Summer 8-16-2024

Applications of Environmental DNA Metabarcoding in Stream Biomonitoring

Beth Y. Davis
University of Maine, bethy.davis4@gmail.com

APPLICATIONS OF ENVIRONMENTAL DNA METABARCODING IN STREAM BIOMONITORING

By:

Beth Yima Davis

B.S. Oregon State University, 2020

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

(In Ecology and Environmental Sciences)

The Graduate School

The University of Maine

August 2024

Advisory Committee:

Hamish Greig, Associate Professor of Stream Ecology, University of Maine; Advisor

Andrew Rominger, Cooperating Assistant Professor, University of Maine; Assistant Professor, University of Hawai'i at Manoa

Erin Grey, Assistant Professor of Aquatic Genetics, University of Maine

LAND ACKNOWLEDGMENT

All sample collection for this thesis and all laboratory work was performed in the homelands of the Penobscot Nation, where issues of water and territorial rights, data sovereignty, and encroachment upon sacred sites are ongoing. The Penobscot Nation is connected to the other Wabanaki Nations– the Passamaquoddy Tribe, Houlton Band of Maliseet Indians, and Mi'kmaq Nation– through kinship, alliances, and diplomacy. The Wabanaki Nations are distinct, sovereign, legal, and political entities with their own powers and rights of self-governance and self-determination. The legacy and continuing actions of systematic erasure and exclusion of Indigenous peoples manifests also in academic exclusion of Indigenous researchers and partnerships with Indigenous communities. While individual researchers have limited power to overcome centuries of structural inequality, we have a responsibility to address and minimize harm wherever possible. All data collected for this thesis are associated with a Biocultural Notice to acknowledge the cultural connections and rights of the Penobscot Nation to define the use and disclosure of information and sequences generated from the genetic material collected from their traditional lands, waters, and territories. The Local Contexts Projects associated with these data are available at the following for Chapter 2 and Chapter 3.

APPLICATION OF ENVIRONMENTAL DNA METABARCODING IN STREAM BIOMONITORING

By: Beth Yima Davis

Thesis Advisor: Dr. Hamish Greig

An Abstract of the THESIS
Presented in Partial Fulfillment of the Requirements for the
Degree of Master of Science
(in Ecology and Environmental Sciences)
August 2024

Advancements in genetic technology and processing allows for the presence of loose genetic material in the environment to become a resource, capable of assisting habitat and wildlife management efforts by providing information about organisms in a region without having to disturb or disrupt the organisms and environment. This use of environmental DNA has gained traction across biomes, with researchers continuing to test extraction and processing of DNA from various environmental media. However, the high variability in media quality, characteristics, and taxonomic knowledge means that the tested capabilities of eDNA vary wildly depending on the application and species of interest. In this thesis, I focus on the use of eDNA metabarcoding in freshwater streams in Maine, examining the ability and existing libraries of two genetic loci to identify Maine fish and macroinvertebrate species. eDNA results are compared against a traditional specimen-based surveying method utilized by the Maine Department of Environmental Protection and the Penobscot Nation's Department of Natural Resources, and over time to monitor the success of stream restoration initiatives. While eDNA samples successfully detected fish and invertebrate species in both datasets, no strong correlation

was found between benthic macroinvertebrate abundance counts and detected sequence variants. Furthermore, eDNA detection led to highly different community survey results than the specimen-based survey method, and limitations of available reference sequences indicate a strong need for localized references for future eDNA work. While eDNA was able to identify ASVs at a higher clarity than the specimen-based survey method, only 4 taxonomic families were shared between the survey method categorization and eDNA detection. However, eDNA was successful when applied to a broader range of taxa for presence-absence detection and community composition detection, and found that stream communities did change significantly based on installment of large wood addition projects.

ACKNOWLEDGEMENTS

I'd first like to express my appreciation to Dr. Andy Rominger, for first welcoming me into the EcoEvoMatics Lab, and for providing much needed support and reassurance throughout my time at UMaine. To Dr. Hamish Greig, I can't express my gratitude enough for your patience and welcoming me into your lab despite the chaos, and for your level-headedness and ability to keep me even somewhat connected to reality. Thank you both for years of support, patience, and encouragement as this thesis has evolved.

To my other committee members, both past and present, I thank you for your time and generosity in guiding me through the development of my research, even though not all of your inputs were integrated or shown in this thesis.

Dr. Erin Grey has been welcoming and willing to share her knowledge and resources at every step, and her experience at varied eDNA techniques has been invaluable. Thank you truly for always being willing to help, and allowing me to claim so many late night lab hours.

I thank Dr. Darren Ranco for humoring and guiding my inexperienced enthusiasm in working with communities, and introducing me to people who made all of my stress and chaos over the last three years feel worth it. I've learned a lot on how science should be done, and I hope to continue learning and to do justice by the communities I work for in the future.

I thank Dr. Joseph Zydlewski, the only person to make me wish I enjoyed studying fish, for his patience and guidance when I was first stumbling through my research design and adjusting to Maine's streams. His level-headedness was critical in helping me not get too overwhelmed.

In my time in Maine, being able to get to know and work with so many amazing people made the dark winter worth it. Thank you to everyone in the Penobscot Indian Nation

Department of Natural Resources, it was a true delight being able to do fieldwork with you, and your patience and welcome can not be overstated. To Sarah Nelson and Steve Tatko at the Appalachian Mountain Club, as well as everyone on the summer field teams, thank you for working with me through the growing pains and for your confidence in my work. Your passion for the rivers gave me hope and taught me to love the streams, no matter how many black flies were around.

I would not have been able to do the laboratory work, and surely would've lost even more sanity, without Geneva York. Her laboratory, emotional, and troubleshooting support enabled both me and this work to keep going, and I'll miss our late night scheduling conflicts.

Also critical in understanding and finishing laboratory work was Kylie Holt. Thank you for lending me your experience and advice, professionalism, and being willing to hike through the woods with me with bricks.

Thank you especially to Ben Tupper, who has taught me so many ways to be a better coder that I immediately spent a month implementing your tips, and I cannot express how much I appreciate your experience and sense of humor.

I'd like to also thank Heather Richard, Jennifer Smith-Mayo, Julia Sunnarborg, Emily Lancaster, Brandon Henry, and the rest of my fellow Maine-eDNA graduate students for emotional support, laboratory guidance, fieldwork tips, and your friendship.

Thank you to the many talented and passionate individuals who I've met on and through Maine-eDNA, in particular Beth Campbell, Dan Timmerman, Laura Jackson, Melissa Kimble, Sarah Sparks, Ciera Downing, and Jami Downing. Your kindness, especially as I stumbled, was amazing and I wish you all the best.

Thank you to Lenny and Marjorie Johnson for making so much of this possible.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS v

LIST OF TABLES ix

LIST OF FIGURES x

1. BUILDING REFERENCE SEQUENCES DATABASES FOR MAINE TAXA 1

1.1 Current Usage of Environmental DNA Metabarcoding 1

1.2 Primer Decisions and Limitations 3

1.3 Reference Library Construction 4

1.4 Reference Library Discussions 12

2. COMPARISON OF ENVIRONMENTAL DNA METABARCODING TO A SPECIMEN-BASED SURVEY METHOD 14

2.1 Introduction 14

2.1.1 Specimen-Based Survey Methods and eDNA 14

2.1.2 Environmental DNA in Flowing Water Systems 15

2.1.3 Biomonitoring and Stream Explorers 16

2.2 Methods 18

2.2.1 Sampling Protocol 21

2.2.2 eDNA Water Filtration 22

2.2.3 eDNA Extraction 22

2.2.4 PCR and Sequencing 23

2.2.5 Bioinformatic Workflow 24

2.2.6 Gamma and Alpha Diversity 25

2.2.7 Beta Diversity and Abundance Correlation 25

2.3 Results 26

2.3.1 eDNA Workflow Filtering 26

2.3.2 Gamma and Alpha Diversity 30

2.3.3 Beta Diversity and Abundance Correlation 32

2.4 Discussion and Conclusions 35
3. DETECTION OF SHIFTS IN COMMUNITY COMPOSITION BEFORE AND AFTER LARGE ADDITION RESTORATION USING AQUATIC ENVIRONMENTAL DNA 39
3.1 Introduction 39
3.2 Methods 41
3.2.1 Sampling and Field Filtration 41
3.2.2 Water Filtration 44
3.2.3 eDNA Extraction 44
3.2.4 PCR and Sequencing 45
3.2.5 Bioinformatic Workflow 46
3.2.6 Gamma and Alpha Diversity 46
3.2.7 Beta Diversity 47
3.3 Results 47
3.3.1 Workflow Filtering 47
3.3.2 Gamma and Alpha Diversity 53
3.3.3 Beta Diversity 57
3.4 Discussion and Conclusions 59
REFERENCES 63
APPENDIX 69
BIOGRAPHY OF THE AUTHOR 87

LIST OF TABLES

Table 1.1 Resulting counts of species name resolution and identification 7

Table 1.2 Number of resolved Metazoan species by phyla 8

Table 2.1 Calculated alpha diversity metrics for Stream Explorers data 31

Table 2.2 Jaccard Similarity scores for Stream Explorers observations 33

Table 2.3 Bray-Curtis Dissimilarity scores for Stream Explorers organism counts 33

Table 3.1 Site numbers by category 43

Table 3.2 Non-parametric ANOVA results for alpha diversity comparisons 57

Table 3.3 PERMANOVA results for AMC samples based on presence-absence 59

Table A.1 Workflow verification for cleaning PNW 12S reads 73

Table A.2 Workflow verification for cleaning PNW COI reads 74

Table A.3 Total ASVs counted for each clean PNW sample 78

Table A.4 Number of organismal categories found at each Stream Explorers site 79

Table A.5 Calculated alpha diversity metrics for eDNA samples 79

Table A.6 Individual and ASV counts for taxa observed through Stream Explorers and eDNA surveys 79

Table A.7 PERMANOVA results for PNW eDNA samples based on ASV count 80

Table A.8 PERMANOVA results for PNW eDNA samples based on presence-absence 80

Table A.9 Calculated alpha diversity metrics for AMC 12S eDNA samples 92

Table A.10 Calculated alpha diversity metrics for AMC COI eDNA samples 93

LIST OF FIGURES

Figure 1.1	GBIF polygon search reference	5
Figure 1.2	Display of Metazoan species found with 12S reference sequences	10
Figure 1.3	Display of Metazoan species found with COI reference sequences	11
Figure 1.4	Number of species found in both reference libraries	12
Figure 2.1	Site locations for Stream Explorers and PNW eDNA sampling	20
Figure 2.2	Stream Explorers and PNW sampling sites in context	21
Figure 2.3	Workflow read trends for PNW 12S samples	26
Figure 2.4	Workflow read trends for PNW COI samples	27
Figure 2.5	Class identifications of taxa detected in cleaned PNW COI samples	28
Figure 2.6	Class identifications of taxa detected in cleaned PNW 12S samples	28
Figure 2.7	Rarefaction visualization of PNW 12S samples	29
Figure 2.8	Rarefaction visualization of PNW COI samples	30
Figure 2.9	Diversity and ASV abundance of families detected in PNW COI samples	31
Figure 2.10	Plot of alpha diversity metrics conducted on PNW COI samples	32
Figure 2.11	Ordination results for PNW eDNA samples	34
Figure 2.12	Shared taxa and abundances for SE and COI eDNA	34
Figure 2.13	Dissimilarity plot for SE and COI eDNA	35
Figure 3.1	Map of eDNA sample sites and LWA installations in the LWA survey area	42
Figure 3.2	Diagram of eDNA subsample collection plan	43
Figure 3.3	Workflow progression of AMC 12S reads	48
Figure 3.4	Workflow progression of AMC COI reads	49

Figure 3.5 Class identification of taxa detected in cleaned AMC 12S samples 50
Figure 3.6 Class identification of taxa detected in cleaned AMC COI samples 51
Figure 3.7 Rarefaction visualization of AMC 12S samples 52
Figure 3.8 Rarefaction visualization of AMC COI samples 52
Figure 3.9 Diversity and ASV abundance of families detected in AMC LWA samples with the 12S primer 54
Figure 3.10 Diversity and ASV abundance of families detected in AMC LWA samples with the COI primer 55
Figure 3.11 Plot of alpha diversity metrics conducted on AMC 12S samples 56
Figure 3.12 Plot of alpha diversity metrics conducted on AMC COI samples 57
Figure 3.13 Jaccard ordination for AMC sites based on treatment type 58
Figure 3.14 Jaccard ordination for AMC sites based on LWA position 58
Figure A.1 Stream Explorers field data sheet - top 69
Figure A.2 Stream Explorers field data sheet - reverse 70
Figure A.3 Aggregate quality profile plots for PNW raw reads 70
Figure A.4 Error rates for PNW 12S forward reads 71
Figure A.5 Error rates for PNW 12S reverse reads 72
Figure A.6 Error rates for PNW COI forward reads 72
Figure A.7 Error rates for PNW COI reverse reads 73
Figure A.8 Phylum identifications of taxa detected in PNW 12S samples 75
Figure A.9 Phylum identifications of taxa detected in PNW COI samples 76
Figure A.10 Diversity and ASV abundance of families detected in PNW 12S samples 77
Figure A.11 Plot of alpha diversity metrics conducted on PNW 12S samples 78
Figure A.12 Error rates for AMC 12S forward reads 80
Figure A.13 Error rates for AMC 12S reverse reads 81
Figure A.14 Error rates for AMC COI forward reads 82

Figure A.15 Error rates for AMC COI reverse reads 83
Figure A.16 Phylum identifications of taxa detected in AMC 12S samples 84
Figure A.17 Phylum identifications of taxa detected in AMC COI samples. 84

CHAPTER 1

BUILDING REFERENCE SEQUENCE DATABASES FOR MAINE TAXA

1.1 Current Usage of Environmental DNA Metabarcoding

Environmental DNA (eDNA) refers to a category of biological monitoring techniques as well as the physical material consisting of free-floating extracellular DNA, tissue, feces,[1,2] and other shed material[1, 3-6] that persists in environmental media. Successful isolation of eDNA has been reported in air,[7] sediment,[6, 8-11] ethanol from bulk samples,[12-13] and water[2,5-6,8-9] - with water being the most commonly used media for eDNA[1] in the context of both mesocosm experiments[1,5,14-16] or collecting water directly from marine and freshwater sources.[3-4,11,17-18] As every organism sheds eDNA through biological activity,[2,17] this pool of genetic material is ubiquitous throughout the environment, though the concentration of eDNA is not uniformly distributed for each species. [4, 9, 16] Due to the varying biological sources of eDNA, eDNA as a physical entity can be simulated and understood as a collection of particles[9, 15, 19] with high variety in diameter size[6] that are subject to complex combined mechanisms of decay[6, 15] and transport in aquatic systems.[1,4,15-16] Understanding how eDNA decays and the environmental factors influencing these mechanisms is a critical area of study[4,16,20] in the process of linking detected eDNA concentrations to organism abundance[4,8,15-16] and transport distance.[1,4,15-16] Luckily, depending on the research questions selected by each study, a thorough understanding of these mechanics is not always required and studies can be performed with eDNA detection as long as key assumptions and limitations are communicated and addressed in the study design.[1,4,10,15,21]

There are two common approaches to utilizing eDNA - single-species assessment[20] and broad taxonomic assessment. Single-species eDNA studies commonly utilize qPCR or ddPCR,[20] among other techniques, both of which detect the presence of eDNA in the environment and report the DNA concentration or distribution of signal for the respective species. These techniques attempt to relate organism abundance or biomass to eDNA concentration, as well as

checking for the presence of species in a given area.[15,16,20] However, there are tens or hundreds of species of interest globally depending on the research region and question, and restricting a broad study to single species assays can be time-consuming and expensive.[20] Metabarcoding, a technique that applies high-throughput sequencing to enable identification of numerous species within a single sample, is applied to broad-taxa questions instead.[3,12,18,20-21]

Metabarcoding relies upon the successful extraction[8] of DNA from a sample, whereupon a specific gene region or locus is sequenced and the resulting detected amplicon sequence variants (ASVs) are compared against a database of reference sequences,[10,20,22] hereafter referred to as a reference library. This reliance on a reference library makes the quality and completion of the library in relation to the target taxa critically important.[22-23] If an ASV is present in the sample but neither present or identified to a useful taxonomy level in the reference library, it may not be included in the usable output results even if it is a high quality or valid sequence for its species.[3,5] Factors influencing the quality of the reference library include the databases chosen for sourcing references, completeness of metadata in the reference sequences,[20,23] and how well the reference sequences were identified or collected to begin with.[3] Furthermore, regional variation in a species' genome should also be considered as regional hybridization can result in ASVs for a species not being properly identified.[5]

The International Nucleotide Sequence Database Collaboration is the leading repository for raw genomic data[23] and includes the National Center for Biotechnology Information which maintains the GenBank sequence database. Uploading raw genetic data to GenBank is now a requirement or strong recommendation of many journals for publishing; however adequate metadata to support reuse of genetic sequence is severely lacking across the repository.[23] A 2021 metadata assessment by Toczydlowski[23] et al. found that out of 600 terabytes of genomic data, only 13% included basic spatiotemporal metadata such as locale and year of collection that would be crucial for reuse or application of the data in future studies. While manual effort to reach out to dataset contacts was able to fill in basic metadata for roughly 1,500 datasets, the

effort was highly time-consuming and restricted by expiring author contact information and loss of original written records.[23] Certain repositories or data workflows, such as BOLD or GEOME, require a certain level of metadata completion for upload and can reduce the risk of either misidentification or missing metadata, but are not as widely used or geographically complete.[3] Wherever possible, the efficient use of eDNA metabarcoding necessitates the contextualization of genomic[3,5,12,16,20,24] and sampling resources[4,16,20] with the target taxa[3,5,12-14,17] and environment[15-17]; This is to best isolate, detect, and identify eDNA.

1.2 Primer Decisions and Limitations

The Maine-eDNA EPSCoR program (Maine-eDNA) is an RII Track 1 research grant funded by the National Science Foundation to advance ecological and eDNA knowledge to investigate coastal macrosystems and inform ecosystem management and restoration initiatives. Maine-eDNA involves multiple institutions, faculty, and graduate students researching separate projects utilizing eDNA, but certain decisions and processing procedures are common across projects. In order to enable comparison of results and access common troubleshooting resources, I chose two primers that were the same or highly similar to those used in other projects. These primers are MiFish-U,[5,15] designed from Okinawan aquarium and marine species[14] to amplify bony fish using the 12S mitochondrial ribosomal subunit gene, and the pair BF2 + BR2, designed[12] from insects collected in Ontario, Canada[25] to amplify benthic macroinvertebrates using a subset of the cytochrome c oxidase subunit (COI) region.[12,24-25] These primers will be referred to as 12S and COI for the remainder of this thesis. The 12S primer pair involves a 21 bp forward primer (GTCGGTAAAACTCGTGCCAGC) and a 27 bp reverse primer (GTTTGACCCTAATCTATGGGGTGATAC) to amplify a region approximately 170 bp long.[5,14] This pair has been well-documented for its ability to identify fish species when applied even outside of its initial geographic range[10,20] including Maine bony fish taxa,[5] but has a limited reference library size.[18] The COI primer pair involves a 20 bp forward primer

(GCHCCHGAYATRGCHTTYCC) and a 20 bp reverse primer (TCDGGRTGNCCRAARAAYCA) to amplify a region approximately 420 bp long.[12,18,24] This pair benefits from cytochrome oxidase I being a common choice for DNA barcoding invertebrates,[5,12,18,24] resulting in a large potential reference library for use.[5,13] However, BF2+BR2 is known to amplify non-target (namely bacterial) DNA even when the primer matches poorly to the non-target reads, watering down target macroinvertebrate DNA. This is likely due to higher degeneracy that allows the primer effective usage as a universal primer, sacrificing specificity.[18,24] To create a regional reference library for each primer, a species list for taxa in the Maine region was first constructed, then used to create separate reference libraries based on target search term matching in GenBank entries.

1.3 Reference Library Construction

As a regional species list is critical for metabarcoding projects,[5,9] construction of a Maine regional list began as part of the broader Maine-eDNA effort. Initially, species lists used by labs and agencies were requested and compiled together, along with online and text resources that listed Maine flora and fauna.[26] Finally, a polygon search (Figure 1.1) was used to search the GBIF database for species occurrence records within the search area. This search polygon was set larger than the Maine state boundaries to capture invasive, roaming, or migratory species. The results from the search[27] were downloaded and cleaned in R (version 4.3.1) separately from the other lists due to size. All non-GBIF lists were compiled into a single spreadsheet while keeping record of the original source, then cleaned in R to consolidate duplicates and correct case errors. The resulting species list was then processed using the taxize R package[49] to check species names against known taxonomic databases to correct misspellings, out-of-date species binomials, and verify the validity of species names. Only binomials that passed this taxonomy check were kept in the final species list for use in reference library construction, and included 25,007 species names.

