding with very permissive primer sets in general, and arises due to the sheer dominance of microbial DNA in most systems. This microbial component is difficult to fully eliminate, and has the potential to outcompete detection of rarer target sequences, leading to false negatives and added costs associated with greater sequencing depth needed to adequately capture a otherwise minority of target

sequences. Lab efforts to limit off-target amplification for the COI assay provided little return on investment, and thus use of this assay must be balanced against its relative inefficiency in delivering invertebrate versus bacterial sequences. This, combined with the somewhat patchy species coverage of such a taxonomically broad metabarcoding primer set, may ultimately encourage development of a suite of taxonomically more targeted assays.

However, this assay was also shown to have a broad taxonomic range of detectable taxa, including those off-target taxa found in this study. Despite lack of knowledge of their DNA source, the ability to identify different polychaetes, cnidarians, amphipods, cephalopods and many other represented classes may be useful for studies looking at invertebrate diversity in an area, or even species specific identification using metabarcoding. The breadth of invertebrate taxa that exist within the Gulf of Maine far extends those surveyed in this study, and therefore the assay itself may be useful in identifying different taxonomic groups not surveyed here.

3.5 Conclusions

Our results show that the COI primer used in this study has the potential of being a useful tool for biomonitoring and measuring ecosystem diversity of marine invertebrates. However, use of the COI primer set for marine invertebrates must be conducted with the knowledge that while it is able to detect a very diverse pool of species spanning this incredibly large diversity of animal life, it likely misses a significant proportion of individual species that are demonstrably present in an environment. Some of these species gaps may be partly filled by other complementary metabarcoding primer sets, while others may be incalcitrantly linked to particular taxa attributes. More testing to determine if off-target amplification can be reduced and how recovery of certain marine invertebrate taxonomic groups resolve with more stringent lab protocols should be employed as a next step in refining eDNA metabarcoding for GoM marine invertebrates. At the same time, it should remain a priority to evaluate alternative or

complementary primer sets that might overall improve taxonomic resolution for for GoM invertebrates. This could include the 18S primer set that faced inadequate amplification in this study or the alternative COI primer set of Leray et al. (2013). Finally, it should be recognized that the controlled conditions afforded by the NEAQ are not fully representative of wild environments and that some limited scale tests of the current COI assay, and others, in a natural GoM settings would be beneficial prior to widespread applications.

REFERENCES

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), 134–147. https://doi.org/10.1111/2041-210X.12849

Anderson, D. M., Kulis, D. M., Keafer, B. A., Gribble, K. E., Marin, R., & Scholin, C. A. (2005). Identification and enumeration of Alexandrium spp. from the Gulf of Maine using molecular probes. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *52*(19-21 SPEC. ISS.), 2467–2490. https://doi.org/10.1016/j.dsr2.2005.06.015

Andruszkiewicz, E. A., Starks, H. A., Chavez, F. P., Sassoubre, L. M., Block, B. A., & Boehm, A. B. (2017). Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *PLoS ONE*, *12*(4), 1–20. https://doi.org/10.1371/journal.pone.0176343

Baird, D. J., & Hajibabaei, M. (2012). Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, *21*(8), 2039–2044. https://doi.org/10.1111/j.1365-294X.2012.05519.x

Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, *17*(1), 1–17. https://doi.org/10.1007/s10592-015-0775-4

Bessey, C., Jarman, S. N., Berry, O., Olsen, Y. S., Bunce, M., Simpson, T., Power, M., McLaughlin, J., Edgar, G. J., & Keesing, J. (2020). Maximizing fish detection with eDNA metabarcoding. *Environmental DNA*, *2*(4), 493–504. https://doi.org/10.1002/edn3.74

Bruce, K., Blackman, R. C., Bourlat, S. J., Hellström, M., Bakker, J., Bista, I., Bohmann, K., Bouchez, A., Brys, R., Clark, K., & Elbrecht, V. (n.d.). *A practical guide to DNA-based methods for biodiversity assessment*.

Bush, A., Compson, Z. G., Monk, W. A., Porter, T. M., Steeves, R., Emilson, E., Gagne, N., Hajibabaei, M., Roy, M., & Baird, D. J. (2019). Studying Ecosystems With DNA Metabarcoding: Lessons From Biomonitoring of Aquatic Macroinvertebrates. *Frontiers in Ecology and Evolution*, *7*(November), 1–12. https://doi.org/10.3389/fevo.2019.00434

Bush, A., Monk, W. A., Compson, Z. G., Peters, D. L., Porter, T. M., Shokralla, S., Wright, M. T. G., Hajibabaei, M., & Baird, D. J. (2020). DNA metabarcoding reveals metacommunity dynamics in a threatened boreal wetland wilderness. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(15), 8539–8545. https://doi.org/10.1073/pnas.1918741117

Bylemans, J., Furlan, E. M., Gleeson, D. M., Hardy, C. M., & Duncan, R. P. (2018). Does Size Matter? An Experimental Evaluation of the Relative Abundance and Decay Rates of Aquatic Environmental DNA. *Environmental Science and Technology*, *52*(11), 6408–6416. https://doi.org/10.1021/acs.est.8b01071

Clarke, L. J., Soubrier, J., Weyrich, L. S., & Cooper, A. (2014). Environmental metabarcodes for insects: In silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, *14*(6), 1160–1170. https://doi.org/10.1111/1755-0998.12265

Closek, C. J., Santora, J. A., Starks, H. A., Schroeder, I. D., Andruszkiewicz, E. A., Sakuma, K. M., Bograd, S. J., Hazen, E. L., Field, J. C., & Boehm, A. B. (2019). Marine Vertebrate Biodiversity and

Distribution Within the Central California Current Using Environmental DNA (eDNA) Metabarcoding and Ecosystem Surveys. *Frontiers in Marine Science*, *6*(December), 1–17. https://doi.org/10.3389/fmars.2019.00732

Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., Genner, M. J., & Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution*, *10*(11), 1985–2001. https://doi.org/10.1111/2041-210X.13276

Comeau, A. M., Li, W. K. W., Tremblay, J. É., Carmack, E. C., & Lovejoy, C. (2011). Arctic ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS ONE*, *6*(11). https://doi.org/10.1371/journal.pone.0027492

Crane, L. C., Goldstein, J. S., Thomas, D. W., Rexroth, K. S., & Watts, A. W. (2021). Effects of life stage on eDNA detection of the invasive European green crab (Carcinus maenas) in estuarine systems. *Ecological Indicators*, *124*(January), 107412. https://doi.org/10.1016/j.ecolind.2021.107412

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biology Letters*, *10*(9), 2–5. https://doi.org/10.1098/rsbl.2014.0562

Deagle, B. E., Kirkwood, R., & Jarman, S. N. (2009). Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology*, *18*(9), 2022–2038. https://doi.org/10.1111/j.1365-294X.2009.04158.x

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., de Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895. https://doi.org/10.1111/mec.14350

Deiner, K., Renshaw, M. A., Li, Y., Olds, B. P., Lodge, D. M., & Pfrender, M. E. (2017). Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods in Ecology and Evolution*, *8*(12), 1888–1898. https://doi.org/10.1111/2041-210X.12836

Dejean, T., Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P., & Miaud, C. (2011). Persistence of environmental DNA in freshwater ecosystems. *PLoS ONE*, *6*(8). https://doi.org/10.1371/journal.pone.0023398

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E., & Miaud, C. (2012). Improved detection of an alien invasive species through environmental DNA barcoding: The example of the American bullfrog Lithobates catesbeianus. *Journal of Applied Ecology*, *49*(4), 953–959. https://doi.org/10.1111/j.1365-2664.2012.02171.x

Dell'Anno, A., & Corinaldesi, C. (2004). Degradation and turnover of extracellular DNA in marine sediments: Ecological and methodological considerations. *Applied and Environmental Microbiology*, *70*(7), 4384–4386. https://doi.org/10.1128/AEM.70.7.4384-4386.2004

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS ONE*, *10*(7). https://doi.org/10.1371/journal.pone.0130324

Elbrecht, V., & Leese, F. (2017). Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science*, *5*(APR), 1–11.

https://doi.org/10.3389/fenvs.2017.00011

Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., Mahon, A. R., Pfrender, M. E., Lamberti, G. A., & Lodge, D. M. (2016). Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 29–41. https://doi.org/10.1111/1755-0998.12433

Ficetola, Gentile F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., Gielly, L., Lopes, C. M., Boyer, F., Pompanon, F., Rayé, G., & Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, *15*(3), 543–556. https://doi.org/10.1111/1755-0998.12338

Ficetola, Gentile Francesco, Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, *4*(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Fonseca, V. G., Carvalho, G. R., Sung, W., Johnson, H. F., Power, D. M., Neill, S. P., Packer, M., Blaxter, M. L., Lambshead, P. J. D., Thomas, W. K., & Creer, S. (2010). Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications*, *1*(7). https://doi.org/10.1038/ncomms1095

Forsström, T., & Vasemägi, A. (2016). Can environmental DNA (eDNA) be used for detection and monitoring of introduced crab species in the Baltic Sea? *Marine Pollution Bulletin*, *109*(1), 350–355. https://doi.org/10.1016/j.marpolbul.2016.05.054

García-Machado, E., Laporte, M., Normandeau, E., Hernández, C., Côté, G., Paradis, Y., Mingelbier, M., & Bernatchez, L. (2021). Fish community shifts along a strong fluvial environmental gradient revealed by eDNA metabarcoding. *Environmental DNA*, *March*, 1–18. https://doi.org/10.1002/edn3.221

Geerts, A. N., Boets, P., Van den Heede, S., Goethals, P., & Van der heyden, C. (2018). A search for standardized protocols to detect alien invasive crayfish based on environmental DNA (eDNA): A lab and field evaluation. *Ecological Indicators*, *84*(October 2016), 564–572. https://doi.org/10.1016/j.ecolind.2017.08.068

Gerlach, J., Samways, M., & Pryke, J. (2013). Terrestrial invertebrates as bioindicators: An overview of available taxonomic groups. *Journal of Insect Conservation*, *17*(4), 831–850. https://doi.org/10.1007/s10841-013-9565-9

Gielings, R., Fais, M., Fontaneto, D., Creer, S., Costa, F. O., Renema, W., & Macher, J. N. (2021). DNA Metabarcoding Methods for the Study of Marine Benthic Meiofauna: A Review. *Frontiers in Marine Science*, *8*(September), 1–13. https://doi.org/10.3389/fmars.2021.730063

Gold, Z., Choi, E., Kacev, D., Frable, B., Burton, R., Thompson, A., & Barber, P. (2020). FishCARD : Fish 12S California Current Specific Reference Database for Enhanced Metabarcoding Efforts. *Authorea*, 1–14.

Gold, Z., Sprague, J., Kushner, D. J., Marin, E. Z., & Barber, P. H. (2021). eDNA metabarcoding as a biomonitoring tool for marine protected areas. *PLoS ONE*, *16*(2 February 2021), 1–19. https://doi.org/10.1371/journal.pone.0238557

Harris, L. G., & Tyrrell, M. C. (2001). Changing community states in the Gulf of Maine: Synergism between invaders, overfishing and climate change. *Biological Invasions*, *3*(1), 9–21.

https://doi.org/10.1023/A:1011487219735

Harrison, J. B., Sunday, J. M., & Rogers, S. M. (2019). Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1915). https://doi.org/10.1098/rspb.2019.1409

Jackman, J. M., Benvenuto, C., Coscia, I., Oliveira Carvalho, C., Ready, J. S., Boubli, J. P., Magnusson, W. E., McDevitt, A. D., & Guimarães Sales, N. (2021). eDNA in a bottleneck: Obstacles to fish metabarcoding studies in megadiverse freshwater systems. *Environmental DNA*, *3*(4), 837–849. https://doi.org/10.1002/edn3.191

Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, *4*(2), 150–157. https://doi.org/10.1111/j.1755-263X.2010.00158.x

Keeley, N., Wood, S. A., & Pochon, X. (2018). Development and preliminary validation of a multi-trophic metabarcoding biotic index for monitoring benthic organic enrichment. *Ecological Indicators*, *85*(November 2017), 1044–1057. https://doi.org/10.1016/j.ecolind.2017.11.014

Keller, A., Hohlfeld, S., Kolter, A., Schultz, J., Schultz, J., Gemeinholzer, B., Ankenbrand, M. J., & Ankenbrand, M. J. (2020). BCdatabaser: On-The-fly reference database creation for (meta-)barcoding. *Bioinformatics*, *36*(8), 2630–2631. https://doi.org/10.1093/bioinformatics/btz960

Kelly, R. P., Port, J. A., Yamahara, K. M., & Crowder, L. B. (2014). Using environmental DNA to census marine fishes in a large mesocosm. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0086175

Kelly, R. P., Shelton, A. O., & Gallego, R. (2019). Understanding PCR Processes to Draw Meaningful Conclusions from Environmental DNA Studies. *Scientific Reports*, *9*(1), 1–14. https://doi.org/10.1038/s41598-019-48546-x

Klemm, D. J., Blocksom, K. A., Fulk, F. A., Herlihy, A. T., Hughes, R. M., Kaufmann, P. R., Peck, D. V., Stoddard, J. L., Thoeny, W. T., Griffith, M. B., & Davis, W. S. (2003). Development and evaluation of a Macroinvertebrate Biotic Integrity Index (MBII) for regionally assessing Mid-Atlantic Highlands streams. *Environmental Management*, *31*(5), 656–669. https://doi.org/10.1007/s00267-002-2945-7

Klymus, K. E., Marshall, N. T., & Stepien, C. A. (2017). Environmental DNA (eDNA) metabarcoding assays to detect invasive invertebrate species in the Great Lakes. *PLoS ONE*, *12*(5), 1–24. https://doi.org/10.1371/journal.pone.0177643

Lacoursière-Roussel, A., Côté, G., Leclerc, V., & Bernatchez, L. (2016). Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. *Journal of Applied Ecology*, *53*(4), 1148–1157. https://doi.org/10.1111/1365-2664.12598

Lejzerowicz, F., Esling, P., Pillet, L., Wilding, T. A., Black, K. D., & Pawlowski, J. (2015). High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. *Scientific Reports*, *5*(April), 1–10. https://doi.org/10.1038/srep13932

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*(1), 1–14. https://doi.org/10.1186/1742-9994-10-34

McElroy, M. E., Dressler, T. L., Titcomb, G. C., Wilson, E. A., Deiner, K., Dudley, T. L., Eliason, E. J., Evans, N. T., Gaines, S. D., Lafferty, K. D., Lamberti, G. A., Li, Y., Lodge, D. M., Love, M. S., Mahon, A. R., Pfrender, M. E., Renshaw, M. A., Selkoe, K. A., & Jerde, C. L. (2020). Calibrating Environmental DNA Metabarcoding to Conventional Surveys for Measuring Fish Species Richness. *Frontiers in Ecology and Evolution*, *8*, 0–12. https://doi.org/10.3389/fevo.2020.00276

Meyer, A., Boyer, F., Valentini, A., Bonin, A., Ficetola, G. F., Beisel, J. N., Bouquerel, J., Wagner, P., Gaboriaud, C., Leese, F., Dejean, T., Taberlet, P., & Usseglio-Polatera, P. (2021). Morphological vs. DNA metabarcoding approaches for the evaluation of stream ecological status with benthic invertebrates: Testing different combinations of markers and strategies of data filtering. *Molecular Ecology*, *30*(13), 3203–3220. https://doi.org/10.1111/mec.15723

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species. *Royal Society Open Science*, *2*(7). https://doi.org/10.1098/rsos.150088

Miya, Masaki, Gotoh, R. O., & Sado, T. (2020). MiFish metabarcoding: a high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. In *Fisheries Science* (Vol. 86, Issue 6). Springer Japan. https://doi.org/10.1007/s12562-020-01461-x

Morey, K. C., Bartley, T. J., & Hanner, R. H. (2020). Validating environmental DNA metabarcoding for marine fishes in diverse ecosystems using a public aquarium. *Environmental DNA*, *2*(3), 330–342. https://doi.org/10.1002/edn3.76

Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., Green, R. E., & Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources*, *18*(5), 927–939. https://doi.org/10.1111/1755-0998.12895

O'Rourke, D. R., Bokulich, N. A., Jusino, M. A., MacManes, M. D., & Foster, J. T. (2020). A total crapshoot? Evaluating bioinformatic decisions in animal diet metabarcoding analyses. *Ecology and Evolution*, *10*(18), 9721–9739. https://doi.org/10.1002/ece3.6594

Pearman, J. K., Anlauf, H., Irigoien, X., & Carvalho, S. (2016). Please mind the gap - Visual census and cryptic biodiversity assessment at central Red Sea coral reefs. *Marine Environmental Research*, *118*, 20–30. https://doi.org/10.1016/j.marenvres.2016.04.011

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Bris, A. Le, Mills, K. E., Nye, J. A., Record, N. R., Scannell, H. A., Scott, J. D., Sherwood, G. D., & Thomas, A. C. (2016). Response to comments on "Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. *Science*, *352*(6284), 423. https://doi.org/10.1126/science.aae0463

Porazinska, D. L., Giblin-Davis, R. M., Faller, L., Farmerie, W., Kanzaki, N., Morris, K., Powers, T. O., Tucker, A. E., Sung, W., & Thomas, W. K. (2009). Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources*, *9*(6), 1439–1450. https://doi.org/10.1111/j.1755-0998.2009.02611.x

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), 590–596. https://doi.org/10.1093/nar/gks1219

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., & Gough, K. C. (2014). The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool

in ecology. *Journal of Applied Ecology*, *51*(5), 1450–1459. https://doi.org/10.1111/1365-2664.12306

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). EcoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, *39*(21), 1–11. https://doi.org/10.1093/nar/gkr732

Salter, I., Joensen, M., Kristiansen, R., Steingrund, P., & Vestergaard, P. (2019). Environmental DNA concentrations are correlated with regional biomass of Atlantic cod in oceanic waters. *Communications Biology*, *2*(1), 1–9. https://doi.org/10.1038/s42003-019-0696-8

Sard, N. M., Herbst, S. J., Nathan, L., Uhrig, G., Kanefsky, J., Robinson, J. D., & Scribner, K. T. (2019). Comparison of fish detections, community diversity, and relative abundance using environmental DNA metabarcoding and traditional gears. *Environmental DNA*, *1*(4), 368–384. https://doi.org/10.1002/edn3.38

Schmelzle, M. C., & Kinziger, A. P. (2016). Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. *Molecular Ecology Resources*, *16*(4), 895–908. https://doi.org/10.1111/1755-0998.12501

Steyaert, M., Priestley, V., Osborne, O., Herraiz, A., Arnold, R., & Savolainen, V. (2020). Advances in metabarcoding techniques bring us closer to reliable monitoring of the marine benthos. *Journal of Applied Ecology*, *57*(11), 2234–2245. https://doi.org/10.1111/1365-2664.13729

Stoeckle, M. Y., Das Mishu, M., & Charlop-Powers, Z. (2020). Improved Environmental DNA Reference Library Detects Overlooked Marine Fishes in New Jersey, United States. *Frontiers in Marine Science*, *7*(May). https://doi.org/10.3389/fmars.2020.00226

Stoeckle, M. Y., Soboleva, L., & Charlop-Powers, Z. (2017). Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. *PLoS ONE*, *12*(4), 1–15. https://doi.org/10.1371/journal.pone.0175186

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., & Willerslev, E. (2012). Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples. *PLoS ONE*, *7*(8), 1–9. https://doi.org/10.1371/journal.pone.0041732

Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., Orlando, L., & Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, *21*(11), 2565–2573. https://doi.org/10.1111/j.1365-294X.2011.05418.x

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019

Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., & Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological Applications*, *13*(6), 1790–1801. https://doi.org/10.1890/02-5078

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., Bellemain, E., Besnard, A.,

Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G. H., Geniez, P., Pont, D., Argillier, C., Baudoin, J. M., … Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, *25*(4), 929–942. https://doi.org/10.1111/mec.13428

Valsecchi, E., Bylemans, J., Goodman, S. J., Lombardi, R., Carr, I., Castellano, L., Galimberti, A., & Galli, P. (2020). Novel universal primers for metabarcoding environmental DNA surveys of marine mammals and other marine vertebrates. *Environmental DNA*, *2*(4), 460–476. https://doi.org/10.1002/edn3.72

Wangensteen, O. S., Palacín, C., Guardiola, M., & Turon, X. (2018). DNA metabarcoding of littoral hardbottom communities: High diversity and database gaps revealed by two molecular markers. *PeerJ*, *2018*(5), 1–30. https://doi.org/10.7717/peerj.4705

Watts, C., Dopheide, A., Holdaway, R., Davis, C., Wood, J., Thornburrow, D., & Dickie, I. A. (2019). DNA metabarcoding as a tool for invertebrate community monitoring: a case study comparison with conventional techniques. *Austral Entomology*, *58*(3), 675–686. https://doi.org/10.1111/aen.12384

Wilcox, T. M., McKelvey, K. S., Young, M. K., Jane, S. F., Lowe, W. H., Whiteley, A. R., & Schwartz, M. K. (2013). Robust Detection of Rare Species Using Environmental DNA: The Importance of Primer Specificity. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0059520

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D. A., & Cooper, A. (2003). Williserv2003.PDF. *Science*, *300*(2).

Wood, Z. T., Erdman, B. F., York, G., Trial, J. G., & Kinnison, M. T. (2020). Experimental assessment of optimal lotic eDNA sampling and assay multiplexing for a critically endangered fish. *Environmental DNA*, *2*(4), 407–417. https://doi.org/10.1002/edn3.64

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, *3*(4), 613–623. https://doi.org/10.1111/j.2041-210X.2012.00198.x

Zhang, S., Zhao, J., & Yao, M. (2020). A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. *Methods in Ecology and Evolution*, *11*(12), 1609–1625. https://doi.org/10.1111/2041-210X.13485

APPENDICIES

Appendix A: Aquarium census and thermocycling table

Enclosure Name	Class Name	Genus	Species	Common Name	Population Male	Population Female	Population Unknown
Boulder Reef Community (22)	ANTHOZOA	Metridium	senile	Clonal plumose anemone	0	0	25
Boulder Reef Community (22)	ANTHOZOA	Urticina	felina	Northern red anemone	0	0	22
Eastport (25)	ANTHOZOA	Gersemia	rubiformis	Sea strawberry	0	0	1
Eastport (25)	ANTHOZOA	Metridium	senile	Clonal plumose anemone	0	0	52
Eastport (25)	ANTHOZOA	Urticina	felina	Northern red anemone	0	0	15
EOS TP	ANTHOZOA	Stomphia	didemon	Cowardly anemone	0	0	4
EOS TP	ANTHOZOA	Urticina	felina	Northern red anemone	0	0	12
Eastport (25)	ASCIDIACEA	Ciona	intestinalis	Sea vase	0	0	2
Shorebirds	AVES	Calidris	alba	Sanderling	1	0	0
Shorebirds	AVES	Calidris	minutilla	Least sandpiper	0	1	0
Shorebirds	AVES	Calidris	pusilla	Semipalmated sandpiper	0	1	0
Shorebirds	AVES	Charadrius	melodus	Piping plover	0	1	0
Shorebirds	AVES	Charadrius	semipalmatus	Semipalmated plover	0	1	0
Shorebirds	AVES	Sterna	hirundo	Common tern	0	2	0
Eastport (25)	BIVALVIA	Modiolus	modiolus	Northern horse mussel	0	0	16
Eastport (25)	BIVALVIA	Mytilus	edulis	blue mussel	0	0	9