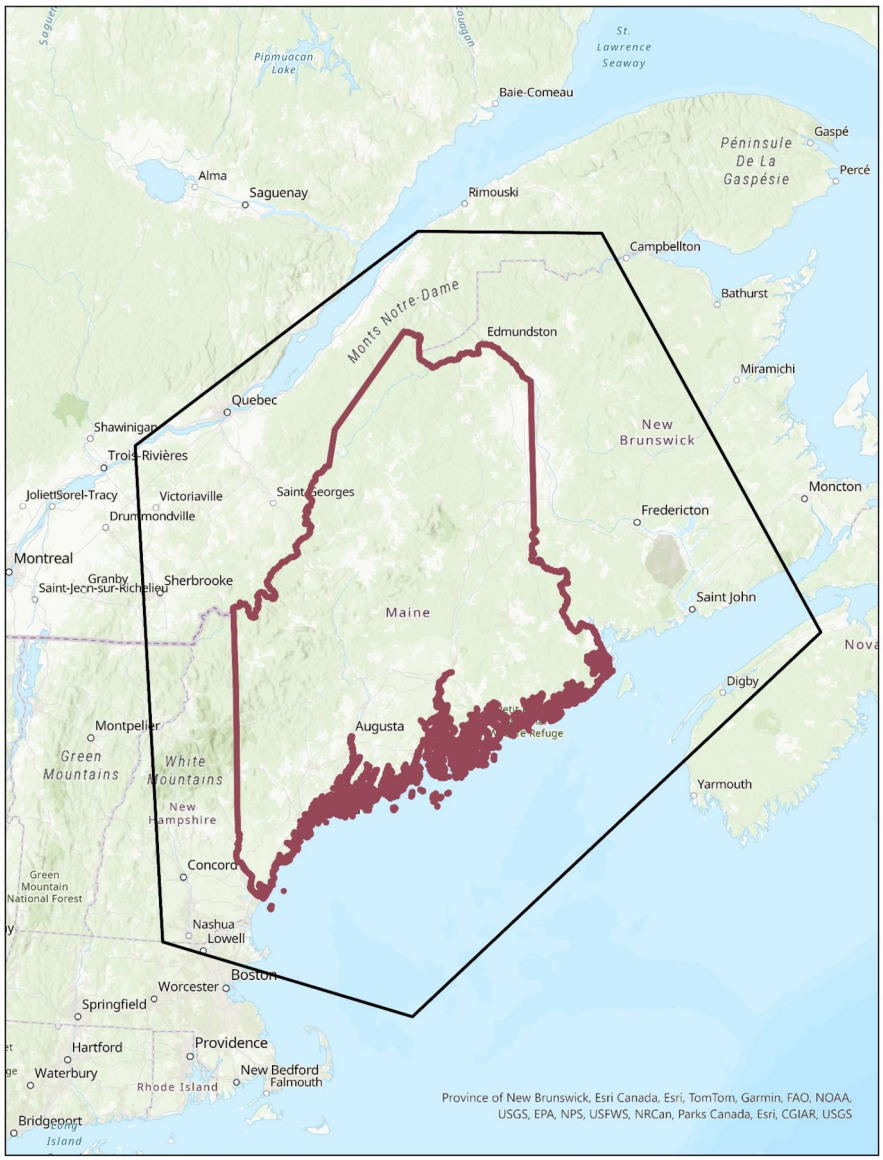

Figure 1.1 GBIF polygon search reference. Polygon reference (black) of the location-based search for GBIF species occurrence records, with outline of the Maine state region (red).

Search terms were created in R as a combination of the target primer region (and synonyms found with AnnotationBustR[48]) and a species name from the species list, and used to scrape accession numbers then ultimately FASTA sequences from search results on GenBank. For both reference libraries, a check was performed to identify taxonomic orders that had no accession numbers found from the search. Missing orders were then randomly sampled to select up to three accession numbers from species in those orders for addition into the reference library. This was done to provide additional comparison points for ASVs and reduce the risk of false identifications. Once sequences were found, they were deduplicated and formatted for DADA2 workflow processing. Code for the search process is available on GitHub.[47]

Overall, the 12S reference library initially lacked 209 taxonomic orders, and the missing order representative search added 524 species to the library, though only 352 of those species were found to have target sequences. The COI reference library lacked 273 taxonomic orders and a search added 606 species to the library, with 483 of those found to have target sequences. There were only 3,218 species present in both the COI and 12S libraries. A breakdown of species by kingdom that passed from the species list to each library is available in Table 1.1. Notably, for the target fish or invertebrate taxa, the resolved species list only contained 10,908 target taxa. Of those, 2,912 target taxa were present in the 12S reference library and 9,712 in the COI library (Table 1.2).

Table 1.1 Resulting counts of species name resolution and identification. Number of Maine species from the initial species list that passed resolution and cleaning, identified to taxonomic kingdom. Compared against the number of species found with available GenBank target loci sequences for the 12S and COI reference libraries, also organized by kingdom. Reference library counts include representative species added for orders missing in the resolved species list.

Category	Resolved Species List	12S Reference Library	COI Reference Library
Total Species	25007	25356	25485
Total Species with Target sequences:	NA	6920	10933
Animalia	555	11	39
Bacteria	531	34	9
Chromista	223	2	1
Fungi	2194	1880	90
Metazoa	10353	2901	9673
Plantae	117	10	3
Protozoa	1	0	0
Viridiplantae	3790	1179	309

Table 1.2 Number of resolved Metazoan species by phyla. Counts of Metazoan species in the resolved Maine species list, and that were found to have target reference sequences for 12S and COI in GenBank. Table does not include any species that were identified at other taxonomic levels without phylum identification.

Phylum	Resolved Species List	12S Reference Library	COI Reference Library
Acanthocephala	7	4	5
Annelida	274	146	230
Arthropoda	7804	1055	7037
Brachiopoda	6	1	1
Bryozoa	49	27	32
Chaetognatha	3	3	2
Chordata	1843	1347	1711
Cnidaria	121	49	87
Ctenophora	4	3	4
Cycliophora	1	0	1
Echinodermata	50	23	44
Entoprocta	2	0	1
Gastrotricha	1	0	1
Gnathostomulida	1	0	1
Hemichordata	4	0	1
Mollusca	432	124	363
Nematoda	24	19	16
Nemertea	22	12	22
Phoronida	1	0	1
Platyhelminthes	56	18	23
Porifera	40	15	29
Priapulida	1	1	1
Rotifera	147	60	88
Tardigrada	13	5	10
Xenacoelomorpha	1	0	0

I found that while, expectedly, the overall number of species with target sequences were different for each library, the taxonomic spread of each reference library was also different. The 12S search found more existing records for vertebrates (Figure 1.2) and the COI search found more records for invertebrates and algae or plants (Figure 1.3). This matches with the typical usage in DNA barcoding for each primer region[5,18,24] and as reported in other studies the COI library[5,13] found far more target accessions than the 12S search.

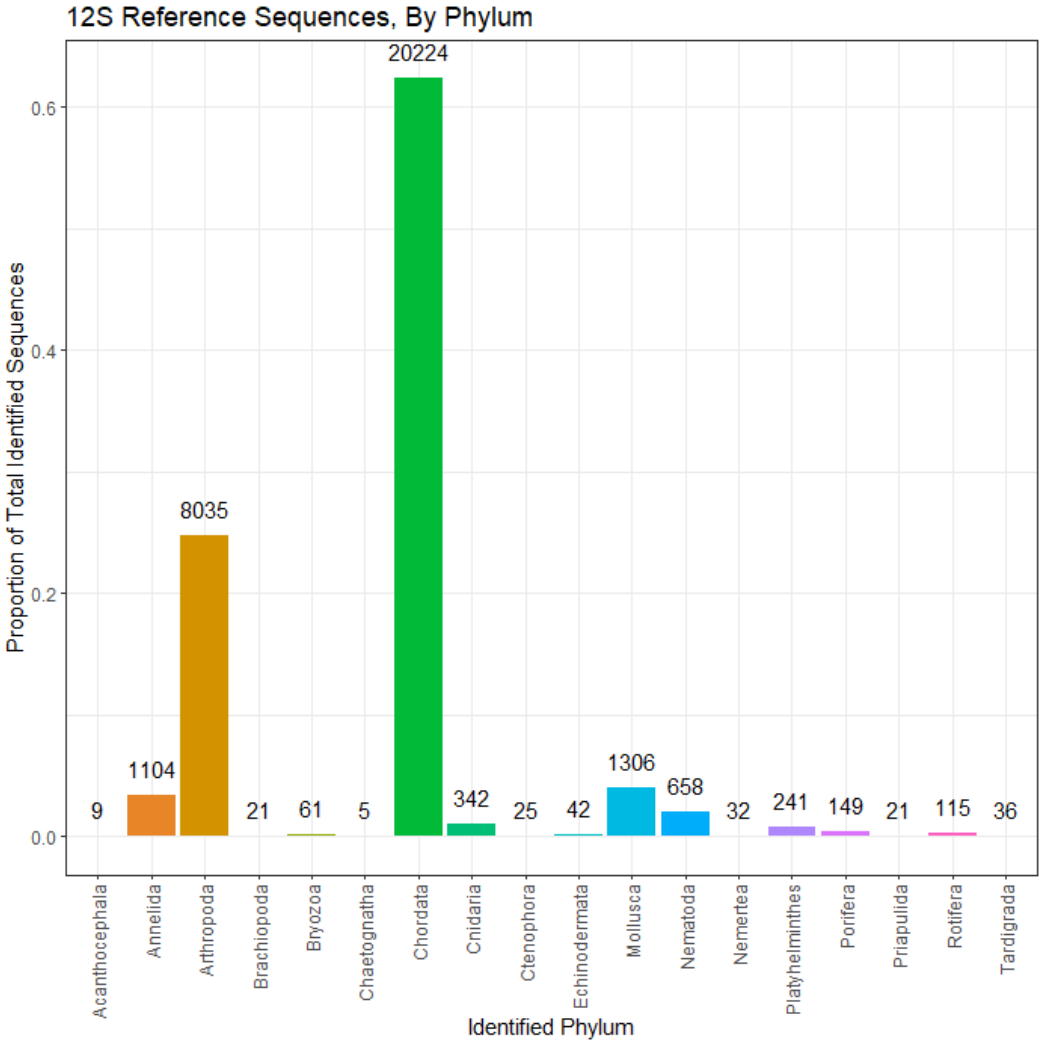

Figure 1.2 Display of Metazoan species found with 12S reference sequences. Proportion of Metazoa phyla in the 12S reference library based on the number of target reference sequences found in GenBank (top number of each bar), divided by the total number of 12S target sequences found.

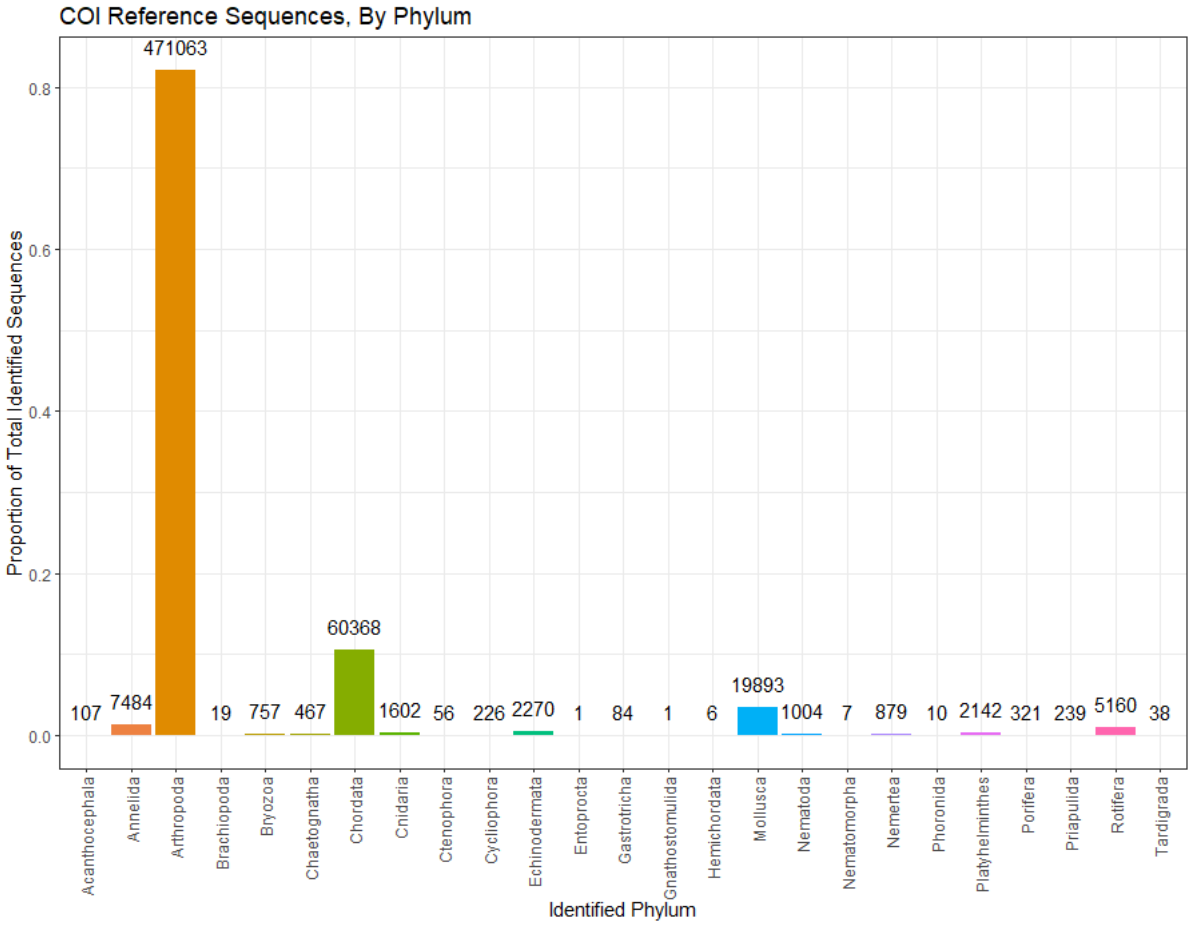

Figure 1.3 Display of Metazoan species found with COI reference sequences. Proportion of Metazoa phyla in the COI reference library based on the number of target reference sequences found in GenBank (top number of each bar), divided by the total number of COI target sequences found.

Of the species occurring in both libraries, the majority were vertebrates, followed by arthropods (Figure 1.4). Based on these results, I expect a considerable amount of identification bias, with low overlap in species detected by each primer, and for 12S to primarily detect vertebrates and bacteria, and COI to detect primarily invertebrates and algae.

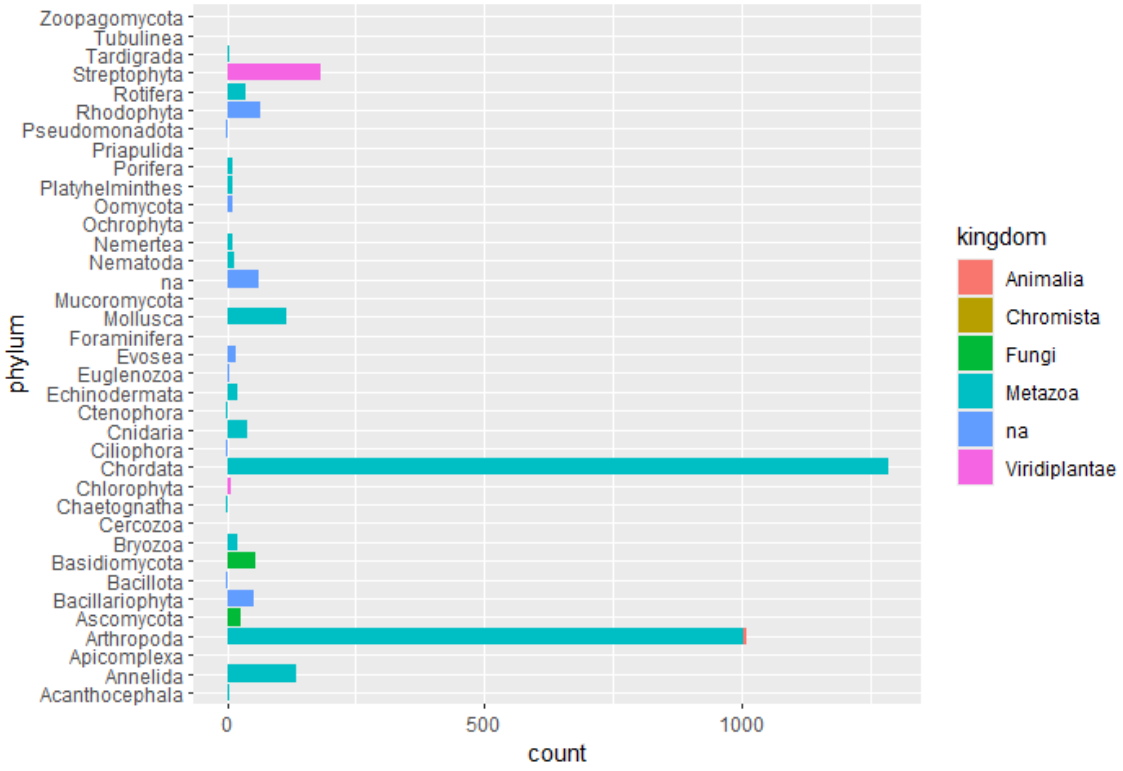

Figure 1.4 Number of species found in both reference libraries. Barplot of the number of resolved species by phylum that were found to have both 12S and COI target sequences in GenBank. Colored by taxonomic kingdom.

1.4 Reference Library Discussions

From the initial species list, the COI reference library retained most of the target Metazoan taxa (Table 1.1), primarily split between arthropods and chordates (Table 1.2). The 12S library included less than a third of the initial number of Metazoan species with 27% of the library consisting of fungal sequences (Table 1.1). The construction process did appear to have a potential naming bias - while Animalia and Metazoa are synonyms along with Plantae and Viridiplantae, resolved taxonomy for the species list tended to identify with the latter kingdoms in both cases. For species identified to the Animalia and Plantae kingdoms, a much lower

proportion of species were found with target reference sequences. These differences may reflect the historical evolution of taxonomic names, though a closer examination would be required to identify specific patterns.

While these libraries were created purely from pre-existing digital records, localized reference libraries are best created by sampling or otherwise collecting material from positively identified species in the region of interest.[3,5,12-17] This can include collaborating with museums to extract sequences from preservation fluid[13] or small amounts of preserved tissue, hosting community bio-blitzes for common taxa, or other dedicated sampling efforts to collect exemplar specimens.[25] As sequences are digital data, a local reference library can grow over the course of years as gaps are identified and funding for projects becomes available. A benefit to eDNA sampling is that so long as data is documented and preserved appropriately,[20,23] sample sequence data can be re-analyzed with new versions of a reference library if future researchers are interested in returning to past samples.

In this thesis, I conduct two proof-of-concept studies using eDNA metabarcoding to monitor species in Maine's mountainous streams using two primers each to collect fish and benthic macroinvertebrate data. The first, in Chapter 2, pairs eDNA metabarcoding with a specimen-based benthic macroinvertebrate survey method to test eDNA detection of Maine species and the comparability of eDNA concentration metrics with abundance counts. The second, in Chapter 3, compares eDNA detection results before and after stream restoration efforts, and tests both the success of eDNA detection and conducts a preliminary survey on the success of restoration. Both projects utilize the same two primers and the same reference libraries for ASV identification.

CHAPTER 2

COMPARISON OF ENVIRONMENTAL DNA METABARCODING TO A SPECIMEN-BASED SURVEY METHOD

2.1 Introduction

2.1.1 Specimen-Based Survey Methods and eDNA

Traditional methods of surveying ecosystems generally involve some level of disruption[4,21] or invasion; trawling,[11,20] electrofishing,[10] kicknet sampling,[28] etc. all require physical collection of specimens to collect species, size, sex, or other target information.[4,10-11,20-21] These methods can involve capture-and-release[11,16,20] or euthanizing organisms for preservation and taxonomic identification.[21] Due to the reliance on directly handling organisms, these methods can be classified as specimen-based survey methods (SBSM) and are often utilized in combination with water quality assessment or other environmental metrics. Combining species presence or abundance data with environmental monitoring enables researchers and ecosystem managers to link environmental factors with impacts on species, or vice versa.[20,24,29] However, SBSMs tend to be expensive,[20] require high amounts of labor or field time,[3,6] and may completely miss rare species[6,9,11,15] (species present in low numbers in the sampling area[45]). SBSMs are also limited to study regions and time periods that are accessible to field crews carrying equipment, such as shallow wadeable stream reaches during calm weather, or which can be accessed by boat. Because of these limitations, SBSMs are generally limited to sampling a fairly small area to most efficiently apply resources, and habitats that are difficult to reach are undersampled and understudied.[21]

While eDNA metabarcoding is still a developing technique particularly in streams, it poses multiple known advantages over SBSMs. Namely, the cost of eDNA metabarcoding has consistently decreased and is relatively inexpensive given the number of species that can be detected with metabarcoding,[20] though studies utilizing multiple primers will require higher

costs for multiple sequencing runs. Collecting eDNA samples can be done quickly and does not require extensive taxonomic or specialized machinery training,[9,21] and the basic equipment necessary are sterile gloves and a sterile water container. This basic equipment can easily be carried in a pack and greatly lowers the amount of equipment needed for sampling, making it easier for field crews to trek in and out of a site. eDNA is also capable of detecting cryptid species that may otherwise be missed by SBSMs,[6,9] though the success of this detection is still reliant on species behavior and eDNA sampling design.[21] For non-cryptid species, eDNA has been reported to have a much higher sensitivity for species detection than most SBSMs.[1,18] This sensitivity may not yield the same detection results as SBSMs, however – studies using metabarcoding on water samples or bulk tissue samples both report distinctly different community results between eDNA methods and SBSM methods.[50-53] These differences vary from low[50-52] to high[53] amounts of overlap between eDNA and SBSM methods, and are often attributed to incomplete reference libraries,[53] lack of sample replication,[51] or methodological limitations such as primer design.[52]

Despite continuing uncertainties around decay,[1,6,15-16] transport,[4,9,15-16,19] and nutrient interaction, eDNA is still a promising technique for use in ecological monitoring, particularly in areas that are difficult or budgetarily restrictive for SBSM sampling[21] such as remote streams.

2.1.2 Environmental DNA in Flowing Water Systems

As a complex particle,[15] eDNA is subject to flow rate, deposition, transportation, and suspension, particularly in flowing stream systems.[9,15-16] The precise extent to which eDNA can travel while still providing quality results is currently unknown, with previous literature recording distances of less than 10 meters[15] to hundreds of kilometers.[4,9] Transport generally takes the shape of a conical plume[4] from the eDNA source, and simulations of transport report that without decay, eDNA ultimately grows in concentration downstream of the source along the stream banks.[4,16] While shorter DNA fragments have been found to decay more slowly than

longer fragments, eDNA water samples can show up to 90% decline in concentration after one day in the environment.[1] However, eDNA is able to persist long enough for sediment eDNA to record years or even decades of some genetic information, and as such disturbances to stream substrate such as storms or floods can resuspend eDNA whilst diluting the overall eDNA concentration by increasing water volume. In addition, abiotic and biotic factors[10,16] have been found to influence rates of eDNA decay, likelihood of deposition, or output rates of eDNA from a source.

Previous studies have found that water temperature,[6,16] benthic biofilm cover,[1,10,16,19] chlorophyll concentration,[1] UV light, water oxygen levels,[17] pH,[16] and soluble nutrient concentrations[10] can all impact detected eDNA concentration, though the strength and nature of the relationships found vary between studies. Biological factors influencing eDNA concentration or shedding are related to conditions or behaviors that increase the amount of loose biological material in the environment,[3,17] including reproductive behaviors,[11] predation,[3] aging,[3,11] and death. Other factors found to influence eDNA shedding are organism size,[8-13] surface area,[8] biomass,[11] exoskeleton presence,[5,11,13] and placement within the water column.[6] Soft-bodied organisms such as most bony fish, vertebrates, and soft macroinvertebrates may have higher rates of eDNA output than crustaceans or other macroinvertebrates with chitinous exoskeletons.[5] Such hard-shelled benthic macroinvertebrates have been shown to emit little[5] to no detectable eDNA[11] in water samples, with paired sediment samples showing higher eDNA concentration and detection rates.[11] It is highly recommended that eDNA studies take target taxa behaviors and traits into consideration when creating sampling designs,[9,11] as results can significantly vary between microhabitat differences even at a single site.[11]

2.1.3 Biomonitoring and Stream Explorers

Under the federal Clean Water Act[24,30] and growing global awareness[24,31] of the importance of healthy freshwater systems, federal, state, and tribal agencies have increased

efforts to identify and restore impaired surface water systems.[24,30,32] This identification generally combines the monitoring of chemical and biological identifiers[9,24,31-32] to classify the health of a water body or stream[30,32] and guide future action for restoration or maintenance.[30-31] In Maine, the Penobscot Nation Department of Natural Resources (DNR) and the Maine Department of Environmental Protection (DEP) are two agencies working both individually and in collaboration to monitor stream and water body health. The state of Maine uses four classes for freshwater streams (AA, A, B, C)[30-32] to convey a hierarchy of risk of stream degradation due to natural or anthropogenic events.[30] As of DEP's Draft 2024 Integrated Water Quality Monitoring and Assessment Report, 7.5% of the total river and stream miles in Maine were classified as AA, 46.8% as A, 44.6% as B, and 1.1% as C out of 40,791 miles assessed in 2024.[30] In order to enable stream assessment on such a large scale, DEP's assessment system allows for and includes SBSM techniques on disturbance- or pollution-sensitive species.[30-33] This practice of biomonitoring has been widely applied globally[3,9,24,31] to use sensitive species to estimate the environmental quality of an area. For streams, the EPT group (Ephemeroptera, Plecoptera, Trichoptera)[31-32] of macroinvertebrates is commonly used, though certain fish[10,30,32] and algal[30,32] species can also serve as biomonitoring indicators. The taxa used for biomonitoring can depend on available taxonomists,[3] stream morphology, and available sampling resources.

As identifying organisms is highly time consuming[3] and clarifying precis species is impossible in some cases,[24] Maine Audubon partnered with the Lakes Environmental Association, DEP, and the Portland Water District to develop the Stream Explorers project (SE).[33] Initially designed to train citizen volunteers to survey streams in the Sebago Lake Watershed, SE surveys group benthic macroinvertebrates into non-species-specific categories that are easy for beginners to visually identify (Figures A.3-A.4).[33] These categories use common names and are roughly equivalent to family taxonomic rank; a loose precision level for genetics studies, but still precise enough to separate organisms into Sensitive, Moderately Sensitive, and Tolerant categories for biomonitoring use.[24,33] These categories describe the organisms'

sensitivity level to pollution, with the final SE report combining the number of organismal categories found in each sensitivity category and the categorical abundance (Few, Common, Abundant) of the organisms.[33] This broad categorization allows greater numbers of volunteers and community scientists to take part in water quality assessment and biomonitoring, and requires virtually no wait time for results - all data can be collected and compiled in the field and specimens returned to the water body. Despite the significantly reduced taxonomic resolution of SE results, the benefits for rapid monitoring results, increased collaboration possibilities, and ability to quickly survey multiple sites in a day has encouraged DEP to apply the SE survey system outside of the initial Sebago Lake Watershed area.

I led a team to collect eDNA samples alongside a joint collaboration between DNR and DEP to learn how the SE system worked and compare eDNA and SE results. Though SE only categorizes benthic macroinvertebrates (Figures A.3-A.4), I applied both the 12S and COI primers to eDNA samples to test if there was any overlap in species detection for target fish and benthic macroinvertebrates that would enable a single primer to be used for effective Maine stream biomonitoring. The goals of this comparison were to a) compare the number and diversity of families detected with the two primers to the organismal categories found and counted with the SE method, b) compare the number of organisms counted with SE with the number of ASVs detected with eDNA, and c) examine the number and taxonomic spread of species identified with eDNA.

I hypothesize that eDNA metabarcoding will detect the families of the Stream Explorers' target organisms and a wider overall range of species, but will not show a correlation between ASV abundance and the counted target organism abundances.

2.2 Methods

Sites were selected as part of a joint sampling effort between DNR and DEP for biannual sampling in the spring and fall of each year, and focused on three individual sites on the East

Branch Penobscot River in Maine (Figure 2.1). Each sampling site was located in a shallow, wade-able section of the stream, with typically straight channels and gradual banks, and low to no visible in stream wood beyond nearby trees. Flow at each site was typically slow and visually uniform with a large stream width (6-7 meters+) and straightness. Sampling occurred in run sections of each stream. Due to weather events and scheduling, eDNA samples were only collected at two time periods - September 2022 and May 2023. Each site involved a roughly 15-minute sampling period followed by 45-60 minutes of specimen identification for the SE protocol, and a total of four liters of water for eDNA sampling collected before SE sampling. Each day of sampling used a single liter of DI water as a negative control, which traveled with all field samples and was filtered first as a check for lab equipment.

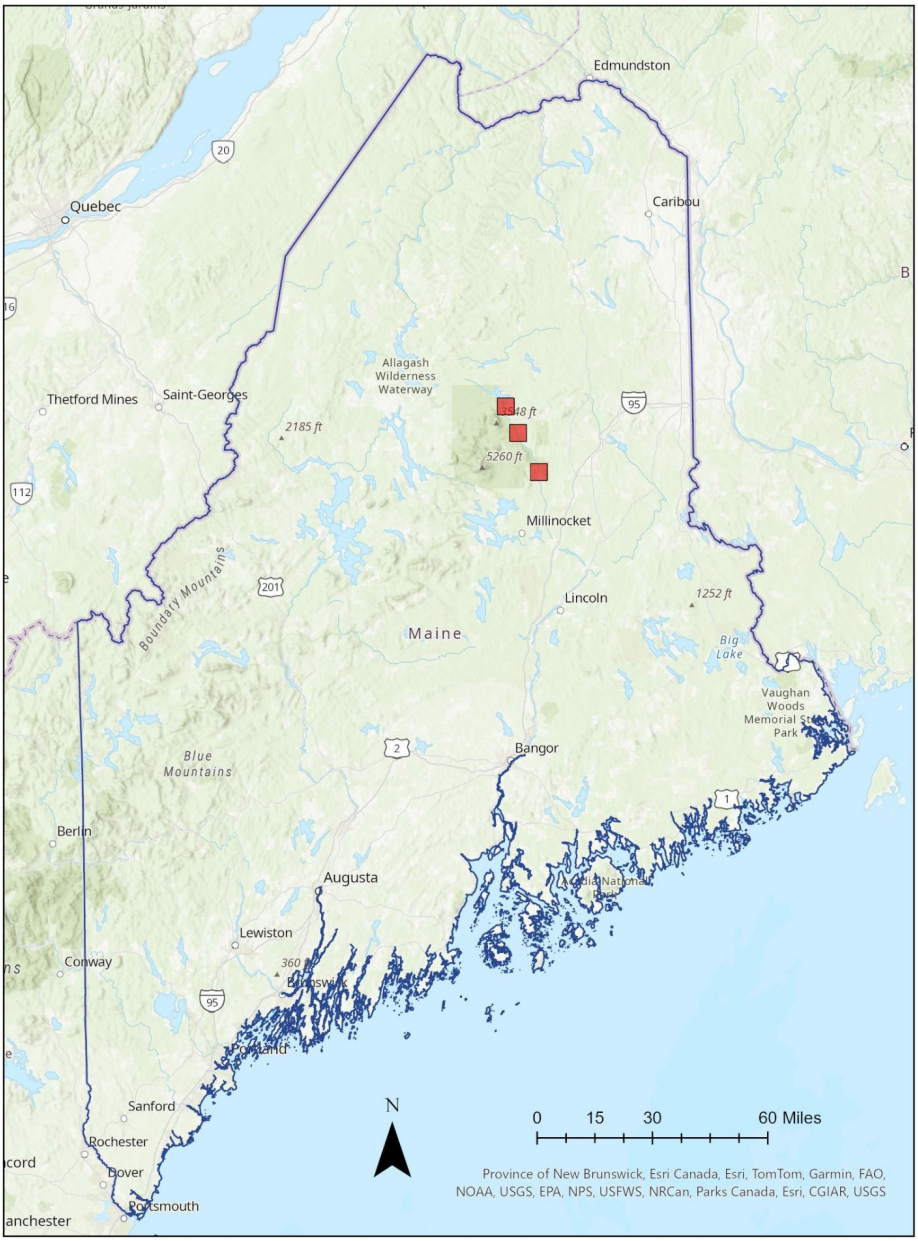

Figure 2.1 Site locations for Stream Explorers and PNW eDNA sampling. Location of sites (red squares) sampled in relation to the Maine state borders.

2.2.1 Sampling Protocol

Water collection utilized 1 liter Nalgene bottles that had been soaked in 10% bleach solution for 10 minutes, rinsed five times with tap water, then rinsed again in site river water just before collection. This water collection was divided into upstream and downstream sampling, with each position distanced 5-10 m away from the SE collection area (Figure 2.2). For both positions, one liter of water was collected near the stream bank, and one liter collected mid-stream, both approximately 15 cm from the stream bottom for water depths ranging from 30-76 cm, for a total of 2 liters of water for each position. Water samples were placed on ice in a cooler for transport back to the lab for filtration.

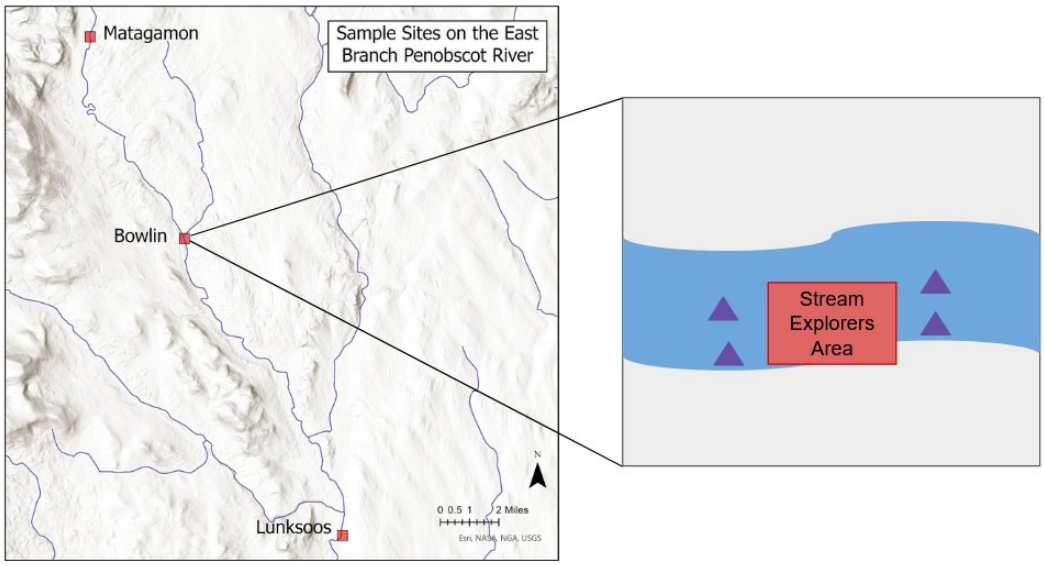

Figure 2.2 Stream Explorers and PNW sampling sites in context. View of sample sites in relation to their placement on the East Branch Penobscot River (left), and diagram of eDNA sampling (right) locations (purple triangles) relative to the Stream Explorers sampling area at each site (red square). Diagram is not to scale.

Once the water samples were collected, Stream Explorers surveying began - six rock samples were taken at each site by placing a net with the opening facing the stream flow and cleaning rocks off so that any present benthic macroinvertebrates flowed into the net. Once collected, nets were emptied into plastic trays of water for sorting. Ice cube trays, plastic syringes, and forceps were used to pick out and identify macroinvertebrates against the SE

sampling sheet (Figures A.3-A.4). All macroinvertebrates were returned to the stream once identification and counting were completed.

2.2.2 eDNA Water Filtration

All samples were placed in a cooler under ice and transported back to the lab for vacuum filtration within 24 hours of collection using sterile single-use 47mm 0.45 μm cellulose nitrate filter papers. All lab surfaces and reusable tools were sterilized with bleach wipes followed by DI water or soaking in 10% bleach solution for at least 10 minutes, and allowed to dry fully.

This was followed by germicidal UV exposure for 30 minutes to an hour before use. For each site position, the upstream and downstream samples were physically aggregated by filtering the bank and midstream sample onto the same filter paper, reducing the number of eDNA samples per site to a single upstream and a single downstream sample. Filter papers were preserved by freezing at -20°C in DNA LoBind tubes until extraction.

2.2.3 eDNA Extraction

eDNA extraction was performed on the filter papers using the Qiagen DNA Powersoil Pro kit, using a protocol developed by Geneva York with some alterations developed by Kylie Holt. All surfaces were wiped down with bleach wipes followed by DI water before use, and pipette tips and other sensitive equipment cleaned with DNA-Off. Other reusable materials such as forceps and tube racks were cleaned by soaking in 10% bleach for 10 minutes, rinsing five times, then allowed to dry fully before use. An hour of germicidal UV exposure was performed in the lab before and after each round of extraction. At least one control sample was extracted in each round of extraction to serve as a laboratory blank as well as field blank. Extracts were eluted to 100 μL volume, then frozen at -20°C before PCR and sequencing.

2.2.4 PCR and Sequencing

For each primer, samples underwent polymerase chain reaction (PCR) and sequencing, with each batch including at least the original field blank for continued use as a negative control. The PCR protocol for both primers used a 25 µL reaction with 9 µL of nuclease-free MB grade H_2O, 1.25 µL of each forward and reverse primer, 12.5 µL of Quantabio HiFi ToughMix 2X, and 1 µL of extracted DNA.

The 12S primer involved a 21 bp forward primer (GTCGGTAAAACTCGTGCCAGC) and a 27 bp reverse primer (GTTTGACCCTAATCTATGGGGTGATAC) to amplify a region approximately 170 bp long.[5,14] Both primers were designed with the addition of the Nextera adapter for MiSeq sequencing. PCR for the 12S samples underwent a protocol of 98°C for 10 seconds, 61°C for 5 seconds, and 68°C for 1 second, repeated for 38 cycles. Samples were refrigerated at 20°C until sequencing.

The COI primer involved a 20 bp forward primer (GCHCCHGAYATRGCHTTYCC) and a 20 bp reverse primer (TCDGGRTGNCCRAARAAYCA) to amplify a region approximately 420 bp long.[12,18,24] Both COI primers were also designed with the addition of the Nextera adapter to prepare for sequencing. PCR for the COI samples underwent a protocol of 95°C for 10 minutes, a 35-cycle repeat of 95°C for 30 seconds, 50°C for 30 seconds, and 72°C for 30 seconds, with a final 72°C for five minute stage. Samples were refrigerated at 20°C until sequencing.

Library prep and sequencing were performed using Illumina Miseq at the University of Maine CORE DNA Sequencing Center by Geneva York and Lindsey Stover. Given the large difference in segment length, 12S and COI samples were sequenced on different runs, though both ran for 300 cycles. Once sequenced, data was sent as FASTA files with the Nextera adapter removed.

2.2.5 Bioinformatic Workflow

All samples were cleaned and prepared for analysis using the dada2 workflow[34,47] in R (version 4.3.1), followed by use of the phyloseq[35] and vegan[36] R packages. To remove low quality reads, quality profile plots were first run on a random sample of both the 12S and COI datasets to guide trimming. For 12S samples, reads less than the default of 20 bp were removed, and forward and reverse reads were trimmed between the primer length (21 bp for forward and 27 bp for reverse, respectively) and 220 bp in raw length. The filterAndTrim function from dada2 was used to trim reads as a pair to enable future read merging. COI samples were trimmed after a total length of 290 bp and 250 bp for the forward and reverse reads, with a left trim of 20 each conducted to remove primers, and also required a minimum read length of 21 bp. Both sets used maxEE = 2, and the default maxN of 0. Default dada2 settings were used for learning errors in both datasets. Once error rates were learned, sample inference, merging of the forward and reverse reads, and removal of chimeras were performed.

The trimmed data were then converted into phyloseq objects, and used assignTaxonomy to identify ASVs against their respective 12S and COI Maine reference library. Any ASVs present in the negative controls were removed from all samples as contamination, and non-target bacterial, fungal, and algal taxa were removed. A check was made to ensure no human DNA was present in the samples, then the cleaned phyloseq objects proceeded for diversity and significance analysis.