EOS TP	BIVALVIA	Crassostrea	virginica	Atlantic oyster	0	0	12
EOS TP	BIVALVIA	Mercenaria	mercenaria	Quahog	0	0	20
EOS TP	BIVALVIA	Mytilus	edulis	blue mussel	0	0	70
EOS TP	BIVALVIA	Placopecten	magellanicus	Atlantic deep-sea scallop	0	0	1
Goosefish (24)	BIVALVIA	Modiolus	modiolus	Northern horse mussel	0	0	1
Shorebirds	BIVALVIA	Mytilus	edulis	blue mussel	0	0	40
Boulder Reef Community (22)	CHONDRICHTHYES	Hydrolagus	colliei	Spotted ratfish	1	0	0
Sandy Bottom Community (23)	CHONDRICHTHYES	Leucoraja	erinacea	Little skate	0	2	0
Sandy Bottom Community (23)	CHONDRICHTHYES	Scyliorhinus	retifer	Chain cat shark	0	2	0
Boulder Reef Community (22)	CRUSTACEA	Homarus	americanus	American lobster	1	0	0
Eastport (25)	CRUSTACEA	Balanus	balanoides	Northern rock barnacle	0	0	6
Eastport (25)	CRUSTACEA	Pagurus	acadianus	Acadian hermit crab	0	0	18
Eastport (25)	CRUSTACEA	Pagurus	arcuatus	Hairy hermit crab	0	0	6
EOS TP	CRUSTACEA	Cancer	borealis	Jonah crab	0	0	1
EOS TP	CRUSTACEA	Libinia	emarginata	Short-clawed spider crab	2	1	0
EOS TP	CRUSTACEA	Pagurus	acadianus	Acadian hermit crab	0	0	14
EOS TP	CRUSTACEA	Pagurus	arcuatus	Hairy hermit crab	0	0	3

EOS TP	CRUSTACEA	Pagurus	longicarpus	Long-armed hermit crab	0	0	47
EOS TP Lower	CRUSTACEA	Homarus	americanus	American lobster	0	1	0
EOS TP Upper	CRUSTACEA	Libinia	dubia	Southern spider crab	0	0	1
Goosefish (24)	CRUSTACEA	Homarus	americanus	American lobster	1	1	0
Shorebirds	CRUSTACEA	Cancer	borealis	Jonah crab	0	0	2
Shorebirds	CRUSTACEA	Hyas	coarctatus	Toad crab	0	0	1
Shorebirds	CRUSTACEA	Libinia	dubia	Southern spider crab	0	0	2
Shorebirds	CRUSTACEA	Pagurus	longicarpus	Long-armed hermit crab	0	0	1
Eastport (25)	ECHINOIDEA	Strongylocentrotus	droebachiensis	Green sea urchin	0	0	7
EOS TP	ECHINOIDEA	Strongylocentrotus	droebachiensis	Green sea urchin	0	0	53
Sandy Bottom Community (23)	ECHINOIDEA	Strongylocentrotus	droebachiensis	Green sea urchin	0	0	11
Shorebirds	ECHINOIDEA	Arbacia	punctulata	Purple sea urchin	0	0	11
Shorebirds	ECHINOIDEA	Strongylocentrotus	droebachiensis	Green sea urchin	0	0	2
Boulder Reef Community (22)	GASTROPODA	Buccinum	undatum	Common northern whelk	0	0	170
Boulder Reef Community (22)	GASTROPODA	Neptunea	lyrata	Common northwest neptune	0	0	15
Eastport (25)	GASTROPODA	Buccinum	undatum	Common northern whelk	0	0	6
EOS TP	GASTROPODA	Buccinum	undatum	Common northern whelk	0	0	3
EOS TP	GASTROPODA	Diodora	aspera	Rough keyhole limpet	0	0	151
EOS TP	GASTROPODA	Euspira	heros	Northern moon snail	0	0	2

EOS TP	GASTROPODA	Littorina	littorea	Common periwinkle	0	0	259
EOS TP	GASTROPODA	Neptunea	lyrata	Common northwest neptune	0	0	14
Goosefish (24)	GASTROPODA			Moon snail	0	0	3
Sandy Bottom Community (23)	GASTROPODA			Moon snail	0	0	3
Boulder Reef Community (22)	HOLOTHUROIDEA	Cucumaria	frondosa	Orange-footed sea cucumber	0	0	51
Boulder Reef Community (22)	HOLOTHUROIDEA	Psolus	fabricii	Scarlet sea cucumber	0	0	2
Eastport (25)	HOLOTHUROIDEA	Chiridota	laevis	Silky cucumber	0	0	2
Eastport (25)	HOLOTHUROIDEA	Cucumaria	frondosa	Orange-footed sea cucumber	0	0	2
Eastport (25)	HOLOTHUROIDEA	Psolus	fabricii	Scarlet sea cucumber	0	0	1
EOS TP	HOLOTHUROIDEA	Cucumaria	frondosa	Orange-footed sea cucumber	0	0	4
Front Seal Pool (FSP)	MAMMALIA	Phoca	vitulina	Harbor seal	2	3	0
EOS TP	MEROSTOMATA	Limulus	polyphemus	Horseshoe crab	1	0	1
Boulder Reef Community (22)	OSTEICHTHYES	Centropristis	striata	Black seabass	0	0	1
Boulder Reef Community (22)	OSTEICHTHYES	Helicolenus	dactylopterus	Blackbelly Rosefish	0	0	8
Boulder Reef Community (22)	OSTEICHTHYES	Macroramphosus	scolopax	Longspine snipefish	0	0	49
Boulder Reef Community (22)	OSTEICHTHYES	Sebastes	fasciatus	Acadian Redfish	0	0	16

Boulder Reef Commun ity (22)	OSTEICHTHYE S	Tautoga	onitis	Tautog	0	0	1
Boulder Reef Commun ity (22)	OSTEICHTHYE S	Tautogolabrus	adspersus	Cunner	0	0	1
Eastport (25)	OSTEICHTHYE S	Pholis	gunnellus	Rock gunnel	0	0	7
Eastport (25)	OSTEICHTHYE S	Sebastes	fasciatus	Acadian Redfish	0	0	10
Eastport (25)	OSTEICHTHYE S	Ulvaria	subbifurcata	Radiated shanny	0	0	2
Eastport (25)	OSTEICHTHYE S	Zoarces	americanus	Ocean eelpout	0	0	1
EOS TP	OSTEICHTHYE S	Myoxocephalus	aenaeus	Grubby sculpin	0	0	9
EOS TP	OSTEICHTHYE S	Pholis	gunnellus	Rock gunnel	0	0	7
EOS TP	OSTEICHTHYE S	Pseudopleuron ectes	americanus	Winter flounder	0	0	3
Goosefis h (24)	OSTEICHTHYE S	Centropristis	striata	Black seabass	0	0	1
Goosefis h (24)	OSTEICHTHYE S	Microgadus	tomcod	Atlantic tomcod	0	0	1
Goosefis h (24)	OSTEICHTHYE S	Myoxocephalus	scorpius	Shorthorn sculpin	0	0	1
Goosefis h (24)	OSTEICHTHYE S	Zoarces	americanus	Ocean eelpout	0	1	0
Sandy Bottom Commun ity (23)	OSTEICHTHYE S	Hippoglossus	hippoglossus	Atlantic halibut	1	0	1
Sandy Bottom Commun ity (23)	OSTEICHTHYE S	Melanogrammu s	aeglefinus	Haddock	0	0	4
Sandy Bottom Commun ity (23)	OSTEICHTHYE S	Morone	saxatilis	Striped seabass	0	0	10
Sandy Bottom Commun ity (23)	OSTEICHTHYE S	Myoxocephalus	octodecemspin osus	Longhorn sculpin	0	0	1
Sandy Bottom Commun ity (23)	OSTEICHTHYE S	Myoxocephalus	scorpius	Shorthorn sculpin	0	0	2

Sandy Bottom Community (23)	OSTEICHTHYES	Pseudopleuronectes	americanus	Winter flounder	0	0	3
Sandy Bottom Community (23)	OSTEICHTHYES	Tautoga	onitis	Tautog	0	0	1
Shorebirds	OSTEICHTHYES	Cyprinodon	variegatus	Sheepshead minnow	0	0	26
Shorebirds	OSTEICHTHYES	Fundulus	heteroclitus	Mummichog	0	0	9
Shorebirds	OSTEICHTHYES	Fundulus	majalis	Striped killifish	0	0	6
Shorebirds	OSTEICHTHYES	Pholis	gunnellus	Rock gunnel	0	0	1
Boulder Reef Community (22)	STELLEROIDEA	Asterias	rubens	Northern sea star	0	0	9
Boulder Reef Community (22)	STELLEROIDEA	Crossaster	papposus	Spiny Sun Star	0	0	8
Boulder Reef Community (22)	STELLEROIDEA	Ophiopholis	aculeata	Daisy brittlestar	0	0	17
Eastport (25)	STELLEROIDEA	Asterias	rubens	Northern sea star	0	0	1
Eastport (25)	STELLEROIDEA	Crossaster	papposus	Spiny Sun Star	0	0	3
Eastport (25)	STELLEROIDEA	Henricia	sanguinolenta	Atlantic blood sea star	0	0	10
Eastport (25)	STELLEROIDEA	Hippasteria	phrygiana	Horse sea star	0	0	1
Eastport (25)	STELLEROIDEA	Ophiopholis	aculeata	Daisy brittlestar	0	0	31
EOS TP	STELLEROIDEA	Asterias	forbesi	Bay sea star	0	0	5
EOS TP	STELLEROIDEA	Asterias	rubens	Northern sea star	0	0	81
EOS TP	STELLEROIDEA	Henricia	sanguinolenta	Atlantic blood sea star	0	0	4
EOS TP	STELLEROIDEA	Ophiopholis	aculeata	Daisy brittlestar	0	0	6
Goosefish (24)	STELLEROIDEA	Henricia	sanguinolenta	Atlantic blood sea star	0	0	37

Sandy Bottom Community (23)	STELLEROIDEA	Solaster	endeca	Purple sun star	0	0	14

Table A1. Aquarium census for all taxa

Locus	Publication	Primer Sequences 5' - 3'	PCR recipe	Thermocycling conditions	Post PCR clean up	Post indexing cleanup
12S	Miya et al. 2015	MiFish-U-F: GTCGGTAAAACTCGTGCCAGC MiFish-U-R: GTTTGACCCTAATCTATGGGGTGATAC	10uL PCR master mix 1uL DNA extract 0.5uL F (10uM) 0.5uL R (10uM) 8uL nuclease free water	95°C for 10:00 13 cycles of 95°C for 0:30 69.5°C* for 0:30 72°C for 1:30 25 cycles of 95°C for 0:30 65°C for 0:30 72°C for 0:45 72°C for 10:00	Double-sided**; 0.85x and 0.80x	Single sided (depletion of small frags)' 0.80x
12S	Riaz et al. 2011	12S-V5-F: ACTGGGATTAGATACCCC R: TAGAACAGGCTCCTCTAG	10uL PCR master mix 2uL DNA extract 0.5uL F (10uM) 0.5uL R (10uM) 8uL nuclease free water	95°C for 10:00 40 cycles of 95°C for 0:30 60°C for 0:30 72°C for 0:30 72°C for 5:00	Single sided; 1.80x	Single sided; 1.20x
16S	Valsecchi et al. 2016	MarVer3F: AGACGAGAAGACCCTRTG MarVer3R: GGATTGCGCTGTTATCCC	10uL PCR master mix 1uL DNA extract 0.5uL F (10uM)	95°C for 10:00 38 cycles of 95°C for 0:30 60°C for 0:30 72°C for 0:30	Double sided; 0.70x, 1.20x	Single sided; 1.20x

			0.5uL R (10uM) 8uL nucleas e free water	72°C for 5:00		
18S	Comeau et al. 2011	E572F: CYGCGGTAATTCCAGCTC E1009R: AYGGTATCTRATCRTCTTYG	10uL PCR master mix 1uL DNA extract 0.5uL F (10uM) 0.5uL R (10uM) 8uL nucleas e free water	95°C for 10:00 35 cycles of 95°C for 0:30 55°C for 0:30 72°C for 0:30 72°C for 5:00	Doubl e sided; 0.70x and 0.80x	Single sided; 0.80x
COI	Elbrecht et al. 2017	BF1: ACWGGWTGRACWGTNTAYCC BR1: ARYATDGTRATDGCHCCDGC	10uL PCR master mix 3uL DNA extract 0.5uL F (10uM) 0.5uL R (10uM) 8uL nucleas e free water	95°C for 10:00 35 cycles of 95°C for 0:30 50°C for 0:30 72°C for 0:30 72°C for 5:00	Doubl e sided; 0.60x and 1.20x	Single sided; 0.80x

Table A2. Primer sequences, thermocycling conditions and recipes, and size selection protocols for each of the primer sets used in this study.

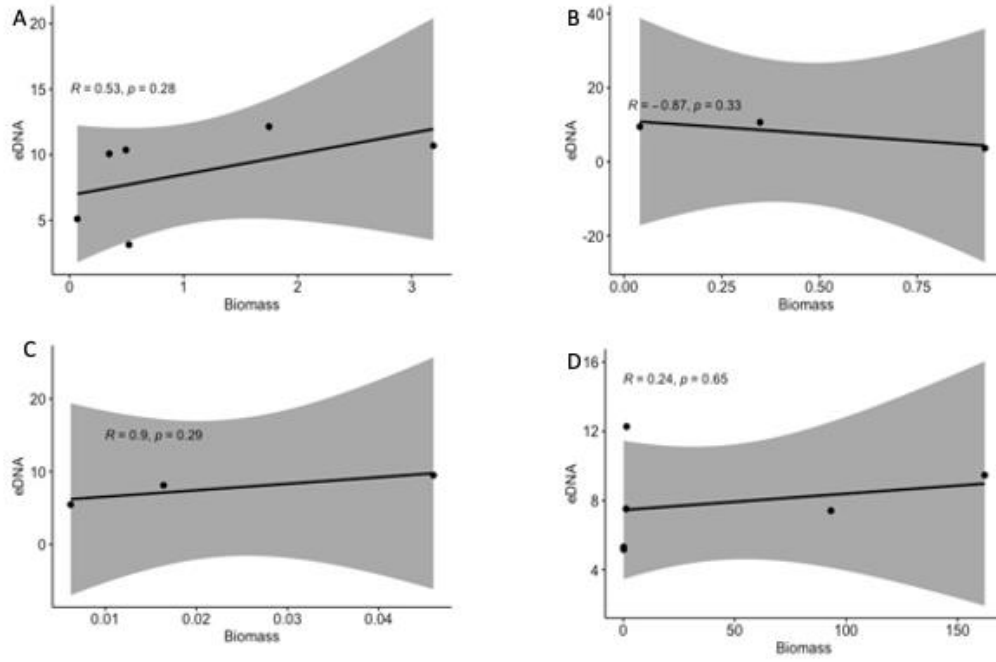

Figure B1. Pearson correlation plots for each tank with a minimum of three analyzable taxa for the MiFish assay; a.) Stellwagen Boulder Reef b.) Isle of Shoals c.) Boston Harbor Islands/ Shorebirds d.) Stellwagen Sandy Bottom

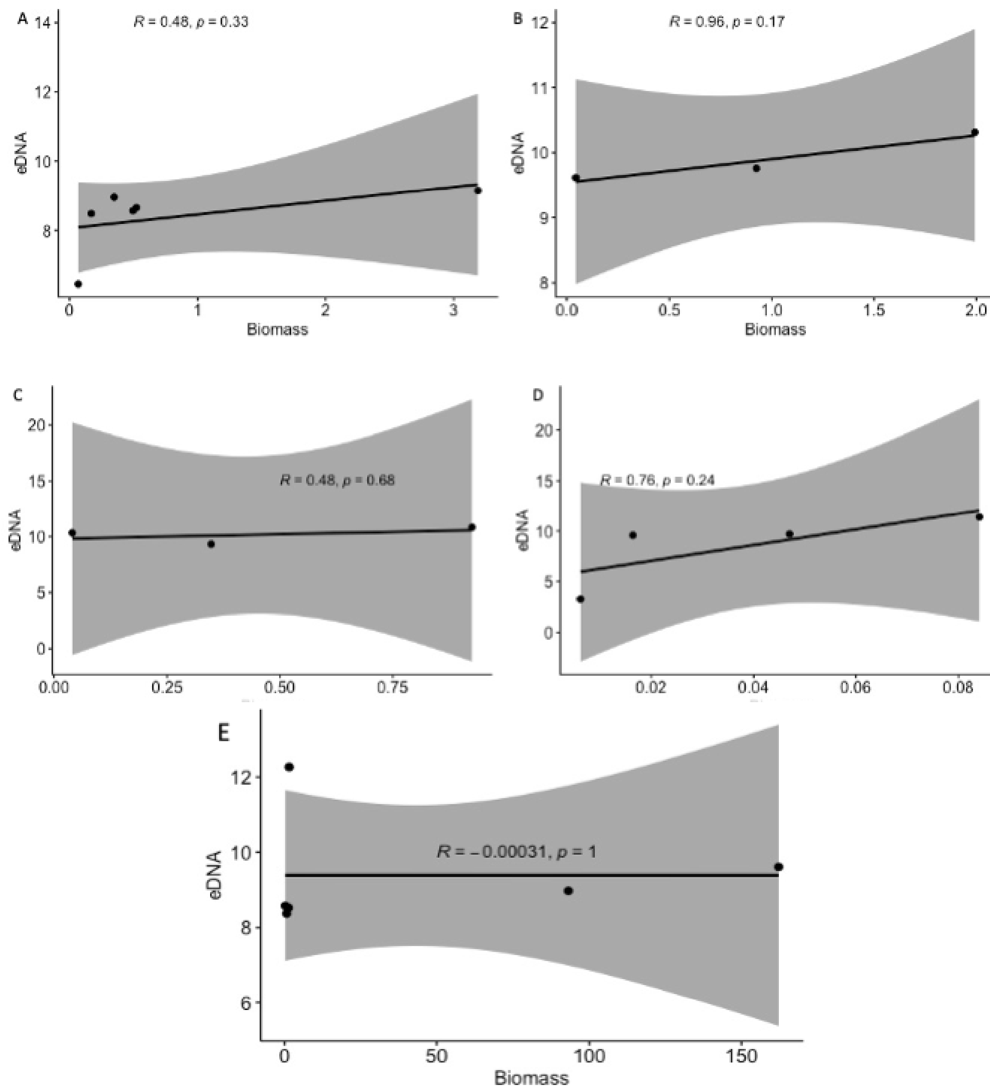

Figure B2. Pearson's correlation plots comparing eDNA sequences to fish biomass for each tank using the MarVer assay; a.) Stellwagen Boulder Reef b.) Eastport, ME c.) Isle of Shoals d.) Boston Harbor Islands/ Shorebirds e.) Stellwagen Sandy Bottom

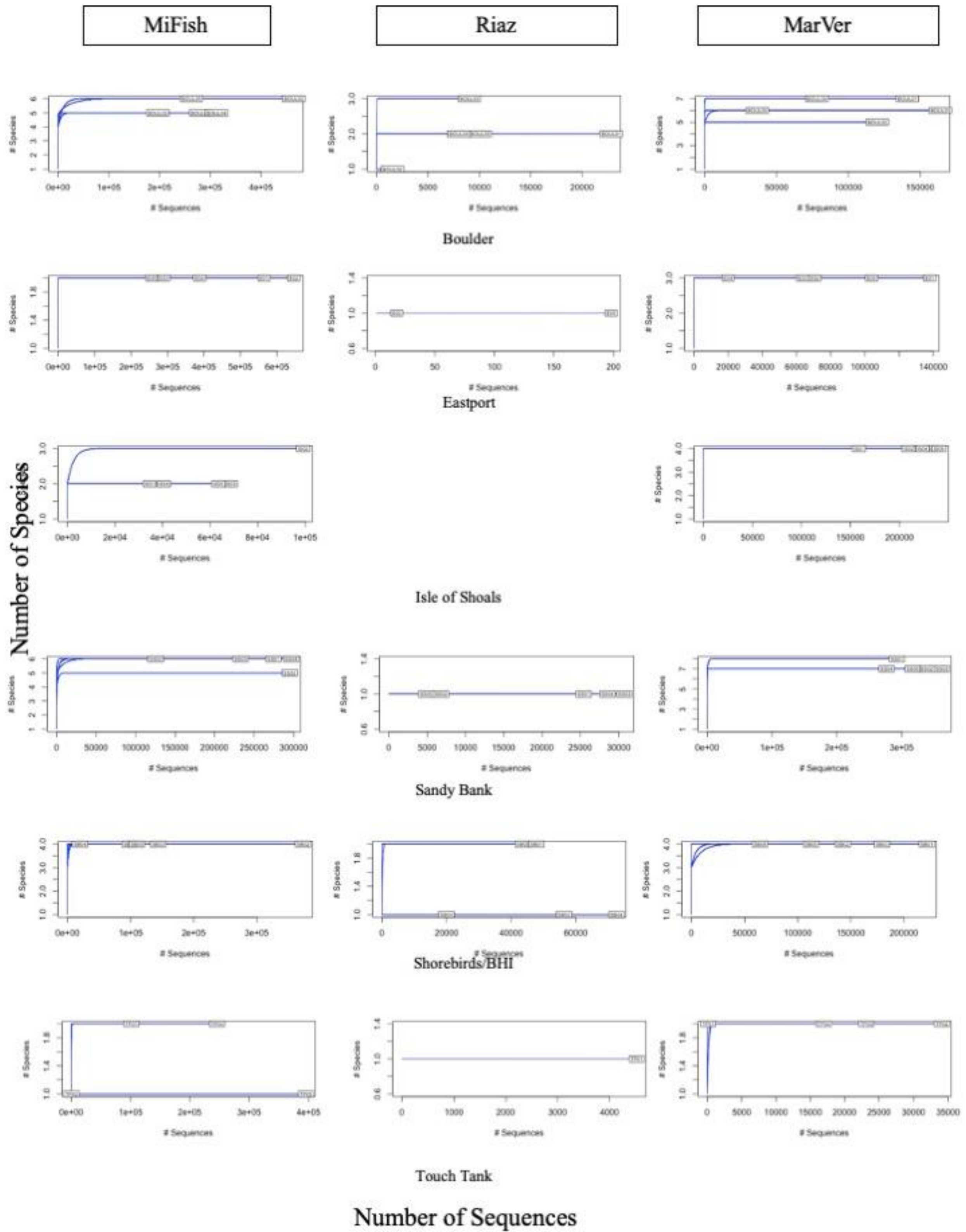

Figure B3*. Read rarefaction for all tanks across all three vertebrate assays. Boxes indicate individual sampling replicates from a given tank. These figures are representative of how many sequences were needed to first detect different taxa for each sampling replicate in a tank.