While taxonomy was resolved for ASVs, discussion of exact species-level identities and changes in species-level composition across samples is not discussed in this thesis to ensure the privacy of natural resources and habitats managed by collaborators on this project; at the time of defense, consent for species-level disclosure was not able to be given by all collaborators. Species-specific analyses and discussion may be added at a future time depending upon full disclosure consent.

2.2.6 Gamma and Alpha Diversity

Gamma diversity was found by calculating the total number of species detected with each primer for each time period of collection and comparing the unique species identifications for each time period. The same process was used for the SE data.

Alpha diversity metrics were calculated for both the SE count data and the eDNA samples, and focused on Chao1, Shannon, and Simpson measures of diversity. Rarefaction was performed to visualize data patterns, but no samples were removed based on rarefaction results.

2.2.7 Beta Diversity and Abundance Correlation

Beta diversity was calculated on eDNA samples using Jaccard's index for presence-absence as well as Bray-Curtis's dissimilarity matrix on ASV data to consider proportional abundance. For the SE data, Jaccard's index and Bray-Curtis were also calculated for comparison purposes. PERMANOVAs were conducted to test relationships between distance for each eDNA sample placement, the filtration amount, sampling time period, and stream reach. Dissimilarity between samples was visualized using Principal Coordinate Analysis (PCoA).

To check if organism count abundances had relationships with ASV counts in the eDNA data, SE organism categories were first matched as closely as possible with distinct taxonomic ranks. Categories that could not be identified to at least the family level were dropped from comparison, and the remaining taxonomic families were used to isolate the number and species identification of detected ASVs belonging to those families in the 12S and COI eDNA samples. Finally, dissimilarity between the SE survey results and the COI eDNA survey results was plotted followed by a Pearson's correlation test.

2.3 Results

2.3.1 eDNA Workflow Filtering

Of the 14 samples sequenced with 12S and COI, all samples made it past the initial filterAndTrim restriction. Trimming removed between 0% and 9.12% of reads in the 12S samples, with a median of 7.50% and average of 6.62% reads removed. For COI samples, the percentage of removed reads ranged between 0% and 78.95% of the total, with a median of 38.92% and an average of 38.77% of reads removed. At the end of the dada2 workflow, one 12S sample had been removed (PNW23_BOWU) for lack of quality reads, with the remainder of samples containing between 35,862 and 223,596 reads (Figure 2.3).

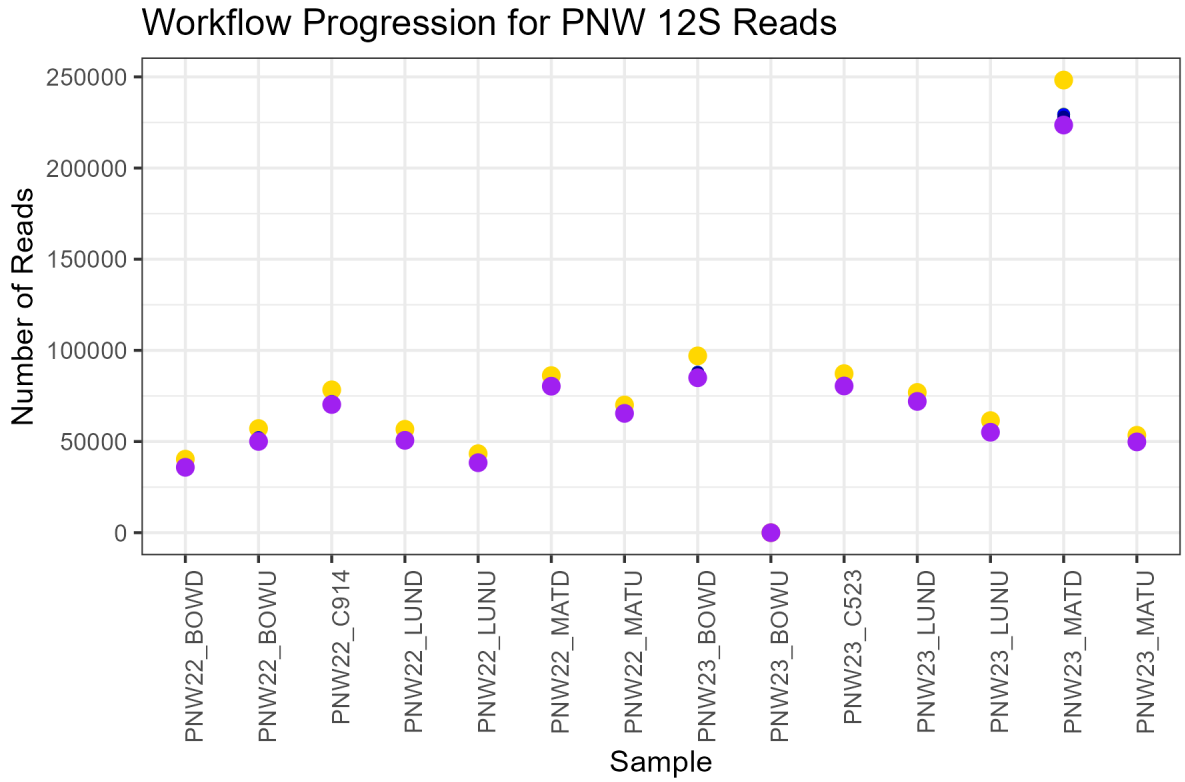

Figure 2.3 Workflow read trends for PNW 12S samples. Display of the total reads found in 12S samples before dada2 processing (gold, top dot) and after (purple, bottom dot) finishing the dada2 workflow with removal of chimeras. Dark blue dot visible with PNW23_MATD refers to the read numbers after merging forward and reverse reads.

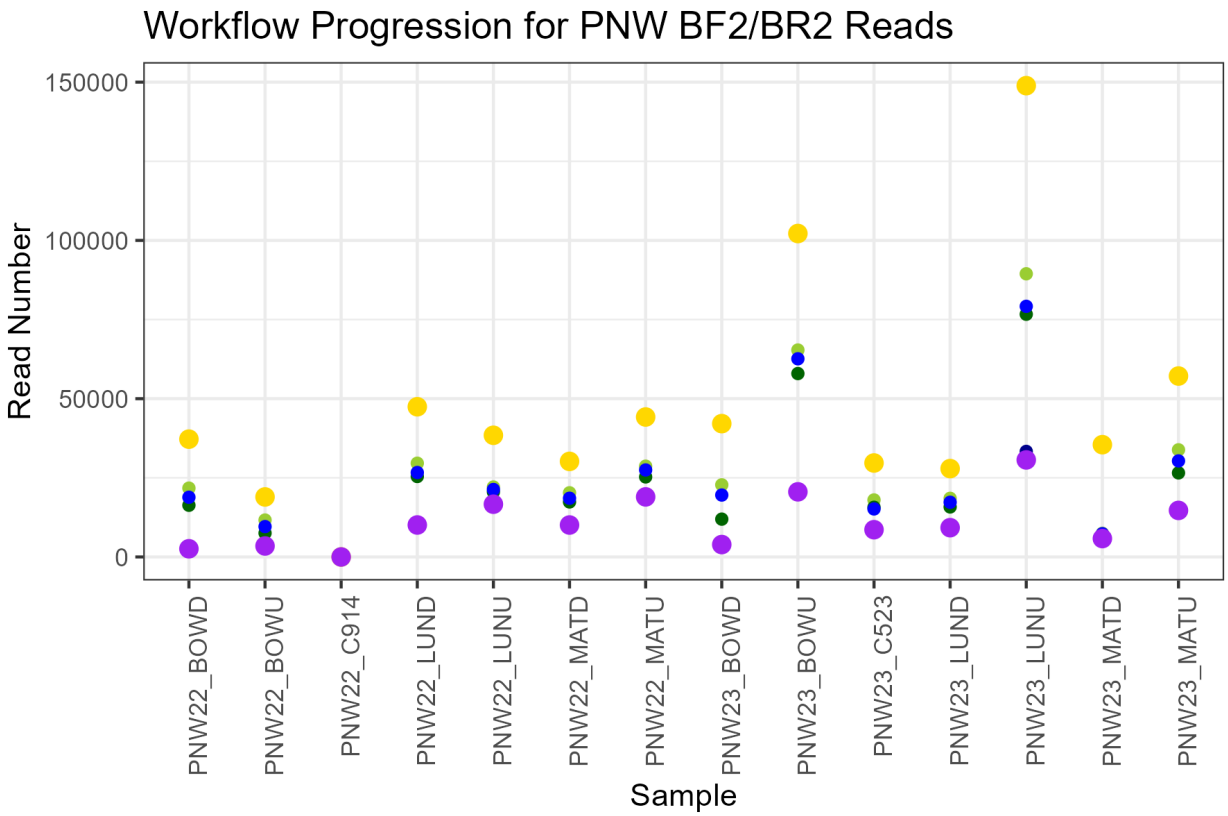

Figure 2.4 Workflow read trends for PNW COI samples. Display of the total reads found in 12S samples before dada2 processing (gold, top dot) and after (purple, bottom dot) finishing the dada2 workflow with removal of chimeras. Other points refer to read numbers after filterAndTrim step (light green), denoised reverse reads (blue), and merging forward and reverse reads (dark blue).

A single COI sample was also removed, though this sample was the negative control (PNW22_C914), with the remaining 13 samples containing between 2,584 and 30,655 reads (Figure 2.4). After structuring results and metadata into phyloseq objects and performing decontamination, the 12S data contained 128 identified OTUs across 11 samples (2 samples being removed during decontam), and the COI data contained 3803 OTUs across 12 samples (with 1 sample removed during decontam).

The COI dataset identified 6 phyla of bacteria, algae, or fungi that made up the overwhelming majority of detected and identified OTUs (Figure 2.5). After removal of these non-target taxa, only 22 OTUs remained within 7 samples.

Figure 2.5 Class identifications of taxa detected in cleaned PNW COI samples. Proportional depiction of taxonomic classes detected in PNW COI samples after all cleaning procedures and removal of contaminants and non-target taxa were completed.

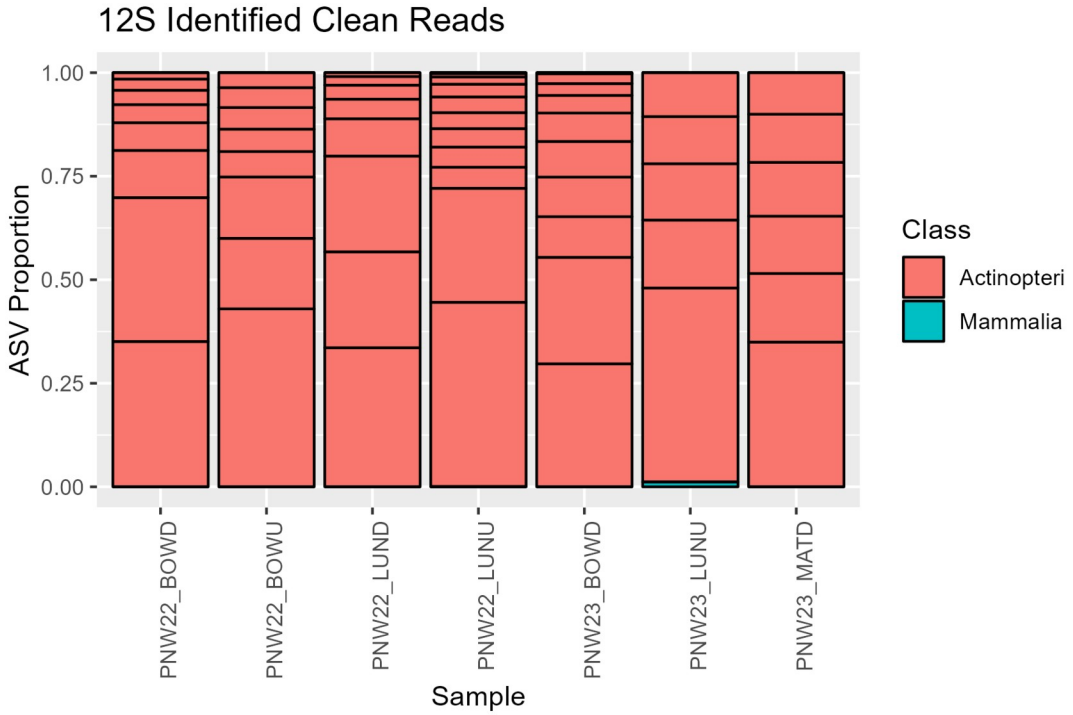

Figure 2.6 Class identifications of taxa detected in cleaned PNW 12S samples. Proportional depiction of taxonomic classes detected in PNW 12S samples after all cleaning procedures and removal of contaminants and non-target taxa were completed.

The 12S dataset experienced similar, though less drastic, alterations. Two non-target phyla were detected in 12S, Chlorophyta and Pseudomonadota, the removal of which left 7 samples containing 26 OTUs (Figure 2.6). Rarefaction found minimum ASV counts for 12S and COI at 9040 and 2 ASVs respectively, and maximum ASV counts at 36,027 and 79 Svs (Figures 2.7-2.8).

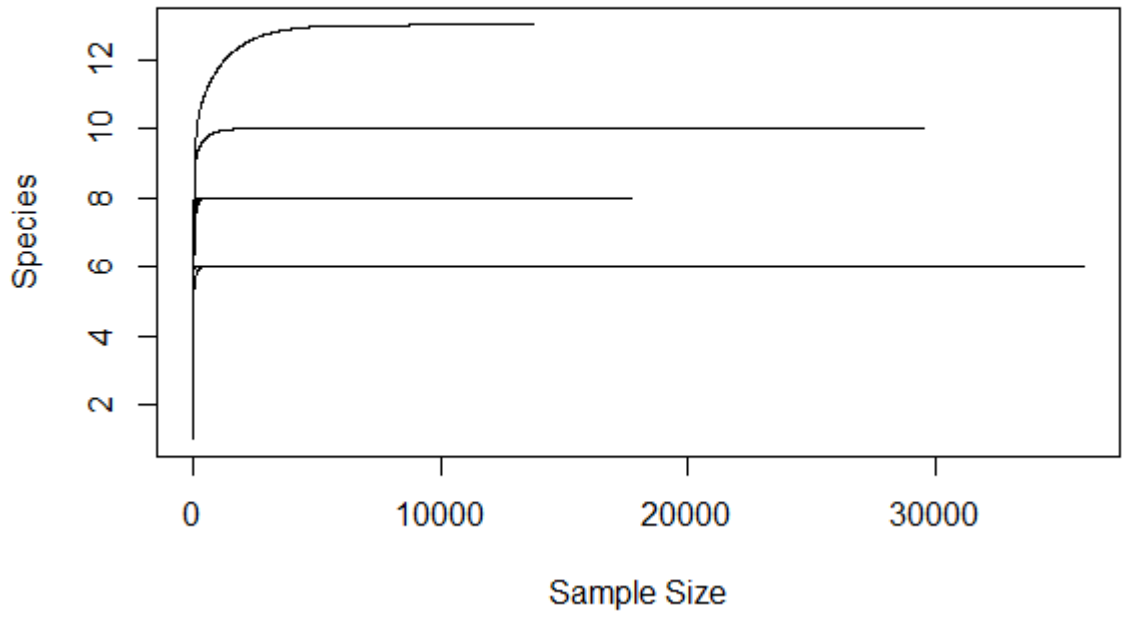

Figure 2.7 Rarefaction visualization of PNW 12S samples. Rarefaction plot of PNW 12S samples created after finishing the dada2 workflow. Samples that were removed due to insufficient target reads are not present on the graph.

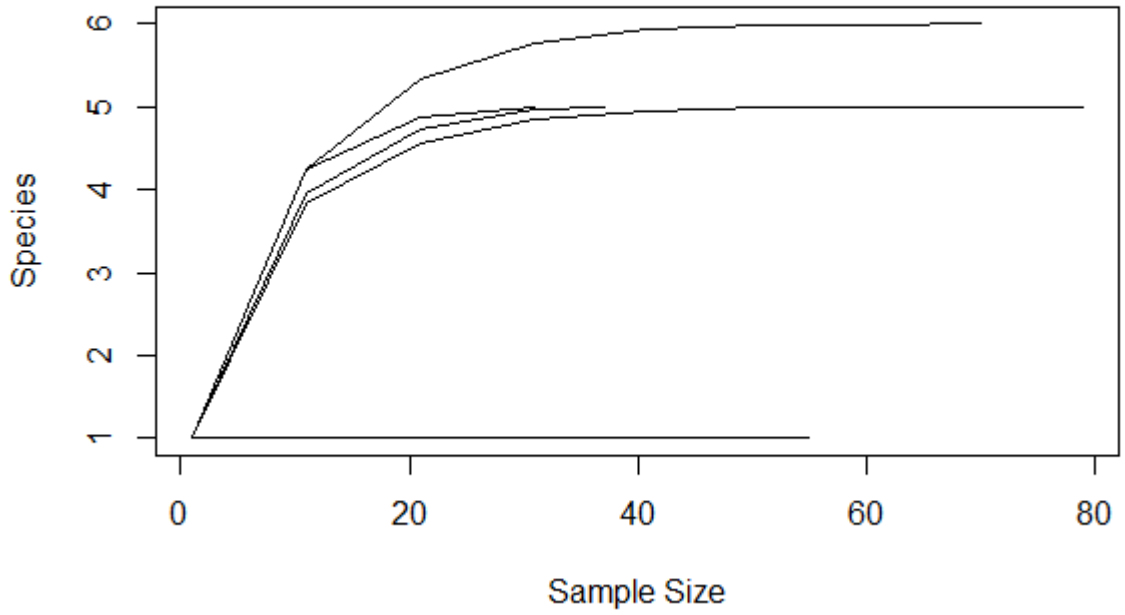

Figure 2.8 Rarefaction visualization of PNW COI samples. Rarefaction plot of PNW COI samples created after finishing the dada2 workflow. Samples that were removed due to insufficient target reads are not present on the graph.

2.3.2 Gamma and Alpha Diversity

The SE survey identifies a total of 43 organismal categories for surveyors to compare specimens against (Figures A.3-A.4), with name clarity ranging from "Other Snails", which could only be clarified to the class Gastropoda, to "Dobsonfly and Alderfly", the former of which is listed as the sole member of the genus *Corydalus*.[37] Overall, these categories could be clarified into 28 unique taxonomic ranks, mostly to the family level. In 2022, SE surveys recorded 28 organismal categories across all three sites, with the category "Fingernail Clam" added in the field under Tolerant organisms. In 2023, 32 categories were recorded across the three sites, with the addition of "Clams", also under Tolerant organisms.

With eDNA, the 12S primer recorded 16 species across 10 families involving 20 unique OTUs for the 2022 sites, and 13 species across 7 families involving 16 unique OTUs for 2023 (A.Figure 2.9). As no benthic macroinvertebrates were detected by 12S, it will not be considered further. Conversely, the COI primer found more species in 2023, with only 6 unique species in 5

families involving 6 unique OTUs for 2022 and 15 species in 10 families involving 13 unique OTUs in 2023 (Figure 2.9).

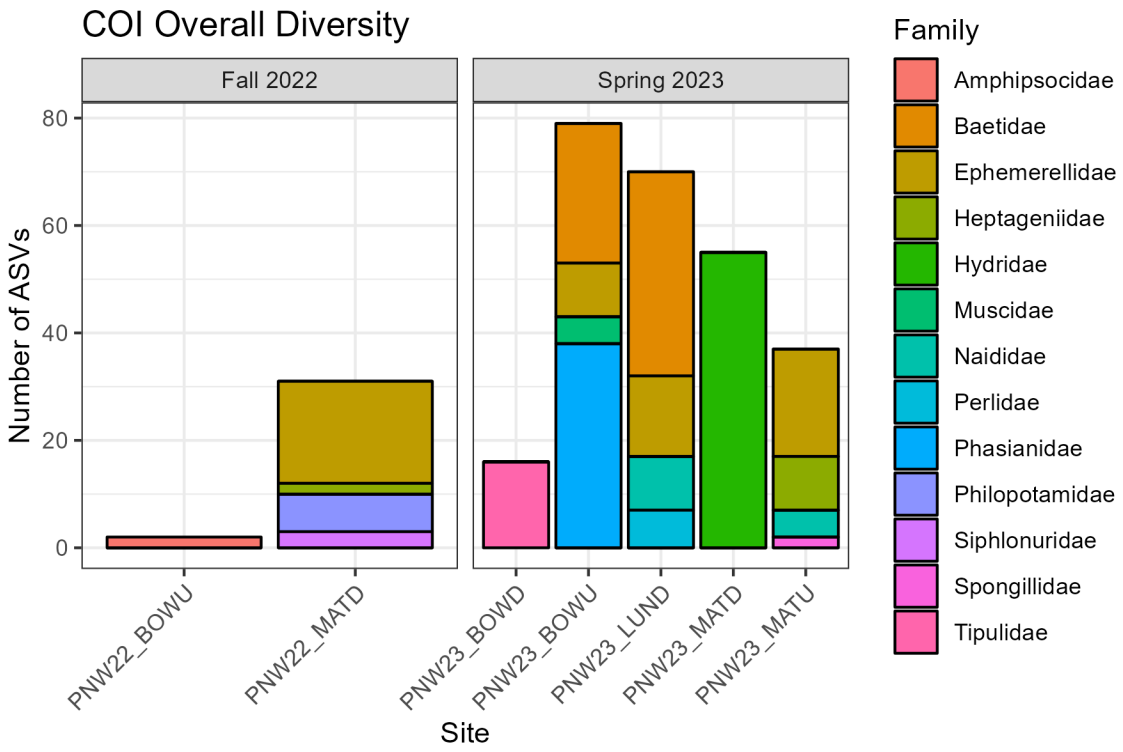

Figure 2.9 Diversity and ASV abundance of families detected in PNW COI samples. Stacked bar plots of the number of ASVs found in each cleaned COI sample as assigned to family rank, divided by sampling season.