*Notes on Figure B3:

Read rarefaction was done using the vegan package in R (version 4.1.1). This was done by randomly sampling each ASV read counts to make a plot comparing number of sequences in a sample to number of species in a sample. The majority of tanks across all three vertebrate assays show saturation of number of species within samples and within tanks, with the exception of the Isle of Shoals tank for the Riaz assay which had no sampling replicates with data. Some tanks also did not have sampling replicates amplify, which results in differing amounts of boxes represented in the graph. Results indicate that in most cases for the vertebrate assays used in this study, read depth in sequencing was sufficient to identify all genetically detectable species.

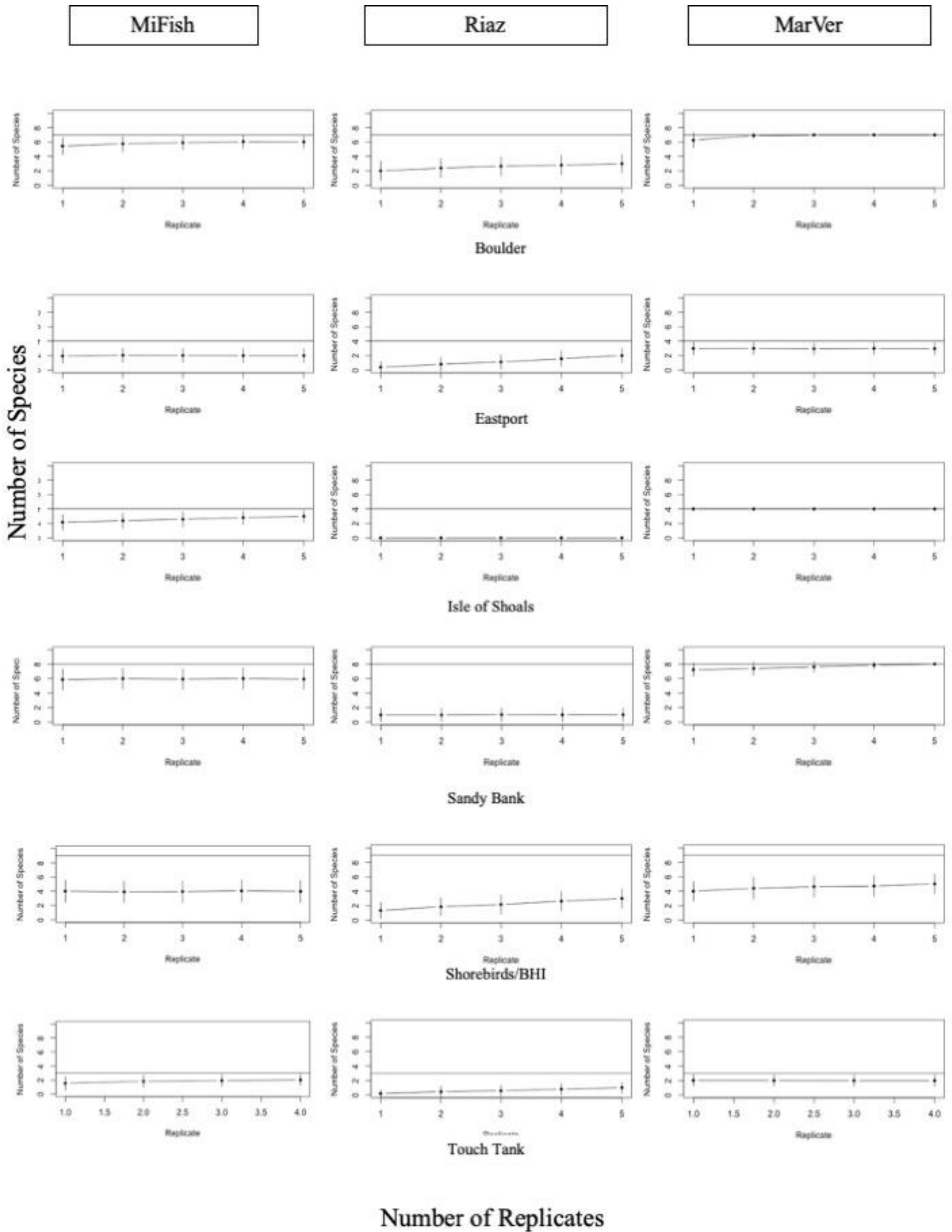

Figure B4*. Traditional rarefaction analysis of number of species revealed with increasing replication in sampling. Error bars represent 95% confidence intervals, x axis represents number of sample replicates,

and y axis represents the number of species in a tank. Horizontal bars represent total census species for that given tank. See above methods and results for Figure B4 for more information.

*Notes on Figure B4:

Traditional rarefaction for vertebrate assays was performed on all assays based on the independent replicate samples collected within each tank. For this analysis, only sequences relating to known census taxa for a given tank were included. Rarefaction effects were determined via bootstrapping individual samples to obtain species lists based on 1-5 samples using 1000 sample randomizations in a for loop in baseR The mean and standard deviation of the number of species per sample was then calculated and plotted to show rarefaction curves. The majority of tanks show that tanks did not necessarily meet the maximum census number (represented by a solid line), but often saturated at the maximum genetically detectable species after 1-3 samples.

Pipeline Assignment	Final Assignment	In Aquarium?
Acadian redfish	Sebastes fasciatus	Y
alewife	Alosa pseudoharengus	N
Ambloplites rupestris	Pomoxis nigromaculatus	N
Atlantic halibut	Hippoglossus hippoglossus	Y
Atlantic herring	Clupea harengus	N
Atlantic salmon	Salmo salar	N
Atlantic silverside	Mendina mendina	N
Atlantic tomcod	Microgadus tomcod	Y
black seabass	Centropristis striata	Y
brook trout	Salvelinus fontinalis	N
capelin	Mallotus villosus	N
cunner	Tautogolabrus adspersus	Y
Cyprinodon macularius	*Cyprinodon variegatus*	Y
Dusky-footed elephant shrew	Homo sapiens	N
grubby sculpin	Myoxocephalus aenaeus	Y
haddock	Melanogrammus aeglefinus	Y
harbor seal	Phoca vitulina	Y
Helicolenus avius	Helicolenus	Y
Hemitremia flammea	Semotilus corporalis	N
longhorn sculpin	Myoxocephalus octodecemspinosus	Y
longspine snipefish	Macroramphosus scolopax	Y
Lycodes ygreknotatus	Zoarces americanus	Y
Macroramphosus sagifue	Macroramphosus scolopax	Y
mouse	Mus musculus	N
mummichog	Fundulus heteroclitus	Y
Myoxocephalus quadricornis	Myoxocephalus quadricornis	N
Phoxinus phoxinus tumensis	Phoxinus phoxinus	N
rock gunnel	Pholis gunnellus	Y
smallmouth blackbass	Micropterus dolomieu	N
spotted ratfish	Hydrolagus collei	Y
spotted seal	Phoca vitulina	Y
striped killifish	Fundulus majalis	Y
striped sea-bass	Morone saxatilis	Y
Tautoga onitis	Tautoga onitis	Y
white sucker	Catostomus commersonii	N
wild boar	Sus scrofa	N
winter flounder	Pseduopleuronectes americanus	Y
Unassigned	uncultured bacterium	N

Table B1. Pipeline assignment of OTUs versus the final assignment based on verification of pipeline

assignment with NCBI and aquarium census list for the MiFish primer set. Presence in the aquarium is also listed.

Pipeline Assignment	Final Assignment	In Aquarium?
antlered sculpin	Myoxocephalus octodecemspinosus	Y
Arctic cod	Gadus morhua	N
Asiatic golden cat	Homo sapiens	N
Atlantic halibut	Hippoglossus hippoglossus	Y
Atlantic herring	Clupea harengus	N
Atlantic silverside	Menidia menidia	N
Atlantic tomcod	Microgadus tomcod	Y
black seabass	Centropristis striata	Y
Bulmer's fruit bat	Homo sapiens	N
capelin	Mallotus villosus	N
Charadrius semipalmatus	Charadrius semipalmatus	Y
Crocidura lasiura	Homo sapiens	N
cunner	Tautogolabrus adspersus	Y
Cyprinodon variegatus variegatus	Cyprinodon variegatus	Y
deepwater redfish	Sebastes	Y
dogs	Canis lupus	N
domestic cow	uncultured eukaryote	N
Dusky-footed elephant shrew	Homo sapiens	N
Eurasian minnow	Phoxinus phoxinus	N
fishing cat	Homo sapiens	N
Flussseeschwalbe	Sterna hirundo	Y
goose-beaked whale	Homo sapiens	N
grey seal	Phoca vitulina	Y
haddock	Melanogrammus aeglefinus	Y
harp seal	Phoca vitulina	Y
Helicolenus avius	Helicolenus dactylopterus	Y
Helicolenus dactylopterus	Helicolenus dactylopterus	Y
Helicolenus hilgendorfi	Helicolenus dactylopterus	Y
Limnodromus sp. NMSU 125X	Calidris	Y
longspine snipefish	Macroramphosus scolopax	Y
Manado fruit bat	Homo sapiens	N
Melonycteris fardoulisi	Homo sapiens	N
mouse	Mus musculus	N
mummichog	Fundulus herteroclitus	Y
Near Eastern wildcat	Felis catus	N
northern searobin	Prionotus carolinus	N
Otter civet	Homo sapiens	N
Phillipine dawn bat	Homo sapiens	N

rock sole	Pseudopleuronectes americanus	Y
round ray	Raja erinacea	Y
short-faced mole	Homo sapiens	N
silver chimaera	Hydrolagus colliei	Y
Spoon-billed sandpiper	Calidris	Y
spotted redshank	Calidris	Y
Steller's sculpin	Myoxocephalus octodecemspinosus	Y
striped killifish	Fundulus majalis	Y
striped sea-bass	Morone saxatilis	Y
Tautoga onitis	Tautoga onitis	Y
thumbprint emperor	bacteria	N
tidepool gunnel	Pholis gunnellus	Y
tiger	Homo sapiens	N
Unassigned	uncultured bacteria	N
viviparous blenny	Zoarces americanus	Y
wild boar	Sus scrofa	N

Table B2. Pipeline assignment of OTUs versus the final assignment based on verification of pipeline assignment with NCBI and aquarium census list for the Riaz primer set. Presence in the aquarium is also listed.

Pipeline Assignment	Final Assignment	In Aquarium?
Acadian redfish	Sebastes fasciatus	Y
Aethaloperca rogaa	Ditylum brightwellii	N
Ancistomus snethlageae	Cerobasis guestfalica	N
antlered sculpin	Myoxocephalus	Y
Arctic cod	Microgadus tomcod	Y
Atlantic cod	Gadus morhua	N
Atlantic halibut	Hippoglossus hippoglossus	Y
Atlantic tomcod	Microgadus tomcod	Y
Australian bull ray	hydroid	N
Australian salmon	bacteria	N
beira	Berkeleya fennica	N
black seabass	Centropristis striata	Y
Bothriechis schlegelii	Homo sapiens	N
capelin	Mallotus villosus	N
Charadrius semipalmatus	Charadrius semipalmatus	Y
Cottus perifretum	Tautoga onitis	Y
creek chub	Semotilus atromaculatus	N
Cyprinodon variegatus variegatus	Cyprinodon variegatus	Y
Danionella mirifica	Cylindrotheca closterium	N
Epigonus angustifrons	Phytophthora	N
European flounder	Pseudopleuronectes americanus	Y
Far Eastern brook lamprey	Macrocystis pyrifera	N
forktongue goby	Phytophthora flexuosa	N
Grant's golden mole	Melitidae	N
gunnel	Pholis gunnellus	Y
haddock	Melanogrammus aeglefinus	Y
harbor seal	Phoca vitulina	Y
Helicolenus hilgendorfi	Helicolenus dactylopterus	Y
Ichthyomyzon gagei	Monomastix sp	N
Indian chameleon	Naineris	N
Indian Ocean caecilian	algae	N
lesser siren	unassigned	N
little skate	Leucoraja erinacea	Y
longspine snipefish	Macroramphosus scolopax	Y
Lycodes ygreknotatus	Zocares americanus	Y
marbled cat	Chordaria flagelliformis	N
mouse	Mus musculus	N
mummichog	Fundulus heteroclitus	Y

northern largemouth blackbass	Mitopus glacialis	N
orange clownfish	Haptoglossa sp	N
prickly sculpin	Myoxocephalus	Y
red deer	spider	N
rockcod	Thalassiosira sp	N
Southeastern four-eyed oppossum	Hemerodromia sp	N
spiky oreo	unassigned	N
spiny-bellied frog	unassigned	N
Spoon-billed sandpiper	Calidris	Y
spotted dogfish	Scyliorhinus	Y
spotted knifejaw	Tautogolabrus adspersus	Y
spotted ratfish	Hydrolagus colliei	Y
striped sea-bass	Morone saxatilis	Y
Sundadanio axelrodi 'red'	Skeletonema cf. pseudocostatum	N
Tamias amoenus canicaudus	unassigned	N
Tsushima leopard cat	Nannochloropsis	N
Unassigned	unassigned	N
Unassigned	Cucumaria frondosa	NA
Unassigned	Hippasteria	NA
Unassigned	Strongylocentrotus droebachiensis	NA
Unassigned	Arbacia punctulata	NA
Unassigned	Buccinum undatum	NA
Unassigned	Henricia	NA
Unassigned	Oscarella pearsei	NA
Unassigned	Solaster endeca	NA
Unassigned	Halichondria panicea	NA
Unassigned	bacteria	N
Unassigned	Asterias rubens	NA
Unassigned	Crossaster papposus	NA
Unassigned	Megathura crenulata	NA
varied sittella	unassigned	N
western European hedgehog	Triparma	N
white sucker	Catostomus commersonii	N
whitecheek monocle bream	Chordariaceae	N
wild boar	Sus scrofa	N

Table B3. Pipeline assignment of OTUs versus the final assignment based on verification of pipeline assignment with NCBI and aquarium census list for the MarVer primer set. Presence in the aquarium is also listed.

Taxon	EXT.BLANK	LAB.BLANK	Summed Reads	Max Reads
Alosa pseudoharengus	38867	0	38867	38867
Catostomus commersonii	0	0	0	0
Centropristis striata	9	22	31	22
Clupea harengus	0	0	0	0
Cyprinodon variegatus	0	0	0	0
Fundulus herteroclitus	0	0	0	0
Fundulus majalis	0	0	0	0
Helicolenus	156	102	258	156
Hippoglossus hippoglossus	0	0	0	0
Homo sapiens	3540	0	3540	3540
Hydrolagus collei	0	0	0	0
Macroramphosus scolopax	13	25	38	25
Mallotus villosus	0	0	0	0
Melanogrammus aeglefinus	0	38	38	38
Mendina mendina	0	0	0	0
Microgadus tomcod	0	0	0	0
Micropterus dolomieu	0	0	0	0
Morone saxatilis	0	0	0	0
Mus musculus	0	0	0	0
Myoxocephalus aenaeus	0	0	0	0
Myoxocephalus octodecemspinosus	0	0	0	0
Myoxocephalus quadricornis	0	0	0	0
Phoca vitulina	25	0	25	25
Pholis gunnellus	10	0	10	10
Phoxinus phoxinus	0	0	0	0
Pomoxis nigromaculatus	0	3489	3489	3489
Pseduopleuronectes americanus	0	0	0	0
Salmo salar	25051	0	25051	25051
Salvelinus fontinalis	97	80676	80773	80676
Sebastes fasciatus	39	35	74	39
Semotilus corporalis	0	0	0	0
Sus scrofa	0	0	0	0
Tautoga onitis	0	0	0	0
Tautogolabrus adspersus	0	0	0	0
uncultured bacterium	190126	9	190135	190126
Zoarces americanus	0	0	0	0

Table B4. MiFish blank reads. Max reads refers to the maximum possible reads for a given taxon that was found in the blanks, and is the value used to subtract from all other OTUs.

Taxon	EXTBLANK	LABBLANKJ	SBFB	SSFB	Summed Reads	Max Reads
bacteria	14	0	0	0	14	14
Calidris	0	0	0	0	0	0
Canis lupus	0	0	0	0	0	0
Centropristis striata	0	0	0	0	0	0
Charadrius semipalmatus	0	0	0	0	0	0
Clupea harengus	0	0	0	0	0	0
Cyprinodon variegatus	0	110	0	147	257	147
Felis catus	0	0	0	0	0	0
Fundulus herteroclitus	0	0	0	0	0	0
Fundulus majalis	12	203	25	0	240	203
Gadus morhua	0	0	0	0	0	0
Helicolenus dactylopterus	3953	26	17	155	4151	3953
Hippoglossus hippoglossus	0	1327	0	0	1327	1327
Homo sapiens	21345	896	102393	5626	130260	102393
Hydrolagus colliei	0	0	0	0	0	0
Macroramphosus scolopax	0	0	0	0	0	0
Mallotus villosus	0	0	0	0	0	0
Melanogrammus aeglefinus	0	0	0	0	0	0
Menidia menidia	189	0	0	0	189	189
Microgadus tomcod	0	0	0	0	0	0
Morone saxatilis	9	805	0	539	1353	805
Mus musculus	12498	0	0	5486	17984	12498
Myoxocephalus octodecemspinosus	6	47156	390	35255	82807	47156
Phoca vitulina	11	9	42020	21835	63875	42020
Pholis gunnellus	52	211	0	11	274	211
Phoxinus phoxinus	0	0	0	0	0	0
Prionotus carolinus	0	0	0	0	0	0
Pseudopleuronectes americanus	0	0	0	0	0	0
Raja erinacea	0	0	0	0	0	0
Sebastes	0	8	70	0	78	70
Sterna hirundo	3323	0	38	0	3361	3323
Sus scrofa	756	0	0	0	756	756

Tautoga onitis	0	15	0	96	111	96
Tautogolabrus adspersus	0	324	0	0	324	324
uncultured bacteria	168285	40375	90499	114499	413658	168285
uncultured eukaryote	0	0	0	0	0	0
Zoarces americanus	0	0	0	0	0	0

Table B5. Riaz blank reads. Max reads refers to the maximum possible reads for a given taxon that was found in the blanks, and is the value used to subtract from all other OTUs.

Removed Taxa
Alosa pseudoharengus
Pomoxis nigromaculatus
Clupea harengus
Salmo salar
Mendina mendina
Salvelinus fontinalis
Mallotus villosus
Homo sapiens
Semotilus corporalis
Mus musculus
Myoxocephalus quadricornis
Phoxinus phoxinus
Micropterus dolomieu
Catostomus commersonii
Sus scrofa
uncultured bacterium
Gadus morhua
Menidia menidia
Canis lupus
uncultured eukaryote
Felis catus
Prionotus carolinus
bacteria
uncultured bacteria
Ditylum brightwellii
Cerobasis guestfalica
hydroid
Berkeleya fennica
Semotilus atromaculatus
Cylindrotheca closterium
Phytophthora
Macrocystis pyrifera
Phytophthora flexuosa
Melitidae
Monomastix sp
Naineris
algae
unassigned
Chordaria flagelliformis

Mitopus glacialis
Haptoglossa sp
spider
Thalassiosira sp
Hemerodromia sp
Skeletonema cf. pseudocostatum
Nannochloropsis
Cucumaria frondosa
Hippasteria
Strongylocentrotus droebachiensis
Arbacia punctulata
Buccinum undatum
Henricia
Oscarella pearsei
Solaster endeca
Halichondria panicea
Asterias rubens
Crossaster papposus
Megathura crenulata
Triparma
Chordariaceae

Table B6. Removed taxa from analyses for vertebrate datasets. This list includes all possible non-target taxa for all 3 primer sets used in this study.

Species	Common Name	MiFish	Riaz	MarVer
Calidris alba	Sanderling	N	Y	Y
Calidris minutilla	Least sandpiper	N	Y	Y
Calidris pusilla	Semipalmated sandpiper	N	Y	Y
Centropristis striata	Black seabass	N	Y	Y
Charadrius melodus	Piping plover	NA	NA	NA
Charadrius semipalmatus	Semipalmated plover	Y	Y	Y
Cyprinodon variegatus	Sheepshead minnow	Y	Y	Y
Fundulus heteroclitus	Mummichog	N	Y	Y
Fundulus majalis	Striped killifish	N	Y	NA
Helicolenus dactylopterus	Blackbelly Rosefish	Y	Y	Y
Hippoglossus hippoglossus	Atlantic halibut	Y	Y	Y
Hydrolagus colliei	Spotted ratfish	N	Y	Y
Macroramphosus scolopax	Longspine snipefish	Y	Y	Y
Melanogrammus aeglefinus	Haddock	Y	y	Y
Microgadus tomcod	Atlantic tomcod	N	Y	Y
Morone saxatilis	Striped seabass	Y	Y	Y
Myoxocephalus aenaeus	Grubby sculpin	Y	Y	NA
Myoxocephalus octodecemspinosus	Longhorn sculpin	Y	Y	NA
Myoxocephalus scorpius	Shorthorn sculpin	Y	Y	Y
Phoca vitulina	Harbor seal	Y	Y	Y
Pholis gunnellus	Rock gunnel	Y	Y	Y
Pseudopleuronectes americanus	Winter flounder	Y	Y	Y
Leucoraja erinacea	Little skate	Y	Y	Y
Scyliorhinus retifer	Chain cat shark	NA	NA	NA
Sebastes fasciatus	Acadian Redfish	Y	Y	Y
Sterna hirundo	Common tern	Y	Y	Y
Tautoga onitis	Tautog	Y	y	Y
Tautogolabrus adspersus	Cunner	Y	Y	Y
Ulvaria subbifurcata	Radiated shanny	N	Y	NA
Zoarces americanus	Ocean eelpout	NA	NA	Y

Table B7. In-silico primer binding expectations for vertebrate assays. Y= represented in a database and primers will bind with 3 or less basepair mismatches; N= will not bind, or has greater than 3 bp mismatches; NA= no reference sequence available in the databases used

Appendix C: Invertebrate supplemental material

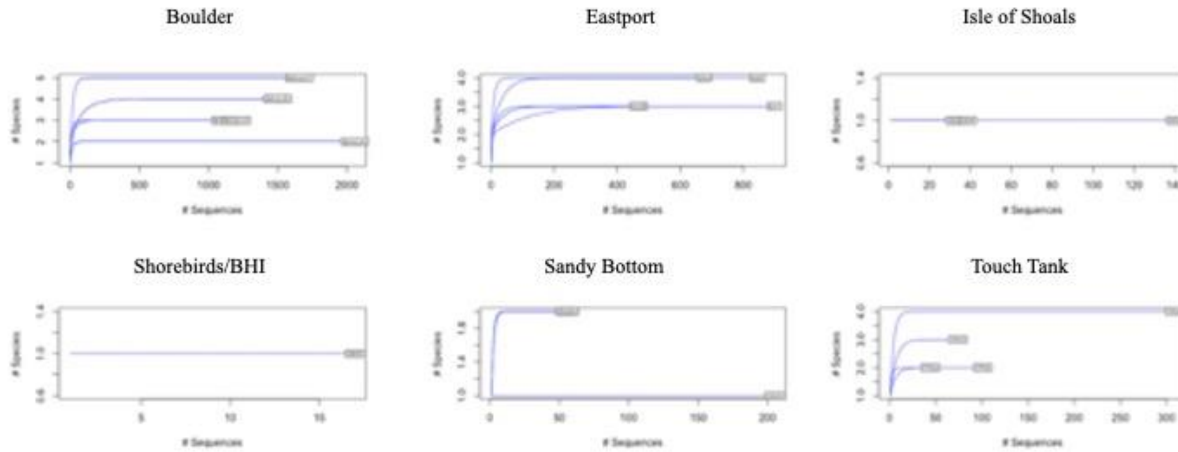

Figure C1*. Rarefaction of sequencing reads versus number of species within each tank per each sampling replicate. The y axis represents number of species, while the x axis represents number of sequences.