Table 2.1 Calculated alpha diversity metrics for Stream Explorers data. Values found for the given alpha diversity metric for each site surveyed with Stream Explorers, calculated using the vegan package.

Stream Explorers Site	Chao1	Shannon	Simpson
Lunksoos 2022	27	1.82	0.75
Lunksoos 2023	28.25	2.4	0.88
Bowlin 2022	23.13	2.28	0.84
Bowlin 2023	26.29	2.59	0.9
Matagamon 2022	21.5	1.93	0.75
Matagamon 2023	26.13	2.55	0.9

COI eDNA results are less consistent, with year comparisons made difficult by an uneven number of samples from an already small sample pool that have reached this point in analysis. Only 2 samples from 2022 and 5 samples from 2023 persisted for analysis, though both 2022

samples have 2023 corollaries. Bowlin upstream (BOWU) experienced a sharp increase across all metrics from 2022 to 2023 while Matagamon downstream (MATD) decreases, once again across all metrics (Figure 2.10). In general, eDNA samples report lower richness and diversity than the SE survey results for most metrics.

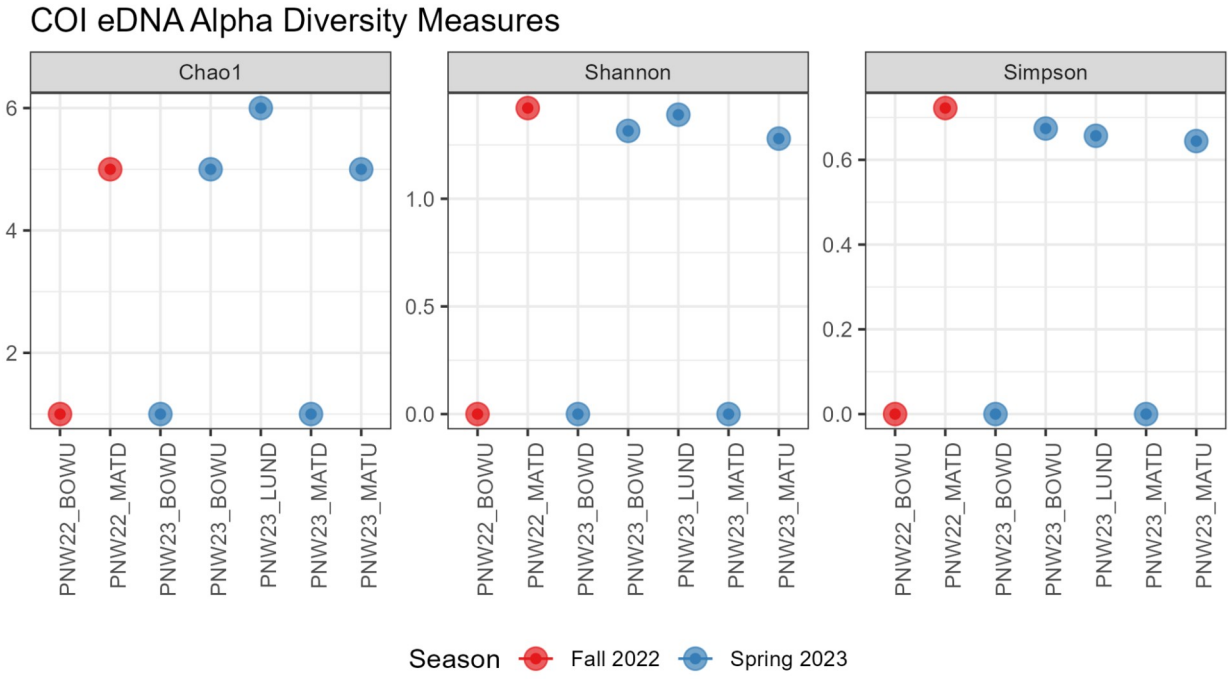

Figure 2.10 Plot of alpha diversity metrics conducted on PNW COI samples. Depiction of Chao1, Shannon, and Simpson diversity values for cleaned PNW COI samples, colored by sampling season.

2.3.3 Beta Diversity and Abundance Correlation

All sites shared taxa for both eDNA and SE survey methods, though the degree of relation and amount of shared taxa varied. Using Jaccard's similarity index for presence-absence, SE survey results range from 0.42 to 0.67 similarity (Table 2.2); when expanded to include the SE count data, Bray-Curtis dissimilarity ranges from 0.19 to 0.64, with no strong correlations between year or reach for either similarity or dissimilarity (Table 2.3).

Table 2.2 Jaccard Similarity scores for Stream Explorers observations. Measures of site similarity based on the presence-absence of organismal categories, as recorded by Stream Explorers surveys.

	Lunksoos 2022	Bowlin 2022	Matagamon 2022	Matagamon 2023	Bowlin 2023
Bowlin 2022	0.42				
Matagamon 2022	0.46	0.67			
Matagamon 2023	0.54	0.5	0.59		
Bowlin 2023	0.43	0.55	0.48	0.66	
Lunksoos 2023	0.46	0.59	0.58	0.64	0.64

Table 2.3 Bray-Curtis Dissimilarity scores for Stream Explorers organism counts. Measures of site dissimilarity based on recorded counts for each organismal category found with Stream Explorers surveys.

	Lunksoos 2022	Bowlin 2022	Matagamon 2022	Matagamon 2023	Bowlin 2023
Bowlin 2022	0.32				
Matagamon 2022	0.77	0.7			
Matagamon 2023	0.71	0.68	0.59		
Bowlin 2023	0.63	0.58	0.78	0.69	
Lunksoos 2023	0.6	0.59	0.77	0.54	0.54

There were similar results found with the eDNA data - while visual patterns appeared to be present in both Jaccard and Bray-Curtis ordinations (Fig 2.11), the only significant ($F < 0.05$) relationship was found in the 12S data where the stream reach had a significant ($F = 0.005$) impact on community composition. With so few identified species and families in the eDNA data, it is not surprising that there was low overlap in detected families with the SE data. Of the 27 total families detected in eDNA, only Ephemerellidae, Heptageniidae, Perlidae, Siphlonuridae, and Tipulidae were also included in the SE survey sheet. Where counts for SE-surveyed individuals and ASVs were available (Figure 2.12), no relationship was found between either abundance. Additionally, there were multiple cases where no ASVs were detected even though specimens were observed, and vice versa. On 3 occasions eDNA was able to identify ASVs from 2 separate species in the family, all other ASV counts came from only a single identified species. Plotted dissimilarity between all categories observed through the SE survey and the COI eDNA detection results (Fig 2.13) shows no similarity between survey results. To confirm this, a Pearson's correlation found insignificant correlation with a p-value = 0.14.

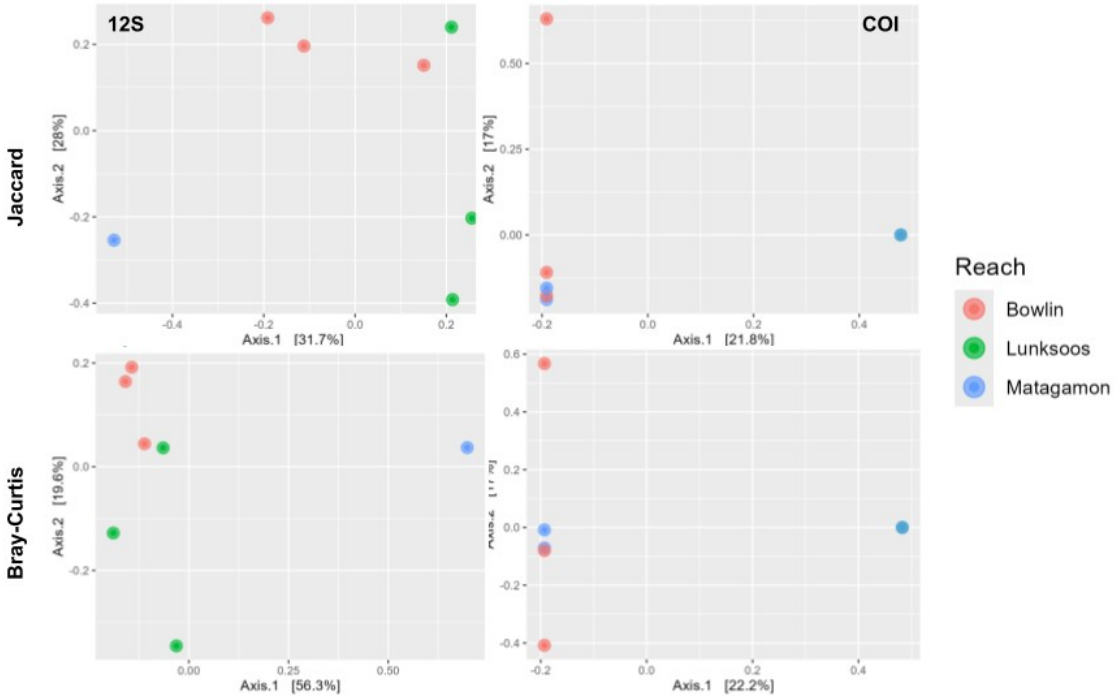

Figure 2.11 Ordination results for PNW eDNA samples. PCoA ordination plots using Jaccard presence-absence similarity (top row) and Bray-Curtis abundance dissimilarity (bottom row) for 12S (left column) and COI (right column) samples. Colored by the stream reach of origin.

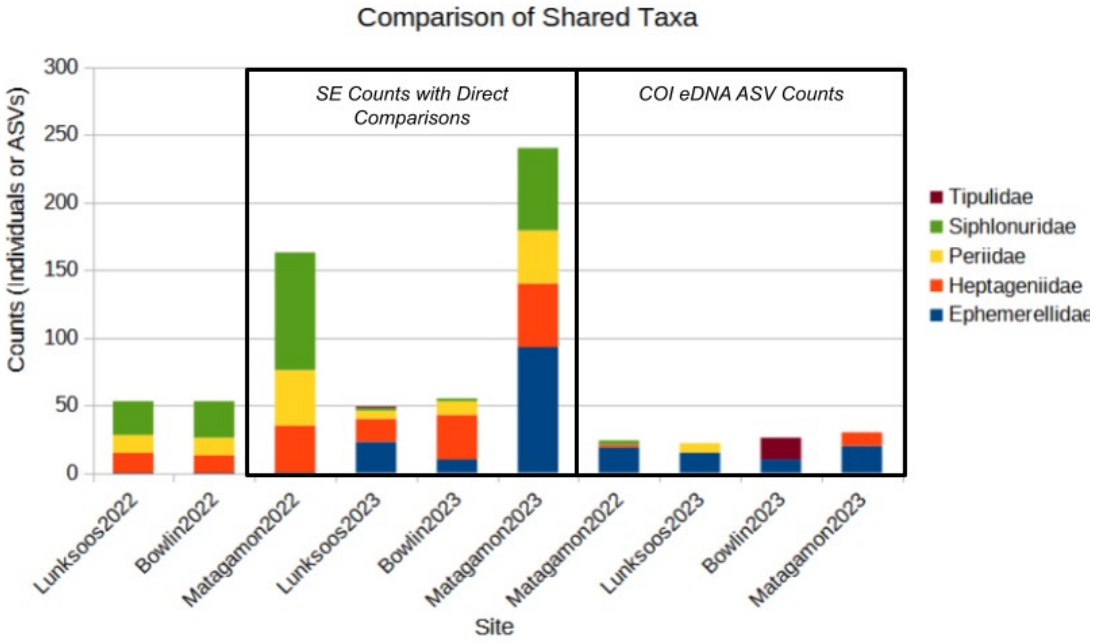

Figure 2.12 Shared taxa and abundances for SE and COI eDNA. Stacked bar plot of organism counts and ASV abundances for individuals and species belonging to families detected by both methods

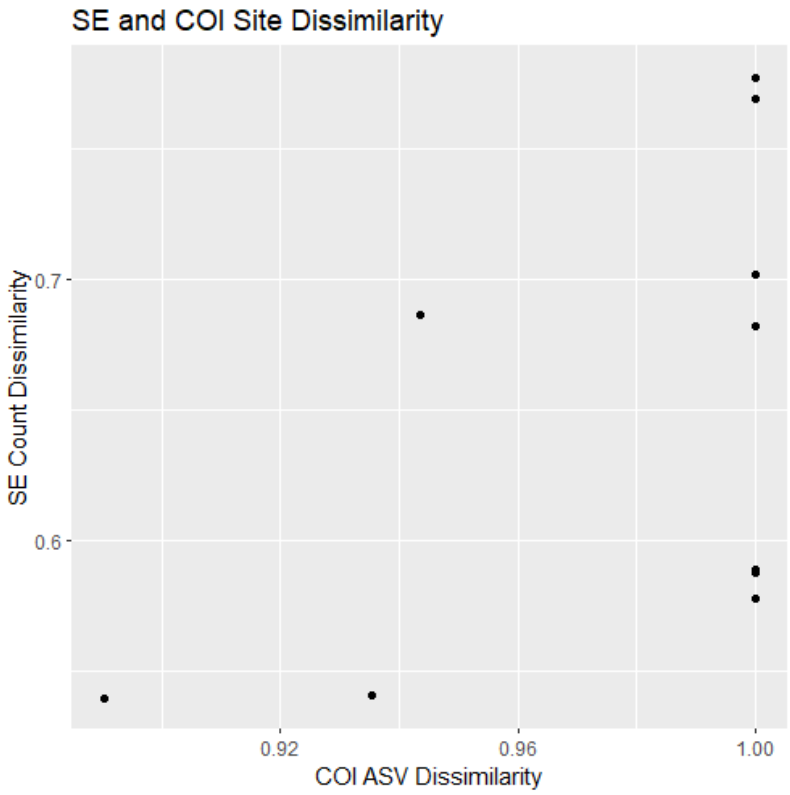

Figure 2.13 Dissimilarity plot for SE and COI eDNA. Similarity plot of count results for each method.

2.4 Discussion and Conclusions

eDNA detected highly variable numbers of reads for both the 12S and COI datasets, though both experienced severe reduction in the final number of reads after cleaning and taxa-removal steps were performed. COI samples consisted primarily of non-target taxa, with counted OTUs decreasing from 3803 to only 25 after taxa removal, raising concerns that amplification of non-target taxa may have drowned out target taxa that may have been detected had a more selective primer been used.[12,18] Despite the reduction, the cleaned COI dataset included more taxonomic classes than were detected with 12S - likely another side effect of the broad detection ability of COI primers. Overall the two primers were successful in identifying their primary target organisms, with COI identifying benthic macroinvertebrates including insects, annelids, sponges, and hydrozoans as well as birds, and 12S identifying bony fish as well

as mammals. No macroinvertebrates were detected with 12S, ruling it out of current consideration as a single primer effective for both fish and invertebrate monitoring.

While the resolution and taxonomic variety of eDNA exceeded that of the Stream Explorers survey, the actual taxa of biomonitoring interest were under-represented or missed entirely in eDNA samples. Only 5 families were found to overlap between the SE and eDNA data. Whether this gap is due to unrecognized ASVs from a non-localized or incomplete reference library, from failure to detect valid sequences, or overwhelming detection of non-target taxa is unknown. It is likely that the cleaning workflow itself removed valid reads that were only present in short sequences or low numbers. This false negative error has been reported in multiple other studies, as cryptid reads may contain the same characteristics as error or low-quality reads.[1,10,12,21] This is likely also the source of the lower Chao1 values for the eDNA datasets, as the singletons and doubletons that the metric relies upon for calculation were broadly removed from the datasets.

While SE data reported diversity and richness increases at all sites from 2022 to 2023, eDNA data was far more mixed, with few samples available in each dataset for year-to-year comparison. SE counted communities gave similar alpha diversity results (Table 2.1), but had higher variation in the number of each observed category. This unevenness follows both the recorded data and general biodiversity observations that most sites involve a few common species and a larger number of rare species,[45] with the common and rare species changing depending on site. eDNA data showed visual patterns where each reach generally appeared to cluster closer together, or along the same axis, than samples from other reaches. The strong dissimilarity of results between the two survey methods indicates that each method will provide highly different views of community composition at sites, likely impacting management decisions depending on which method is chosen. Further testing would be needed to determine if these disparate results support the use of multi-method surveys to complement detection and observation results, or if they are due to complications in eDNA sampling or processing design.

In terms of sampling effort, both the SE survey method and eDNA collection were fairly rapid. The SE survey was complete at each site after a 15 minute sampling period and 45 minutes to an hour for specimen identification, and eDNA samples took less than 10 minutes to complete at each site. The SE survey had the benefit of near-immediate results as specimens were identified and could be returned with minimal destruction to the stream of origin. On the other hand, eDNA samples had to be placed on ice as the stream water was too sedimented to efficiently filter on-site, undergo multiple preparation steps, and required specialized training and equipment to process and analyze data. As a technique for habitat managers, these processing requirements will likely increase initial cost of equipment set-up and the duration of time to receive results, which may hinder large-scale sampling efforts where results are needed to inform short-term management decisions, or other time-sensitive initiatives. These delays would be exacerbated if large-scale projects require outsourcing eDNA processing and analysis to specialized labs, depending on turnaround times and lab availability.

In this study I tested the ability of eDNA to identify Maine stream taxa of interest in biomonitoring, and whether or not observed count data showed a relationship with metabarcoding ASV numbers. While the basic ability of eDNA to detect and identify taxa of interest was confirmed, low sample numbers, low detected OTU and ASV numbers, and difficulty in equating taxonomic ranks to organismal categories of interest restrict the ability to make one-to-one comparisons between eDNA and the Stream Explorers survey method. It is possible that with more sampling periods, stronger patterns or associations could be found; however given the already low number of reference sequence matches for Maine species found in Chapter 1 and the high number of non-target taxa detections, focus should also be placed on creating Maine-localized reference sequences and primers. Another significant limitation for this comparison is the lack of fish count data to check against 12S detections. Though benthic macroinvertebrates were the primary focus, the significance between 12S detected ASV abundance and reach would be more informative if paired fish counts were available. eDNA

remains a potential tool for biomonitoring, but more work remains to be done on a methodological scale to understand how stream transport of eDNA may impact biomonitoring conclusions, and on a local scale to strengthen the ability of Maine researchers and ecosystem managers to utilize eDNA tools effectively.

CHAPTER 3

DETECTION OF SHIFTS IN COMMUNITY COMPOSITION BEFORE AND AFTER LARGE WOOD ADDITION RESTORATION USING AQUATIC ENVIRONMENTAL DNA

3.1 Introduction

Stream restoration often focuses on habitat alterations that target returning a stream reach or habitat area to an earlier or improved state[38-39], typically to increase habitat complexity and/or biodiversity to levels prior to anthropogenic disturbances. Most streams and rivers in Maine and the eastern US are impacted by historic logging activity or other deforestation.[29,38-42] Logging, including use of streams for log driving[38,40], removed mass amounts of instream and bank wood from stream habitats[39] in the 19th and 20th centuries[40]. The removal of instream wood in particular has been strongly linked to the degradation of stream channels[38,41] and fish habitat[40], as well as increased erosion[29,38] and erratic flow fluctuations.[29,41] In combination with clearcutting, urbanization, and deliberate straightening of streams to assist in human infrastructure[38-40], the effects of historic logging remain to impact streams decades after direct logging activity has ceased.[40-41]

Large instream wood, defined as wood pieces more than 10 cm in diameter and 1 m in length,[41-42] serves as a physical mediator[29,38-39,,41], nutrient source[39], and habitat[29,39] in streams. The presence and retainment of large wood can be caused by natural[38-40] or artificial causes,[29,40-42] and impacts stream flow,[29,39,41] retainment of substrate,[29,39] bank stability,[41] and available habitat for fish and macroinvertebrates.[29,39] Though not all disturbed streams are restricted by large wood,[40] higher amounts of large wood has been linked to higher biodiversity,[29,40] lower rates of erosion,[39,41] and increased bank stabilization.[41]

In the northeastern US, where many streams are still impacted from the legacy of logging activity,[38-40] the absence of large wood has been targeted as a goal for stream restoration via

large wood addition (LWA) projects.[29,38-42] LWA projects involve the physical addition of downed trees, logs, and other woody debris to increase the amount of instream large wood,[29,39-42] however the success of these additions have been mixed. Projects installed in warm areas[29] or in wide streams[29,41] have experienced difficulty in retaining their LWA over time as warm conditions increase the rate of wood decay and wider stream channels make anchoring systems difficult.[29] While other projects have recorded increases in fish biomass and size,[29,40] the sampling effort and difficulty to access target streams limits the frequency and thoroughness with which follow-up monitoring surveys can be conducted. [29] The ability to conduct monitoring is critical to understanding if and how an LWA is impacting the stream and its biodiversity.