Notes on Figure C1:

Read rarefaction was done using the vegan package in R (version 4.1.1). This was done by randomly sampling each ASV read counts to make a plot comparing number of sequences in a sample to number of species in a sample. The majority of tanks across all three vertebrate assays show saturation of number of species within samples and within tanks. Some tanks also did not have sampling replicates amplify, which results in differing amounts of boxes represented in the graph. Results indicate that in most cases for the vertebrate assays used in this study, read depth in sequencing was sufficient to identify all genetically detectable species.

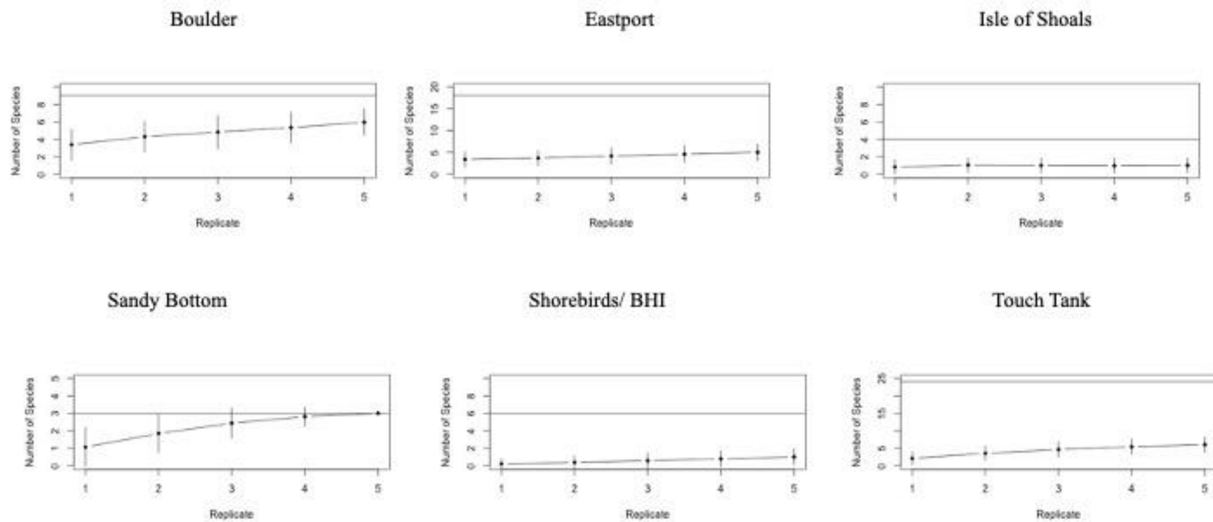

Figure C2*. Rarefaction of sample replication versus number of species within each tank. The y axis represents number of species, and the x axis the number of sampling replicates. Error bars indicate 95% confidence intervals. Horizontal line indicates maximum species number per census data in the tank.

Notes on Figure C2:

Traditional rarefaction for vertebrate assays was performed on all assays based on the independent replicate samples collected within each tank. For this analysis, only sequences relating to known census taxa for a given tank were included. Rarefaction effects were determined via bootstrapping individual samples to obtain species lists based on 1-5 samples using 1000 sample randomizations in a for loop in baseR The mean and standard deviation of the number of species per sample was then calculated and plotted to show rarefaction curves. The majority of tanks show that tanks did not necessarily meet the maximum census number (represented by a solid line), but often saturated at the maximum genetically detectable species after 1-3 samples, with the exception of the Boulder tank which did not saturate at all.

Ref Database Assignment	Final Assignment
Asterias forbesi	Asterias forbesi
Asterias rubens	Asterias rubens
Buccinum undatum	Buccinum
Buccinum undatum	Buccinum undatum
Crossaster papposus	Crossaster papposus
Cucumaria frondosa	Cucumaria frondosa
Homarus americanus	Homarus americanus
Lunatia heros	Euspira heros
Metridium senile	Metridium senile
Mytilus edulis	Mytilus edulis
Ophiopholis aculeata	Ophiopholis aculeata
Pagurus acadianus	Pagurus acadianus
Solaster endeca	Solaster endeca
Strongylocentrotus droebachiensis	Strongylocentrotus droebachiensis
Unassigned	Anthozoa
Unassigned	Amphipholis squamata
Unassigned	Malacostraca
Unassigned	Animalia
Unassigned	Polychaeta
Unassigned	Arachnida
Unassigned	Branchiopoda
Unassigned	Cephalopoda
Unassigned	Copepoda
Unassigned	Diplopoda
Unassigned	Fungi
Unassigned	Henricia
Unassigned	Hydrozoa
Unassigned	Insecta
Unassigned	Gastropoda
Unassigned	Littorina
Unassigned	Hoplonemertea
Unassigned	Palaeonemertea
Unassigned	Heteronemertea
Unassigned	Nudibranchia
Unassigned	Phoca vitulina
Unassigned	Demospongiae
Unassigned	Homosclerophorida

Unassigned	Protozoa
Unassigned	Psuedopleuronectes americanus
Unassigned	Pycnogonida
Unassigned	Scyphozoa
Unassigned	Staurozoa
Unassigned	Unassigned

Table C1. COI taxonomic assignment from the pipeline, and final assignment based on quality checking from the original database, aquarium list and NCBI/BOLD databases.

Pipeline Assingment	Final Assignment
d__Archaea; p__Crenarchaeota; c__Nitrososphaeria; o__Nitrosopumilales; f__Nitrosopumilaceae; g__Candidatus_Nitrosopumilus; s__uncultured_crenarchaeote	archaea
d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__SAR11_clade; f__Clade_I; g__Clade_Ia; s__uncultured_bacterium	bacteria
d__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__SAR11_clade; f__Clade_II; g__Clade_II; s__uncultured_Alphaproteobacteria	bacteria
d__Eukaryota	eukaryota
d__Eukaryota; p__Annelida; c__Polychaeta; o__Sabellida; f__Sabellida; g__Sabellida	polychaeta
d__Eukaryota; p__Annelida; c__Polychaeta; o__Sabellida; f__Sabellida; g__Sabellida; s__Spirorbis_bifurcatus	polychaeta
d__Eukaryota; p__Apicomplexa; c__Conoidasida; o__Coccidia; f__Eimeriorina; g__Rhytidocystis; s__Rhytidocystis_cyamus	apicomplexa
d__Eukaryota; p__Apicomplexa; c__Conoidasida; o__Coccidia; f__Eimeriorina; g__Rhytidocystis; s__Rhytidocystis_polygordiae	apicomplexa
d__Eukaryota; p__Ascomycota; c__Eurotiomycetes; o__Chaetothyriales; f__Herpotrichiellaceae; g__Exophiala	eukaryota
d__Eukaryota; p__Ascomycota; c__Eurotiomycetes; o__Chaetothyriales; f__Herpotrichiellaceae; g__Exophiala	cercozoa
d__Eukaryota; p__Cercozoa	cercozoa
d__Eukaryota; p__Cercozoa; c__Chlorarachniophyta; o__Chlorarachniophyta; f__Chlorarachniophyta; g__uncultured; s__uncultured_eukaryote	cercozoa
d__Eukaryota; p__Cercozoa; c__Thecofilosea; o__Thecofilosea; f__Thecofilosea; g__Thecofilosea; s__uncultured_marine	cercozoa
d__Eukaryota; p__Cercozoa; c__Vampyrellidae; o__Vampyrellidae; f__Vampyrellidae; g__uncultured; s__uncultured_alveolate	cercozoa
d__Eukaryota; p__Chlorophyta; c__Ulvophyceae; o__Ulvophyceae; f__Ulvophyceae; g__Ulvophyceae; s__Lychaete_pellucidoidea	chlorophyta
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Conthreep	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Conthreep; f__Oligohymenophorea	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Conthreep; f__Oligohymenophorea; g__Hyalophysa	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Conthreep; f__Phyllopharyngea; g__Acineta	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Conthreep; f__Phyllopharyngea; g__Suctoria; s__uncultured_eukaryote	ciliophora

d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Arcuospathidium; s__uncultured_ciliate	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Chaenea; s__Chaenea_sp.	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Chaenea; s__Chaenea_vorax	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Chaenea; s__uncultured_litostomatid	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Haptoria	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Haptoria; s__Litonotus_gracilis	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Haptoria; s__Litonotus_pictus	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Hemiophrys; s__Hemiophrys_procera	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Hemiophrys; s__uncultured_eukaryote	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Litonotus; s__Litonotus_paracygnus	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Loxophyllum	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Loxophyllum; s__Loxophyllum_perihoplophorum	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__Trachelotractus; s__uncultured_eukaryote	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Haptoria; g__uncultured; s__uncultured_eukaryote	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Mesodiniidae	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Mesodiniidae; g__FV18-2A2; s__uncultured_marine	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Mesodiniidae; g__Mesodiniidae	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Mesodiniidae; g__Mesodiniidae; s__uncultured_freshwater	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Mesodiniidae; g__Mesodiniidae; s__uncultured_marine	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Mesodiniidae; g__Myrionecta	ciliophora

d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Litostomatea; f__Mesodiniidae; g__uncultured	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Spirotrichea	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Spirotrichea; f__Euplotia; g__Aspidisca; s__Aspidisca_sp.	ciliophora
d__Eukaryota; p__Ciliophora; c__Intramacronucleata; o__Spirotrichea; f__Hypotrichia; g__Protogastrostyla; s__Protogastrostyla_pulchra	ciliophora
d__Eukaryota; p__Ciliophora; c__Postciliodesmatophora; o__Karyorelictea; f__Karyorelictea	ciliophora
d__Eukaryota; p__Ciliophora; c__Postciliodesmatophora; o__Karyorelictea; f__Karyorelictea; g__Karyorelictea; s__Wilbertomorpha_colpoda	ciliophora
d__Eukaryota; p__Ciliophora; c__Postciliodesmatophora; o__Karyorelictea; f__Karyorelictea; g__Kovalevaia; s__uncultured_eukaryote	ciliophora
d__Eukaryota; p__Ciliophora; c__Postciliodesmatophora; o__Karyorelictea; f__Karyorelictea; g__Tracheloraphis	ciliophora
d__Eukaryota; p__Ciliophora; c__Postciliodesmatophora; o__Karyorelictea; f__Karyorelictea; g__Tracheloraphis; s__uncultured_eukaryote	ciliophora
d__Eukaryota; p__Ciliophora; c__Postciliodesmatophora; o__Karyorelictea; f__Karyorelictea; g__Tracheloraphis; s__uncultured_Trachelocercidae	ciliophora
d__Eukaryota; p__Cnidaria; c__Anthozoa; o__Actiniaria; f__Actiniaria; g__Actiniaria	Anthozoa
d__Eukaryota; p__Diatomea	Diatomea
d__Eukaryota; p__Diatomea; c__Bacillariophyceae; o__Bacillariophyceae; f__Bacillariophyceae; g__Pleurosigma; s__uncultured_marine	Diatomea
d__Eukaryota; p__Diatomea; c__Mediophyceae; o__Mediophyceae; f__Mediophyceae; g__Thalassiosira	Diatomea
d__Eukaryota; p__Dinoflagellata	Dinoflagellata
d__Eukaryota; p__Dinoflagellata; c__Dinophyceae	Dinoflagellata
d__Eukaryota; p__Dinoflagellata; c__Dinophyceae; o__Gymnodiniphycidae	Dinoflagellata
d__Eukaryota; p__Echinodermata; c__Holothuroidea; o__Holothuroidea; f__Holothuroidea; g__Holothuroidea	HOLOTHUROIDEA
d__Eukaryota; p__Eukaryota; c__Eukaryota; o__Eukaryota; f__Eukaryota; g__Eukaryota; s__Ciliophrys_infusionum	eukaryota
d__Eukaryota; p__Gastrotricha; c__Gastrotricha; o__Chaetonotida; f__Chaetonotida; g__Chaetonotida	gastrotricha
d__Eukaryota; p__Gastrotricha; c__Gastrotricha; o__Chaetonotida; f__Chaetonotida; g__Chaetonotida	eukaryota