The Appalachian Mountain Club (AMC) is a conservation and recreation organization committed to active conservation, research, and education of the environment, and maintains the Maine Woods Initiative (MWI).[43] The MWI is a multi-use recreation and conservation project in the 100-Mile Wilderness and includes historically logged areas targeted for restoration through LWA.[43] To monitor community changes before and after LWA, AMC invited and supported us to collect eDNA samples on stream reaches in the MWI. Most of the sampled reaches had undergone little to no previous biodiversity sampling due to the difficulty of access and impracticality of specimen-based survey method (SBSM) sampling effort, and so our focus with eDNA is restricted to presence-absence detection. The primary research interests were to examine if there were any measurable shifts in community composition before and after LWA installation and if there were detectable differences in community composition in stream communities that had undergone LWA and those that had not.

I hypothesize that stream reaches that have undergone LWA will have higher alpha diversity and be more similar communities to each other than reaches that have not undergone LWA, and that reaches without LWA will have lower alpha diversity. I further hypothesize that reaches downstream of LWA will have more similar communities as detected by eDNA to reaches containing the LWA than to reaches upstream or without LWA.

3.2 Methods

3.2.1 Sampling and Filtration

eDNA samples were collected in August of 2022 and July of 2023, with near-identical collection and processing protocols but different sampling regions. Sampling sites in 2022 were divided between sites that would offer a view of large wood additions (in the east of the 100 Mile Wilderness) and sites aimed at collecting baseline survey data for the Gulf Hagas and western 100 Mile Wilderness areas (which will not be discussed in this thesis beyond workflow verification). 2023 sites were focused only at examining LWA and did not repeat or include any sites in the Gulf Hagas area.

eDNA was chosen as the sole survey method for all sites due to access and available time, and several sites were removed from consideration as weather and accessibility challenges occurred. A total of 31 sites had samples collected in 2022, and 25 sites in 2023, with 11 sites sampled in both 2022 and 2023 (Figure 3.1). LWA sites in both years included stream reaches that had already had LWA installed between 2020 and 2022 ("Post" sites), reaches that were targeted for 2022 LWA installment ("Pre" sites), and streams nearby with no LWA history or plans ("Control" sites). Each year of eDNA sampling occurred before LWA installation for that year. Sites along the same reach were also divided into Position (or LWAPosition), the spatial relationship of the site in regards to the LWA on the reach - sites upstream of the LWA are listed as "Upstream", sites downstream of the LWA listed as "Downstream", and sites located inside the LWA area listed as "Internal." In cases where two sampled streams of the same order merged, and only one stream experienced LWA installation, sites along the non-LWA stream are listed as "Parallel". The total number of sample sites belonging to each category are available in Table 3.1.

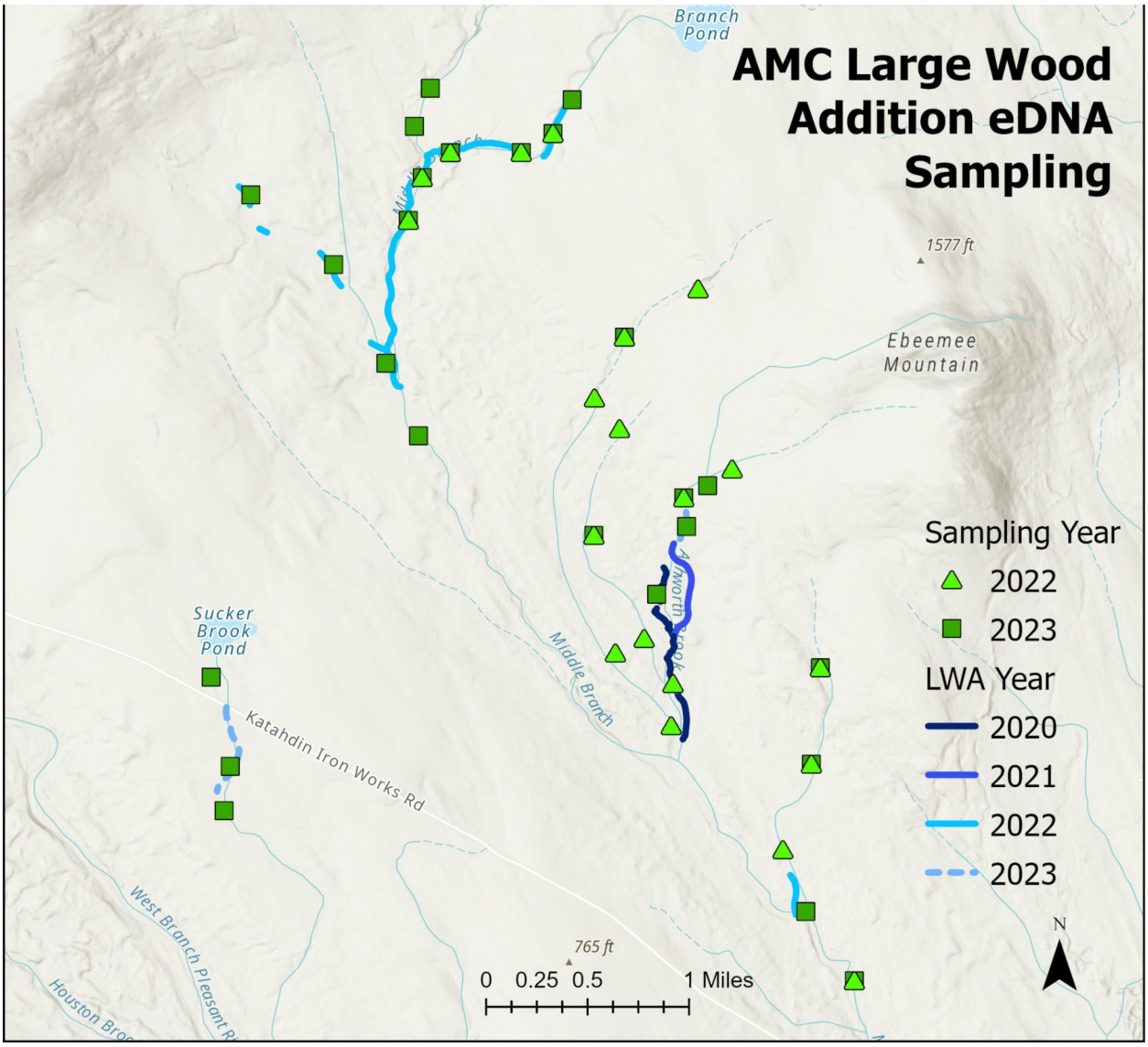

Figure 3.1 Map of eDNA sample sites and LWA installations in the LWA survey area. Map showing the collection sites for eDNA samples in 2022 and 2023 with LWA installation extent. LWAs installed in 2023 (dashed line) were installed after 2023 eDNA collection.

Table 3.1 Site numbers by category. Breakdown of site types for each sampling year

Year	Pre-LWA	Control	Post-LWA	Downstream	Internal	Upstream	Parallel
2022	9	8	3	1	6	5	8
2023	6	4	15	7	11	7	4

At each sampling site, eDNA samples consisted of three subsamples of one liter each for a total of three liters of stream water used to sample eDNA at each site. Subsamples were collected midstream just underneath the stream surface, midstream 7-10 cm above the stream substrate, and on the side of the stream near a bank just underneath the surface (Figure 3.2). This placement of samples was chosen to capture any microhabitat variation in eDNA caused by either species' stream use or differing flow rates.

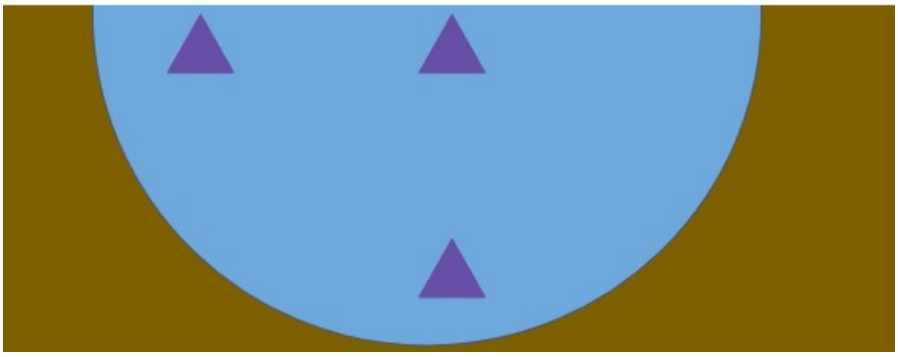

Figure 3.2 Diagram of eDNA subsample collection plan. Not-to-scale depiction of the locations of the surface side, surface middle, and bottom middle subsamples relative to a given stream cross-section.

Each subsample was either filtered in the field using a hand pump or placed on ice and transported to the lab for vacuum filtration within 24 hours of collection, with site subsamples aggregated by filtering onto the same filter paper(s). For each day of sample collection, one liter of DI or other sterile water was filtered first on the filtering equipment to serve as a negative control for that day, and was carried along with the field team until all other samples for the day were collected.

3.2.2 Water Filtration

Filtration of eDNA samples utilized single-use 47 mm 0.45 µm cellulose nitrate filter papers. All lab surfaces and reusable tools were sterilized by either soaking in 10% bleach solution for at least 10 minutes and drying fully before use, or applying bleach wipes and allowing the bleach to stand for 10 minutes before applying a DI water rinse. Before and after each use of the lab for filtration, a germicidal UV light was run for 30 minutes to an hour. In 2022, filter papers were frozen in absolute ethanol at -20°C before extraction. These samples required additional filtration to remove ethanol, which was followed by letting samples evaporate in the cleaned lab in sterilized petri dishes to remove any remaining ethanol. Afterwards samples were frozen at -20°C until extraction. In 2023, samples were simply frozen in DNA LoBind tubes at -20°C before extraction, and did not require ethanol removal. When a site required use of multiple filter papers due to high sedimentation in the subsamples, all filters for the site were kept within the same tube and physically aggregated during extraction. For each round of filtration, at least one negative control sample was filtered alongside site samples for use as a laboratory blank.

3.2.3 eDNA Extraction

eDNA extraction was performed on the filter papers using the Qiagen DNA Powersoil Pro kit, using the same protocol as in Chapter 2 that was developed by Geneva York with some alterations developed by Kylie Holt. All surfaces were wiped down with bleach wipes followed by DI water before use, and pipette tips and other sensitive equipment cleaned with DNA-Off. Other reusable materials such as forceps and tube racks were cleaned by soaking in 10% bleach for 10 minutes, rinsing five times, then allowed to dry fully before use. An hour of germicidal UV exposure was performed in the lab before and after each round of extraction, and extraction times were chosen to minimize overlap with other projects' lab usage. At least one negative control sample was extracted in each round of extraction to continue serving as a laboratory

blank as well as field blank. Extracts were eluted to a 100 µL volume, then frozen at -20°C before PCR and sequencing.

3.2.4 PCR and Sequencing

For each primer, samples underwent polymerase chain reaction (PCR) and sequencing, with each batch including at least the original field blank for continued use as a negative control. The PCR protocol for both primers used a 25 µL reaction with 9 µL of nuclease-free MB grade H_2O, 1.25 µL of each forward and reverse primer, 12.5 µL of Quantabio HiFi ToughMix 2X, and 1 µL of extracted DNA.

The 12S primer pair involved a 21 bp forward primer (GTCGGTAAAACTCGTGCCAGC) and a 27 bp reverse primer (GTTTGACCCTAATCTATGGGGTGATAC) to amplify a region approximately 170 bp long.[5,14] Both 12S primers were designed with the addition of the Nextera adapter for MiSeq sequencing. PCR for the 12S samples underwent a protocol of 98°C for 10 seconds, 61°C for 5 seconds, and 68°C for 1 second, repeated for 38 cycles. Samples were refrigerated at 20°C until sequencing.

The COI primer pair involved a 20 bp forward primer (GCHCCHGAYATRGCHTTYCC) and a 20 bp reverse primer (TCDGGRTGNCCRAARAAYCA) to amplify a region approximately 420 bp long.[12,18,24] Both COI primers were also designed with the addition of the Nextera adapter to prepare for sequencing. PCR for the COI samples underwent a protocol of 95°C for 10 minutes and a 35-cycle repeat of 95°C for 30 seconds, 50°C for 30 seconds, and 72°C for 30 seconds, with a final 72°C for five minute stage. Samples were refrigerated at 20°C until sequencing.

Library prep and sequencing were performed using Illumina Miseq at the University of Maine CORE DNA Sequencing Center by Geneva York and Lindsey Stover. Given the large difference in segment length, 12S and COI samples were sequenced on different runs, though both ran for 300 cycles. Once sequenced, data was sent as FASTA files with the Nextera adapter removed.

3.2.5 Bioinformatic Workflow

All samples were cleaned and prepared for analysis using the dada2 workflow[34,47] in R version 4.3.1, followed by use of the phyloseq[35] and vegan[36] R packages. To remove low quality reads, quality profile plots were first created for a random sample of both the 12S and COI datasets to guide trimming. For 12S samples, reads less than the default of 20 bp were removed, and forward and reverse reads were trimmed between the primer length (21 bp for forward and 27 bp for reverse, respectively) and 220 bp in raw length. COI samples were trimmed to remove the 20 bp forward and reverse primers, and to a total length of 280 and 250 bp for the forward and reverse reads, respectively. Similarly to the 12S samples, reads less than 20 bp were removed. Both datasets used maxEE = 2 and maxN = 0, and default dada2 settings were used for learning errors. Once error rates were learned, sample inference, merging of the forward and reverse reads, and removal of chimeras were performed.

The trimmed data were then converted into phyloseq objects, and used assignTaxonomy to identify ASVs against the respective 12S and COI Maine reference libraries created in Chapter 1. Any ASVs present in the negative controls were removed from all samples as contamination, and non-target bacterial, fungal, and algal taxa were removed. A check was made to ensure no human DNA was present in the samples and phyloseq objects were limited to only the sites in the LWA region.

3.2.6 Gamma and Alpha Diversity

Gamma diversity was examined briefly through the species and taxonomic families found in each year of sampling in the LWA region. Alpha diversity metrics utilized Chao1, Shannon, and Simpson measures, and non-parametric ANOVAs were performed to test the significance of LWA treatment type on alpha diversity metric results. Rarefaction was performed

to visually confirm if the number of sequence reads was sufficient to capture diversity present in samples. No samples were removed based on rarefaction results.

3.2.7 Beta Diversity

Beta diversity was calculated as Jaccard's similarity index for presence-absence, focusing on the impacts of LWA treatment type (Pre, Post, or Control) and LWA position (Downstream, Internal, Parallel, or Upstream). PERMANOVAs were conducted to test significant relationships between distance, LWA treatment type, year of collection, and LWA position. Additionally, interactions between treatment type and LWA position, and treatment type and year were tested for significance. Dissimilarity between samples was visualized using principal component analysis (PCoA).

3.3 Results

3.3.1 Workflow Filtering

77 total samples were initially run through the dada2 and analysis workflows for both 12S and COI assignment. These included a total of 19 negative control samples, 45 LWA region samples, and 13 survey samples from the Gulf Hagas region. Of these, one negative control sample from the 12S dataset and three samples from the COI dataset were removed after the initial filterAndTrim step. The percentage of reads removed from the initial trim ranged between 4.44% and 91.86% for the 12S samples, with a median of 8.88% and an average of 10.65%. For COI, trimming removed between 9.82% and 100% of reads, with a median of 39.01% and an average of 46.10%. ASVs present in negative control samples were then removed from each dataset as assumed contaminants before final ASV results and metadata were structured into phyloseq objects for the 12S and COI datasets. This decontamination process completely removed 18 samples from the 12S dataset and 22 samples from the COI dataset due to high amounts of non-target taxa (Figures 3.3-3.4).

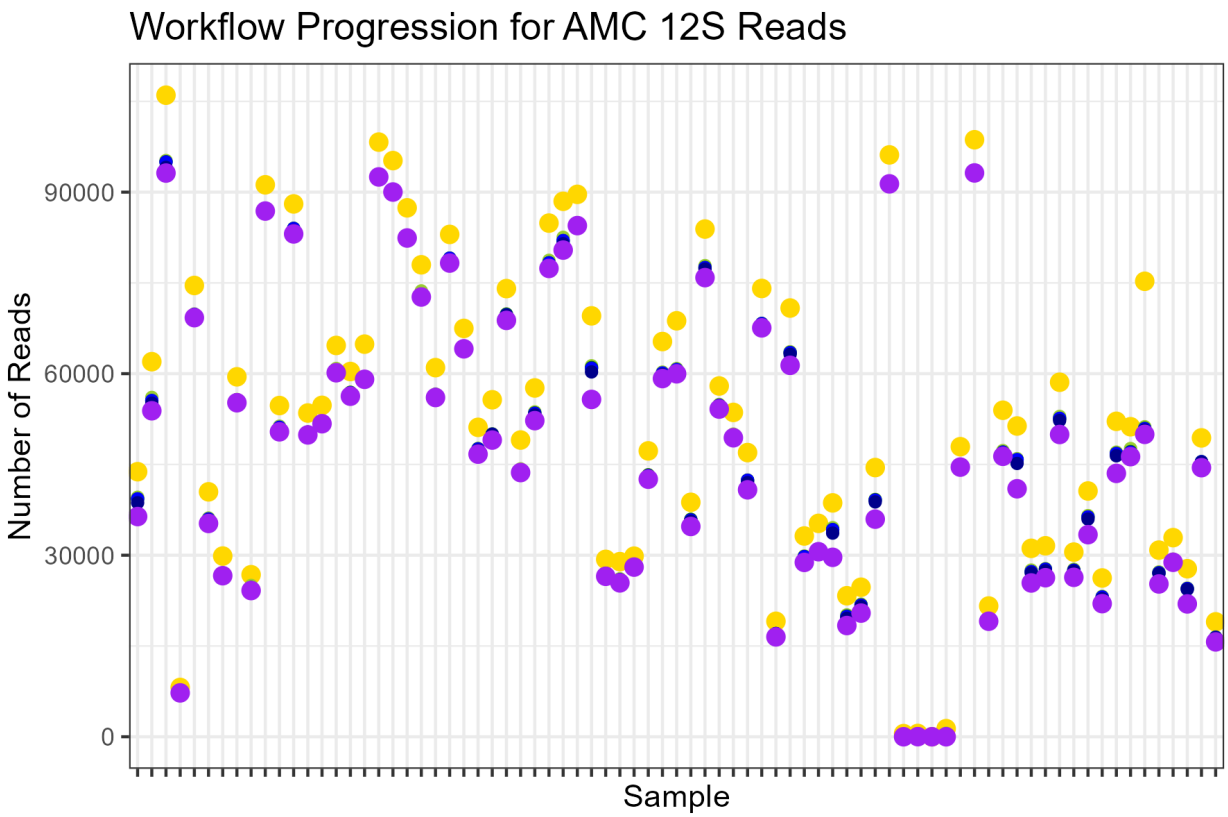

Figure 3.3 Workflow progression of 12S reads. Scatterplot of the initial (gold dot, top) number of reads for each sample and final output number of reads (purple, lowest) after completing the dada2 workflow. Intermediate dots represent read numbers after individual steps of the dada2 workflow (light green = read number after filterAndTrim, blue = read number after denoising reverse reads, dark blue = read number after merging forward and reverse reads).