d__Eukaryota; p__Holozoa; c__Ichthyosporea; o__Abeoformidae; f__Abeoformidae; g__Abeoformidae; s__uncultured_Choanoflagellida	eukaryota
d__Eukaryota; p__Kathablepharidae; c__Kathablepharidae; o__Kathablepharidae; f__Kathablepharidae; g__uncultured	eukaryota
d__Eukaryota; p__Microsporidia; c__Microsporidia; o__Microsporidia; f__Microsporidia; g__Microsporidia; s__Pseudonosema_cristatellae	eukaryota
d__Eukaryota; p__Nemertea; c__Anopla; o__Heteronemertea; f__Heteronemertea; g__Heteronemertea	Nemertea
d__Eukaryota; p__Peronosporomycetes; c__Peronosporomycetes; o__Peronosporomycetes; f__Peronosporomycetes	eukaryota
d__Eukaryota; p__Porifera; c__Calcarea; o__Clathrinida; f__Clathrinida; g__Clathrinida	Porifera
d__Eukaryota; p__Porifera; c__Calcarea; o__Clathrinida; f__Clathrinida; g__Clathrinida; s__Clathrina_blanca	Porifera
d__Eukaryota; p__Porifera; c__Calcarea; o__Leucosolenida; f__Leucosolenida; g__Leucosolenida	Porifera
d__Eukaryota; p__Porifera; c__Demospongiae; o__Suberitida; f__Suberitida; g__Suberitida	Porifera
d__Eukaryota; p__Protalveolata; c__Syndiniales; o__Syndiniales; f__Syndiniales_Group_I; g__Syndiniales_Group_I	eukaryota
d__Eukaryota; p__Retaria; c__Foraminifera; o__Rotaliida	eukaryota
d__Eukaryota; p__Tunicata; c__Ascidiacea; o__Stolidobranchia; f__Stolidobranchia; g__Stolidobranchia; s__Molgula_citrina	Ascidiacea
d__Eukaryota; p__Tunicata; c__Ascidiacea; o__Stolidobranchia; f__Stolidobranchia; g__Stolidobranchia; s__Molgula_complanata	Ascidiacea
d__Eukaryota; p__Vertebrata	Vertebrata
d__Eukaryota; p__Vertebrata; c__Actinopterygii; o__Teleostei; f__Teleostei; g__Teleostei	Teleostei
Unassigned	unassigned

Table C2. 18S taxonomic assignment from the pipeline, and final assignment based on quality checking from the original database, aquarium list and NCBI/BOLD databases.

Taxon	Max Reads
Anthozoa	0
apicomplexa	0
archaea	0
Ascidiacea	0
bacteria	0
cercozoa	3
chlorophyta	0
ciliophora	0
Diatomea	0
Dinoflagellata	0
eukaryota	314
gastrotricha	0
HOLOTHUROIDEA	0
Nemertea	0
polychaeta	0
Porifera	0
Teleostei	0
unassigned	34
Vertebrata	0

Table C3. 18S blank maximum reads and the associated taxa.

Removed Taxa
Actinaria
Amphiuridae
Animalia
Annelida
Arachnida
Branchiopoda
Cephalopoda
Copepoda
Diplopoda
Fungi
Hydrozoa
Insecta
Lepetellida
Malacostraca
Mollusca
Nemertea
Nudibranchia
Pachychilidae
Phoca vitulina
Porifera
Protozoa
Psuedopleuronectes americanus
Pycnogonida
Scyphozoa
Staurozoa
Unassigned

Table C4. Removed taxa from the COI dataset for further analysis.

Enclosure Name	Species	Common Name	Class Name	Population Count	Lowest taxonomic rank detection
Boulder Reef Community (22)	Metridium senile	Clonal plumose anemone	ANTHOZOA	25	Species
Boulder Reef Community (22)	Urticina felina	Northern red anemone	ANTHOZOA	22	No detect
Boulder Reef Community (22)	Homarus americanus	American lobster	CRUSTACEA	1	No detect
Boulder Reef Community (22)	Buccinum undatum	Common northern whelk	GASTROPODA	170	Species
Boulder Reef Community (22)	Neptunea lyrata	Common northwest neptune	GASTROPODA	15	No detect
Boulder Reef Community (22)	Cucumaria frondosa	Orange-footed sea cucumber	HOLOTHUROIDEA	51	Species
Boulder Reef Community (22)	Psolus fabricii	Scarlet sea cucumber	HOLOTHUROIDEA	2	No detect
Boulder Reef Community (22)	Asterias rubens	Northern sea star	STELLEROIDEA	9	Species

Boulder Reef Community (22)	Crossaster papposus	Spiny Sun Star	STELLEROIDEA	8	Species
Boulder Reef Community (22)	Ophiopholis aculeata	Daisy brittlestar	STELLEROIDEA	17	Species
Eastport (25)	Gersemia rubiformis	Sea strawberry	ANTHOZOA	1	No detect
Eastport (25)	Metridium senile	Clonal plumose anemone	ANTHOZOA	52	Species
Eastport (25)	Urticina felina	Northern red anemone	ANTHOZOA	15	No detect
Eastport (25)	Ciona intestinalis	Sea vase	ASCIDIACEA	2	No detect
Eastport (25)	Modiolus modiolus	Northern horse mussel	BIVALVIA	16	No detect
Eastport (25)	Mytilus edulis	blue mussel	BIVALVIA	9	No detect
Eastport (25)	Semibalanus balanoides	Northern rock barnacle	CRUSTACEA	6	No detect
Eastport (25)	Pagurus acadianus	Acadian hermit crab	CRUSTACEA	18	Species
Eastport (25)	Pagurus arcuatus	Hairy hermit crab	CRUSTACEA	6	No detect
Eastport (25)	Strongylocentrotus droebachiensis	Green sea urchin	ECHINOIDEA	7	Species
Eastport (25)	Buccinum undatum	Common northern whelk	GASTROPODA	6	Species

Eastport (25)	Chiridota laevis	Silky cucumber	HOLOTHUROIDEA	2	No detect
Eastport (25)	Cucumaria frondosa	Orange-footed sea cucumber	HOLOTHUROIDEA	2	No detect
Eastport (25)	Psolus fabricii	Scarlet sea cucumber	HOLOTHUROIDEA	1	No detect
Eastport (25)	Asterias rubens	Northern sea star	STELLEROIDEA	1	No detect
Eastport (25)	Crossaster papposus	Spiny Sun Star	STELLEROIDEA	3	No detect
Eastport (25)	Henricia sanguinolenta	Atlantic blood sea star	STELLEROIDEA	10	Genus (sp?)
Eastport (25)	Hippasteria phrygiana	Horse sea star	STELLEROIDEA	1	No detect
Eastport (25)	Ophiopholis aculeata	Daisy brittlestar	STELLEROIDEA	31	Species
EOS TP	Stomphia didemon	Cowardly anemone	ANTHOZOA	4	No detect
EOS TP	Urticina felina	Northern red anemone	ANTHOZOA	12	No detect
EOS TP	Crassostrea virginica	Atlantic oyster	BIVALVIA	12	No detect
EOS TP	Mercenaria mercenaria	Quahog	BIVALVIA	20	No detect
EOS TP	Mytilus edulis	blue mussel	BIVALVIA	70	Species
EOS TP	Placopecten magellanicus	Atlantic deep-sea scallop	BIVALVIA	1	No detect
EOS TP	Cancer borealis	Jonah crab	CRUSTACEA	1	No detect

EOS TP	Libinia emarginata	Short-clawed spider crab	CRUSTACEA	3	No detect
EOS TP	Pagurus acadianus	Acadian hermit crab	CRUSTACEA	14	No detect
EOS TP	Pagurus arcuatus	Hairy hermit crab	CRUSTACEA	3	No detect
EOS TP	Pagurus longicarpus	Long-armed hermit crab	CRUSTACEA	47	No detect
EOS TP	Strongylocentrotus droebachiensis	Green sea urchin	ECHINOIDEA	53	Species
EOS TP	Buccinum undatum	Common northern whelk	GASTROPODA	3	No detect
EOS TP	Diodora aspera	Rough keyhole limpet	GASTROPODA	151	No detect
EOS TP	Euspira heros	Northern moon snail	GASTROPODA	2	No detect
EOS TP	Littorina littorea	Common periwinkle	GASTROPODA	259	No detect
EOS TP	Neptunea lyrata	Common northwest neptune	GASTROPODA	14	No detect
EOS TP	Cucumaria frondosa	Orange-footed sea cucumber	HOLOTHUROIDEA	4	No detect
EOS TP	Limulus polyphemus	Horseshoe crab	MEROSTOMATA	2	No detect
EOS TP	Asterias forbesi	Bay sea star	STELLEROIDEA	5	Species
EOS TP	Asterias rubens	Northern sea star	STELLEROIDEA	81	Species

EOS TP	Henricia sanguinolenta	Atlantic blood sea star	STELLEROIDEA	4	No detect
EOS TP	Ophiopholis aculeata	Daisy brittlestar	STELLEROIDEA	6	Species
EOS TP Lower	Homarus americanus	American lobster	CRUSTACEA	1	Species
EOS TP Upper	Libinia dubia	Southern spider crab	CRUSTACEA	1	No detect
Goosefish (24)	Modiolus modiolus	Northern horse mussel	BIVALVIA	1	No detect
Goosefish (24)	Homarus americanus	American lobster	CRUSTACEA	2	Species
Goosefish (24)	Euspira heros	Northern moon snail	GASTROPODA	3	No detect
Goosefish (24)	Henricia sanguinolenta	Atlantic blood sea star	STELLEROIDEA	37	No detect
Sandy Bottom Community (23)	Strongylocentrotus droebachiensis	Green sea urchin	ECHINOIDEA	11	Species
Sandy Bottom Community (23)	Euspira heros	Northern moon snail	GASTROPODA	3	Species
Sandy Bottom Community (23)	Solaster endeca	Purple sun star	STELLEROIDEA	14	Species
Shorebirds	Mytilus edulis	blue mussel	BIVALVIA	40	No detect
Shorebirds	Cancer borealis	Jonah crab	CRUSTACEA	2	No detect

Shorebirds	Hyas coarctatus	Toad crab	CRUSTACEA	1	No detect
Shorebirds	Libinia dubia	Southern spider crab	CRUSTACEA	2	No detect
Shorebirds	Pagurus longicarpus	Long-armed hermit crab	CRUSTACEA	1	No detect
Shorebirds	Arbacia punctulata	Purple sea urchin	ECHINOIDEA	11	No detect
Shorebirds	Strongylocentrotus droebachiensis	Green sea urchin	ECHINOIDEA	2	Species

Table C5. Taxonomic detection of each species within a given tank, as well as population abundance measures.

BIOGRAPHY OF THE AUTHOR

Samantha Silverbrand was born in Bourne, Massachusetts in 1997. She graduated high school in 2015 from Sturgis Charter Public School in Hyannis, MA and then began her bachelor's degree at the University of Maine. In 2019, she graduated from the University of Maine with a degree in Marine Science. In September of 2019, she returned to UMaine to pursue a graduate degree. Samantha is a candidate for the Master of Science in Marine Biology from the University of Maine in December 2021.