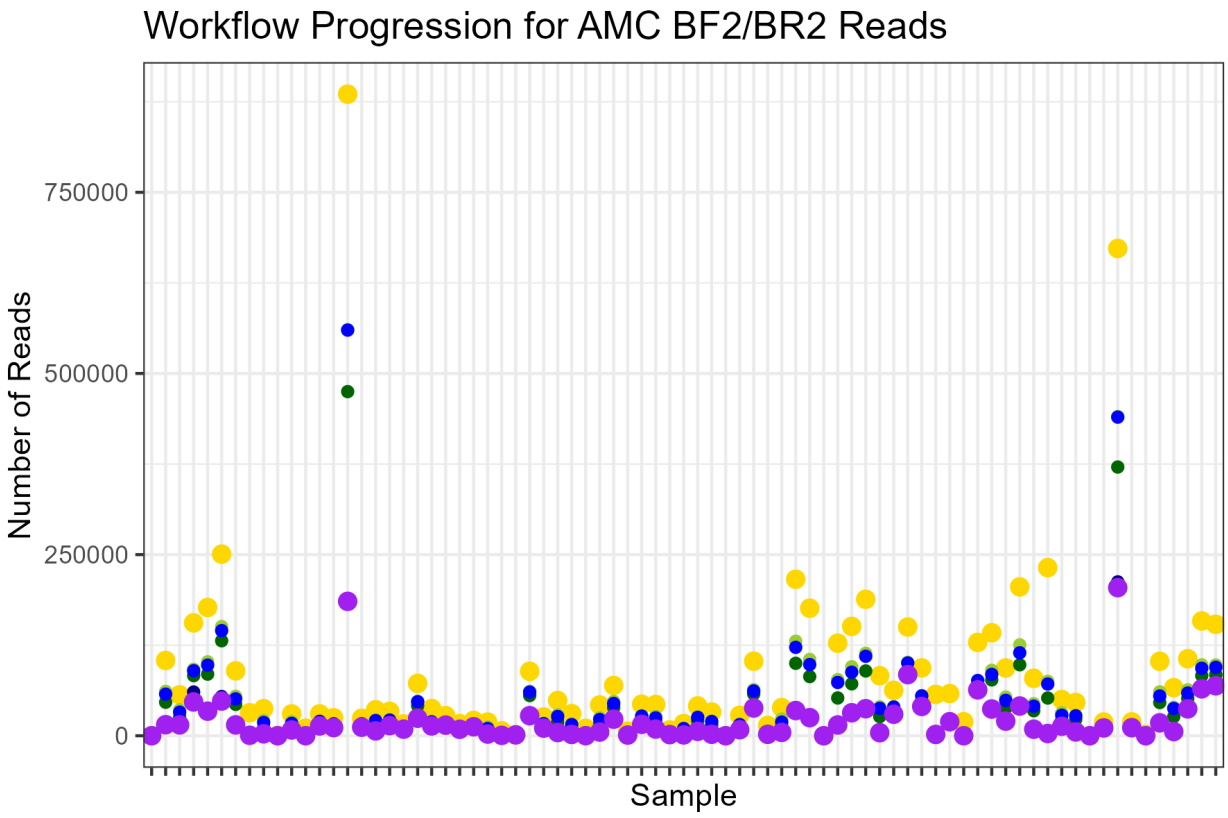

Figure 3.4 Workflow progression of COI reads. Scatterplot of the initial (gold dot, top) number of reads for each sample and final output number of reads (purple, lowest) after completing the dada2 workflow. Intermediate dots represent read numbers after individual steps of the dada2 workflow (light green = read number after filterAndTrim, blue = read number after denoising reverse reads, dark blue = read number after merging forward and reverse reads).

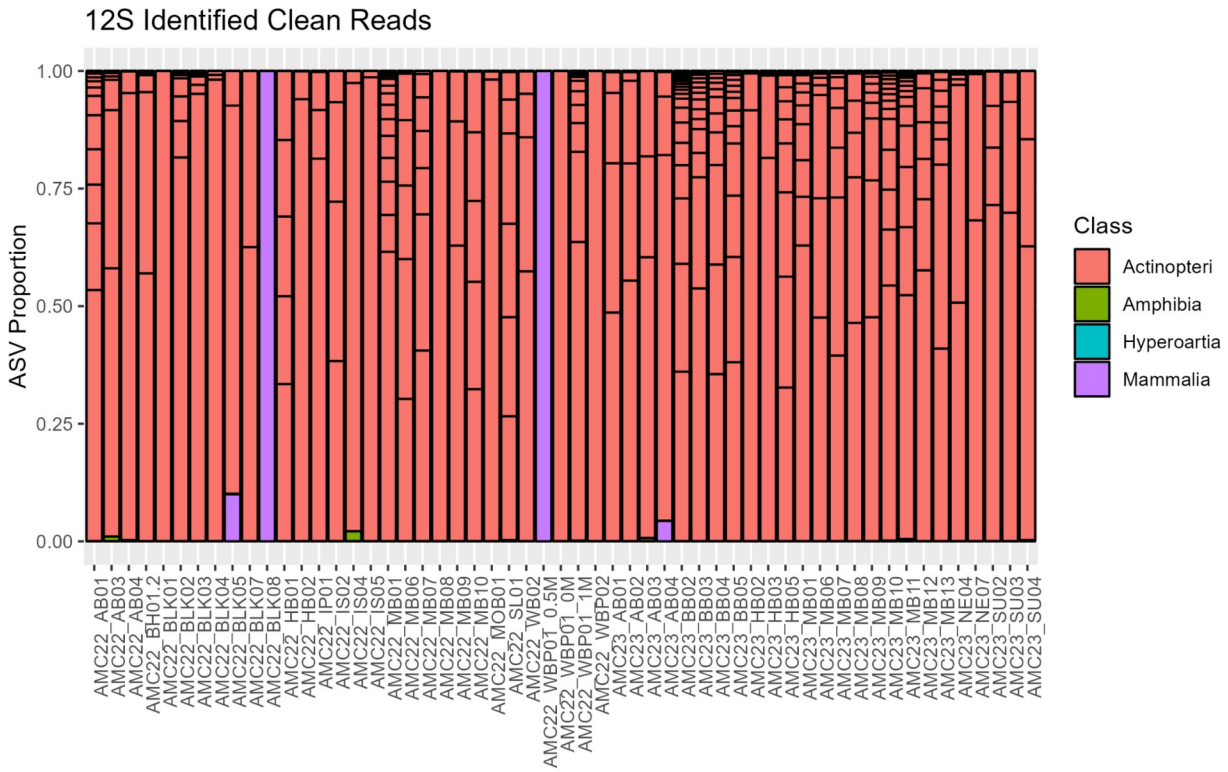

Figure 3.5 Class identifications of taxa detected in cleaned AMC 12S samples. Proportional depiction of taxonomic classes detected in AMC 12S samples after all cleaning procedures and removal of contaminants and non-target taxa were completed. Abundance on the y-axis refers to the number of ASVsidentified to each class. Samples are not region-restricted.

The post-decontamination 12S phyloseq object began with 508 OTUs across 58 samples, and the COI object with 19,025 OTUs across 52 samples. The 12S dataset identified 6 prevalent non-target phyla including bacteria, algae, and plant taxa, the removal of which resulted in only 55 samples containing 136 OTUs (Figure 3.5). The COI dataset identified 10 prevalent non-target phyla, the removal of which left only 134 taxa across 31 samples (Figure 3.6).

Figure 3.6 Class identifications of taxa detected in cleaned AMC COI samples. Proportional depiction of taxonomic classes detected in AMC COI samples after all cleaning procedures and removal of contaminants and non-target taxa were completed. Abundance on the y-axis refers to the number of ASVsidentified to each class. Samples are not region-restricted.

A final check to remove any occurrences of human DNA from either dataset was performed, and phyloseq objects were then limited to only include LWA region samples. This resulted in a final count of 113 OTUs across 43 samples for the 12S dataset and 123 OTUs across 25 samples for the COI dataset. Rarefaction found ASV numbers ranging from 7 to 74,165 for 12S OTUs and 3 to 5,307 for COI OTUs (Figures 3.7 and 3.8).

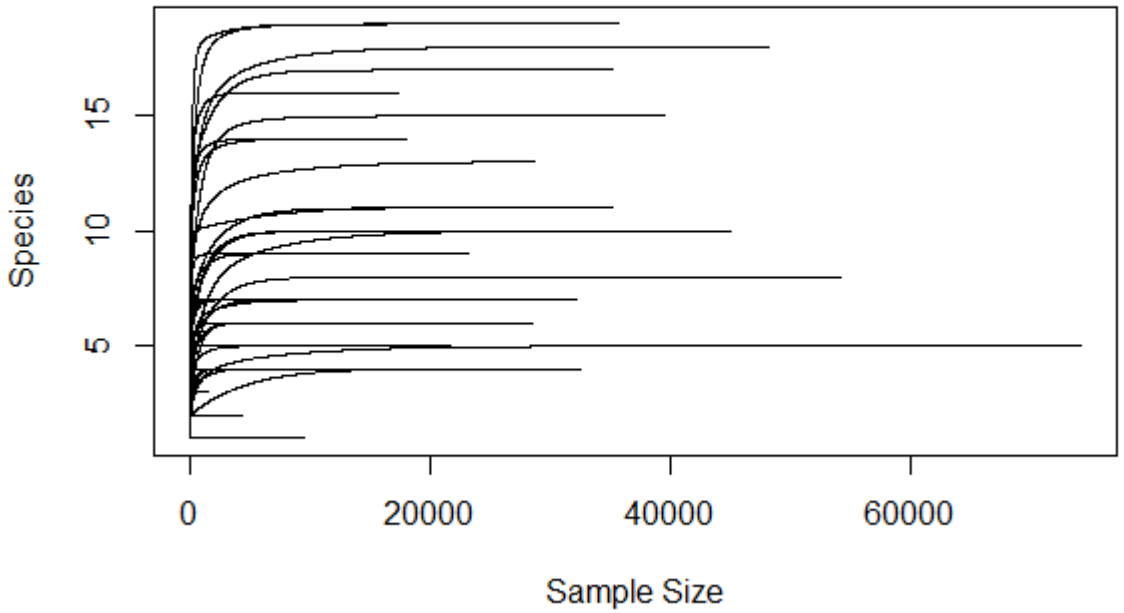

Figure 3.7 Rarefaction visualization of AMC 12S samples. Rarefaction plot for AMC sites collected as part of the LWA study, amplified with the 12S primer. Samples removed during the dada2 workflow due to insufficient target read numbers are excluded from the plot.

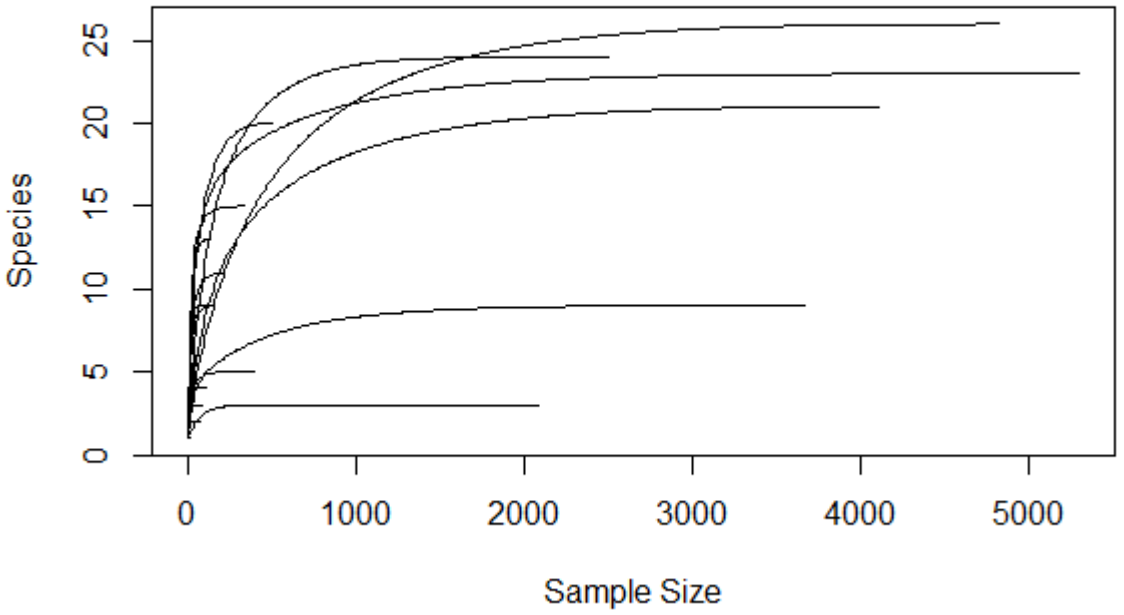

Figure 3.8 Rarefaction visualization of AMC COI samples. Rarefaction plot for AMC sites collected as part of the LWA study, amplified with the COI primer. Samples removed during the dada2 workflow due to insufficient target read numbers are excluded from the plot.

3.3.2 Gamma and Alpha Diversity

In 2022, 20 species were detected in LWA regions with the 12S primer, originating from 15 families and including 54 unique OTUs. In 2023, 12S eDNA detected 21 species across 15 families containing 84 unique OTUs. There were 5 species detected in 2022 that were not detected in 2023, and 6 species detected in 2023 that had not been detected in 2022. ASVswere primarily categorized as Leuciscidae (true minnows), Salmonidae (salmonids), or Cottidae (sculpins) (Figure 3.9).

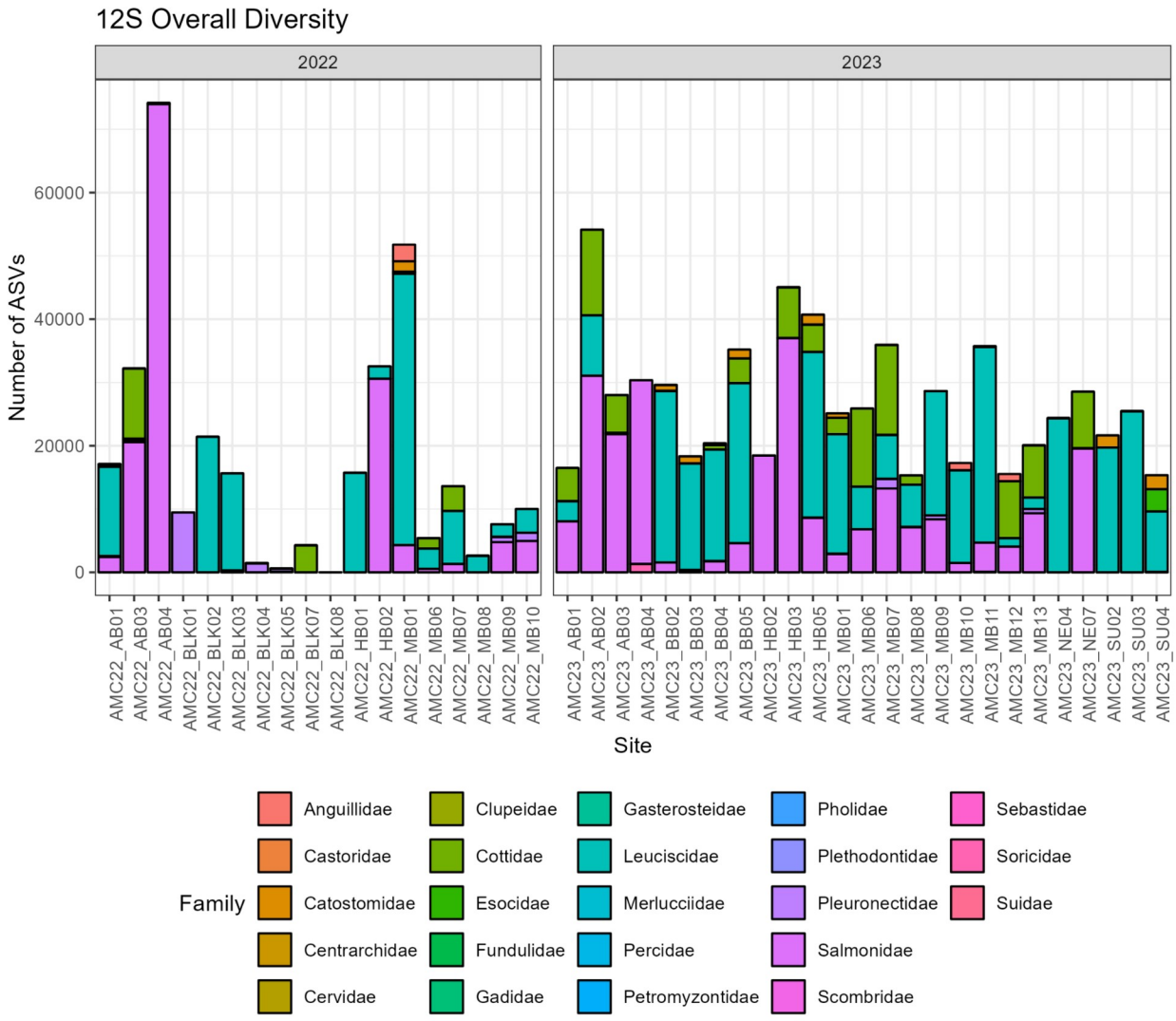

Figure 3.9 Diversity and ASV abundance of families detected in AMC LWA samples with the 12S primer. Stacked bar plots of the number of ASVs found in each cleaned 12S sample as assigned to family rank, divided by sampling season. Limited to samples collected in the LWA study region. Prevalent taxonomic families are Leuciscidae (teal), Salmonidae (purple), Cottidae (green), and small proportions of Castoridae (orange).

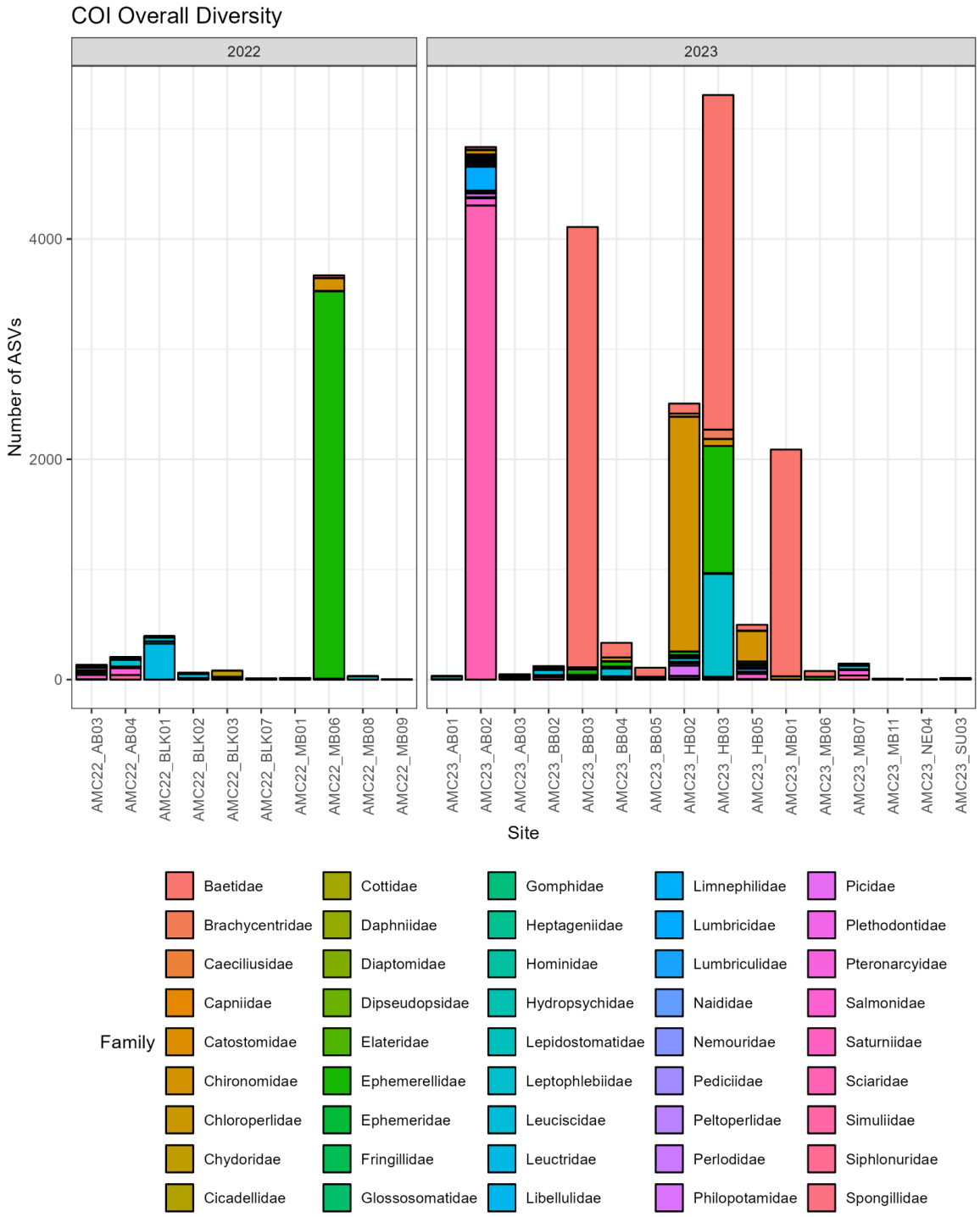

Figure 3.10 Diversity and ASV abundance of families detected in AMC LWA samples with the COI primer. Stacked bar plots of the number of ASVs found in each cleaned COI sample as assigned to family rank, divided by sampling season. Limited to samples collected in the LWA study region. Prevalent taxa belong to the Baetidae (mayflies), Chironomidae (non-biting midges), Pteronarcyidae (stoneflies), and Ephemerellidae families (spiny crawler mayflies).

COI detections reported a larger difference with 33 species across 22 families identified in 2022 containing 44 unique OTUs, compared to 63 species across 35 families with 103 unique OTUs in 2023 (Figure 3.10). Of these COI identifications, 9 species detected in 2022 were not detected in 2023, and 39 species detected in 2023 were not found in 2022.

Alpha diversity indices were calculated and plotted using the phyloseq estimate_richness and plot_richness functions, with 12S alpha diversity metrics shown in Figure 3.11 and COI metrics in Figure 3.12. Precise values are available in Table A.11 for 12S alpha diversity metrics and Table A.12 for COI alpha diversity metrics. Non-parametric ANOVAs found significant ($p < 0.05$) differences in site alpha diversity detected with the 12S primer based on treatment type for all metrics, and significant differences in alpha diversity detected with the COI primer based on treatment type for the Chao1 and Shannon metrics (Table 3.2).

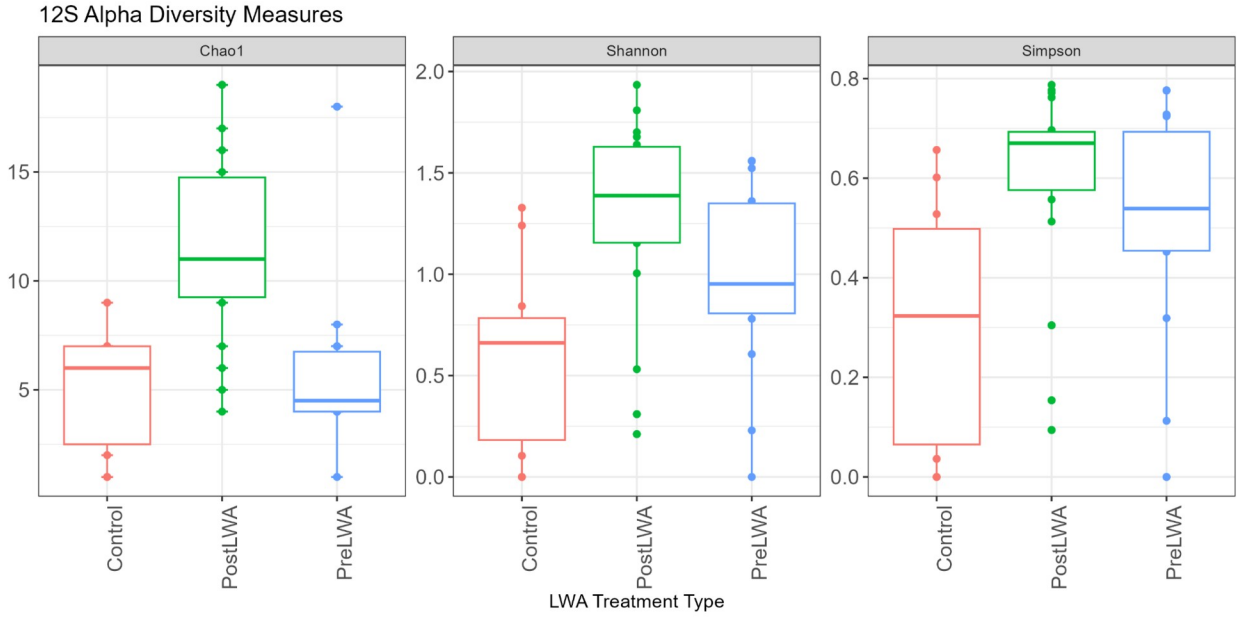

Figure 3.11 Plot of alpha diversity metrics conducted on AMC 12S samples. Depiction of Chao1, Shannon, and Simpson diversity values for cleaned AMC 12S samples in the LWA survey region. Colored by LWA treatment type.

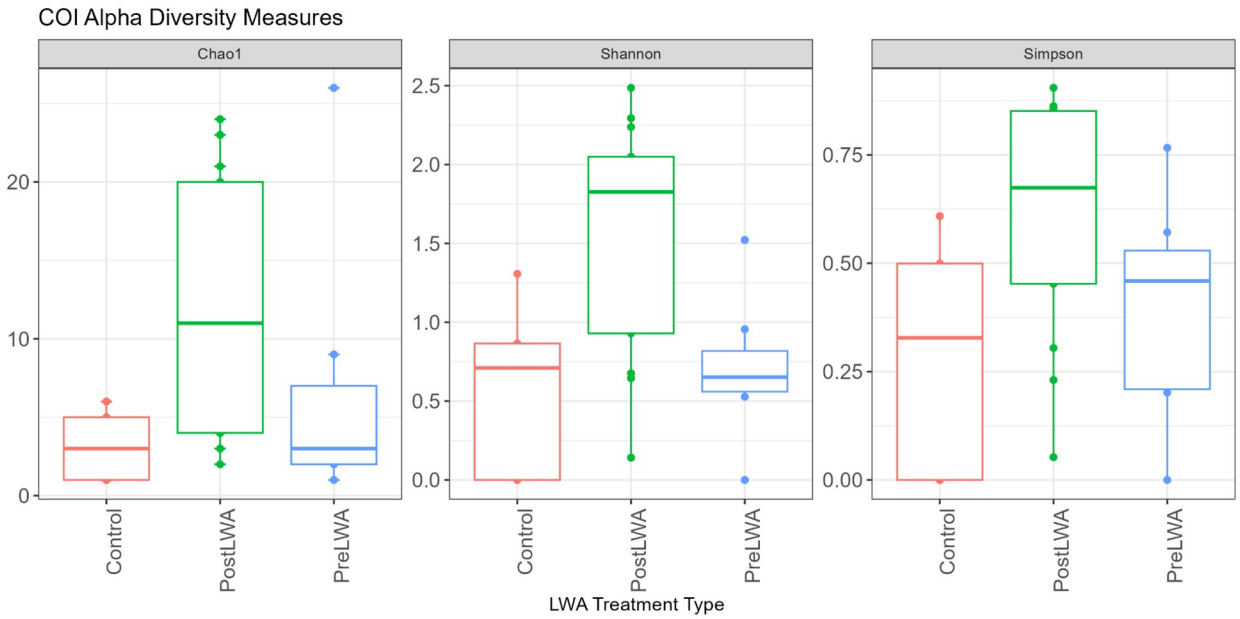

Figure 3.12 Plot of alpha diversity metrics conducted on AMC COI samples. Depiction of Chao1, Shannon, and Simpson diversity values for cleaned AMC COI samples in the LWA survey region. Colored by LWA treatment type.

Table 3.2 Non-parametric ANOVA results for alpha diversity metric comparisons. Table of p-values for alpha diversity metric ~ TreatType for both primer datasets of LWA eDNA samples.

	12S	COI
Chao1	0.00014	0.045
Shannon	0.0024	0.028
Simpson	0.0056	0.053

3.3.3 Beta Diversity

PCoA ordination using Jaccard's similarity index for presence-absence utilized ellipses drawn around the closest 50% of datapoints to reduce visual noise and highlight relationships. Ordination of the 12S data based on site LWA treatment type shows a general clustering of the Control, Pre, and Post treatment sites, though a considerable amount of overlap is present (Figure 3.13). When taking the site position in relation to LWA installation into consideration, sites cluster with Parallel and Upstream sites showing the most overlap with each other, followed by Internal and Downstream sites. LWA sites sequenced with the 12S primer showed the most dissimilarity between Upstream and Downstream sites (Figure 3.14).

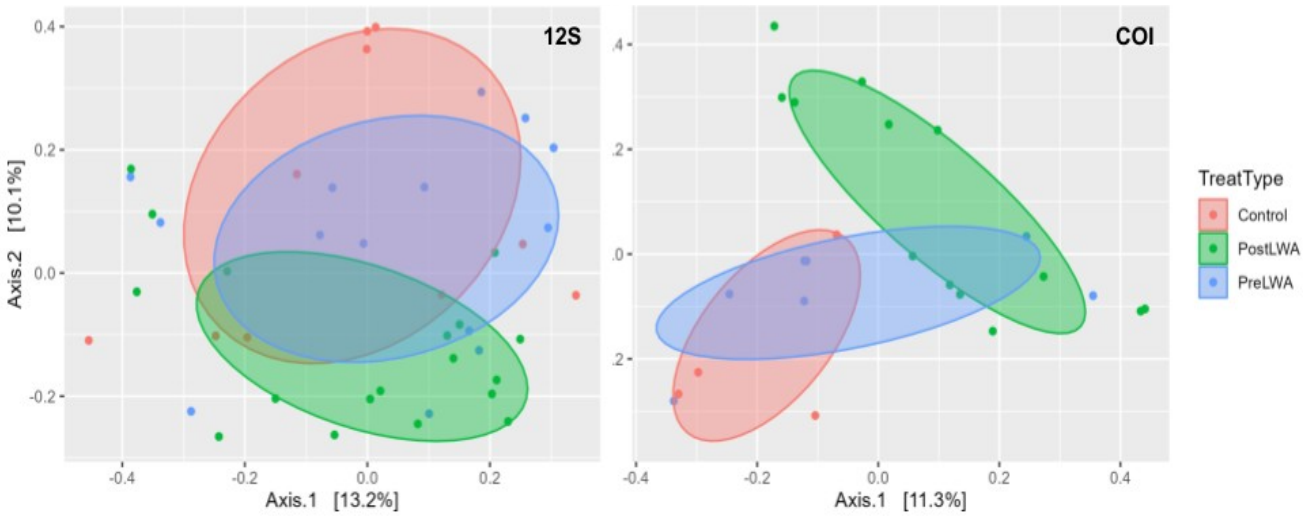

Figure 3.13 Jaccard ordination for AMC sites based on treatment type. PCoA Jaccard ordination, using a 50% best match criteria for the ellipse. 12S (left) and COI (right) samples are colored by LWA treatment type.

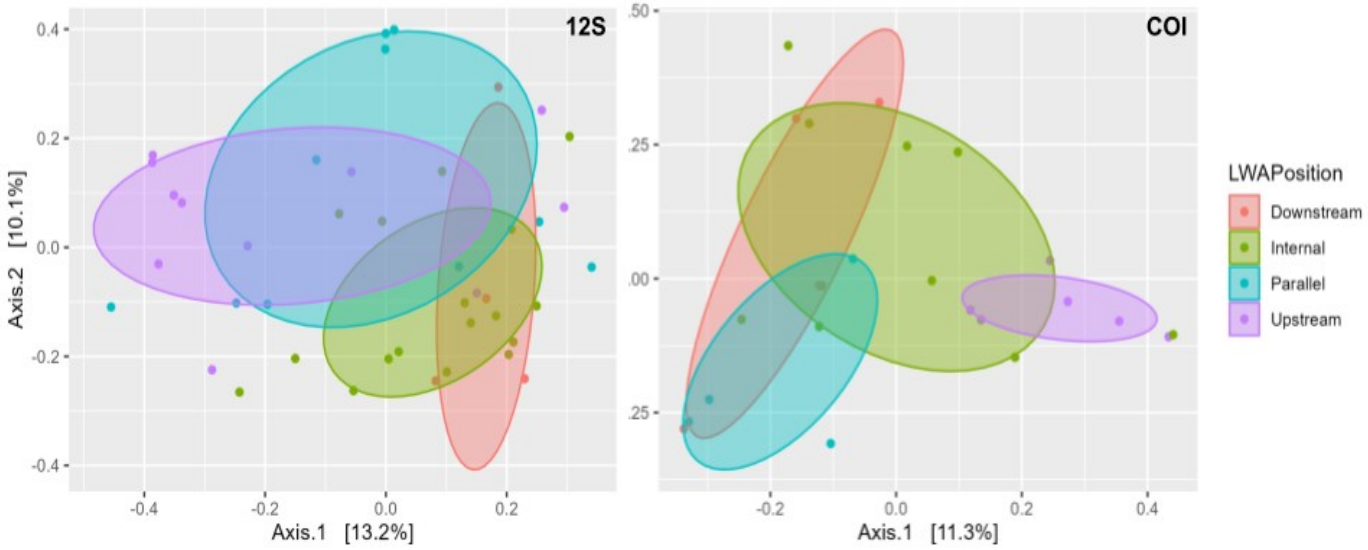

Figure 3.14 Jaccard ordination for AMC sites based on LWA position. PCoA Jaccard ordination, using a 50% best match criteria for the ellipse. 12S (left) and COI (right) samples are colored by LWA position.

Ordination results on the COI dataset also show clustering, but to a greater visual extent than the 12S dataset. Post-LWA and Control sites show little overlap with each other, though overlap for both categories with the Pre sites can be seen - Control sites having the most overlap with Pre sites (Figure 3.13). With site position as the focus, Internal sites have some level of overlap with all other categories, though least with Parallel sites. Parallel and Downstream sites

show overlap with each other but no overlap using the 50% match criteria with Upstream sites (Figure 3.14).

PERMANOVAs for the 12S data found that treatment type, year of collection, and site position all had significant impacts on community composition ($P < 0.05$), as did the interaction between treatment type and year (TreatType*Year). Interaction between treatment type and site position, however, did not have a significant influence on on community composition ($F > 0.05$). For the COI data, the same relationships were found, with treatment type, year, site position, and the interaction between treatment type and year all reporting F-values below 0.05 (Table 3.3).

Table 3.3 PERMANOVA results for AMC samples based on presence-absence. F-value results of PERMANOVA tests using calculated Jaccard distances.

Test	12S	COI
Distance ~ TreatType	0.01	0.02
Distance ~ Year	0.01	0.01
Distance ~ LWAPosition	0	0.01
Distance ~ TreatType*LWAPosition	0.74	0.62
Distance ~ TreatType*Year	0	0.01

3.4 Discussion and Conclusions

Post-LWA sites for both 12S and COI datasets typically showed higher alpha diversity and richness than Control sites, with Pre-LWA sites reporting similarly high or intermediate richness and evenness. Beta diversity ordination is more informative - 12S detections are most similar between Control and Pre-LWA sites, as well as between Parallel and Upstream sites, indicating that community composition shifts after LWA installation. Since Parallel sites are part of the Control treatment type, this could indicate that community shifts due to LWA installation are not reflected upstream of the installation area. The overlap between Internal and Downstream sites may indicate either the transport of eDNA from the LWA-impacted area or that LWA impacts on communities persist downstream of the installation. The significance of

LWA installation is reflected in PERMANOVA results, as both treatment type and site position were found to be significant to 12S detected community composition.

The COI data also shows strong trends for both treatment type and site position. The greater visual separation between Control and Post-LWA sites indicates a significant correlation between LWA installation and community composition, supported by a higher rate of similarity between Pre-LWA and Control sites (Figure 3.13). Meanwhile, the overlap in communities between Parallel and Downstream, and Downstream and Internal sites initially appear at odds with the dissimilarity between Control sites (which include Parallel sites) and Post-LWA sites. Given known issues[5,11,13] in collecting water eDNA samples of benthic macroinvertebrates, this pattern of clustering could indicate a shorter transport distance of COI-matched ASVs compared to the 12S matches. Additionally, Upstream sites showed no visual overlap with either Downstream or Parallel sites, which could also support a shorter transport distance or a lower movement ability between COI-detected benthic macroinvertebrates and 12S-detected bony fish. As PERMANOVA confirmed the significance of treatment, position, collection year, and the interaction between treatment type and year, it is strongly supported that LWA installation impacts community dynamics of macroinvertebrate communities.

Despite an initial larger number of reads in the raw COI FASTA files, reads had significant quality issues, particularly for reverse reads, and samples experienced large drops in read, OTU, and ASV numbers as the workflow progressed (Figures 3.2-3.4). Curiously, despite a large difference in initial OTU identification, the 12S and COI datasets resulted in similar numbers of target taxa found. Non-target 12S identifications were dominated by plants with smaller proportions of bacteria and foraminifera, but 12S samples overall still had significant proportions of vertebrate identification even before decontam and non-target removal (Figure A.16). COI non-target taxa were overwhelmingly plants, fungi, and algae, and made up 99.2% of initially identified OTUs for the dataset (Figure A.17).

Both 12S and COI datasets detected more species, SVs, and OTUs at sites from 2023 compared to 2022, though it is important to remember that not all 2023 sites were repeats of 2022 sites. Unfortunately, due to all 2022 samples undergoing refiltration to minimize ethanol interfering with extraction, it is unclear whether or not the refiltration impacted detection results through loss of eDNA. Since the refiltration factor matches with collection year, PERMANOVA of community composition tested against refiltration yields the same significant response as collection year and is uninformative as to the importance of refiltration specifically as a confounding factor.

The ability to quickly train and collect eDNA samples enabled field teams to survey a wide range of the MWI, including reaches with no prior survey data due to the impracticality of SBSM application. Though no comparison data is available to test for undetected species, the collection of presence-absence data allows for future studies to prioritize sites for further monitoring or validation based on detected taxa. This study is limited by being unable to verify if the lower diversity and richness measured in 2022 samples is valid or a result of processing reducing extractable DNA, a lack of species behavioral data to supplement analysis of beta diversity patterns, and equipment issues between the 2022 and 2023 sampling seasons that reduce the strength of direct comparisons. Sampling was also complicated by weather conditions that made certain sites too dangerous to access, reducing the pool of repeat sites, and a mapping error that led to sampling the wrong stream in 2022 and missing the pre-LWA sampling period for the reach.

Though biodiversity conclusions are limited to presence-absence, eDNA surveys show promising results for the measurable impact of large wood addition restoration in Maine streams. Both the 12S and COI primers, despite varying difficulties in initial species library size, read quality, and detection of non-target taxa, detected their target taxa of fish and benthic macroinvertebrates and recorded significant community composition changes in response to LWA installation. Responses included not only the status of LWA (Control or Pre vs. Post) but

also the spatial relationship between the sampled site and LWA installation. The spatial relationships between detected communities, sites, and LWA installation in particular point to species characteristics impacting eDNA detection. Future monitoring and species surveys are recommended to monitor the continuing impacts of LWA on stream reaches and to potentially link species behaviors or characteristics to community composition changes.

REFERENCES

1. Shogren AJ, Tank JL, Egan SP, August O, Rosi EJ, Hanrahan BR, et al. Water Flow and Biofilm Cover Influence Environmental DNA Detection in Recirculating Streams. Environ Sci Technol [Internet]. 2018 Aug 7;52(15):8530–7. Available from: http://dx.doi.org/10.1021/acs.est.8b01822

2. de Groot NF. Capturing human environmental DNA: Ethical challenges. Digit Soc [Internet]. 2023 Dec 16 [cited 2024 Jul 5];2(3):1–9. Available from: https://link.springer.com/article/10.1007/s44206-023-00077-9

3. Bush A, Compson ZG, Monk WA, Porter TM, Steeves R, Emilson E, et al. Studying ecosystems with DNA metabarcoding: Lessons from biomonitoring of aquatic macroinvertebrates. Front Ecol Evol [Internet]. 2019 Nov 8 [cited 2024 Jul 5];7:484625. Available from: https://www.frontiersin.org/journals/ecology-and-evolution/articles/10.3389/fevo.2019.00434/pdf

4. Wood ZT, Lacoursière-Roussel A, LeBlanc F, Trudel M, Kinnison MT, Garry McBrine C, et al. Spatial Heterogeneity of eDNA Transport Improves Stream Assessment of Threatened Salmon Presence, Abundance, and Location. Frontiers in Ecology and Evolution [Internet]. 2021;9:168. Available from: https://www.frontiersin.org/article/10.3389/fevo.2021.650717

5. Silverbrand S. Use of the New England Aquarium to Evaluate Environmental DNA Metabarcoding of Gulf of Maine Vertebrates and Invertebrates [Internet]. The University of Maine; 2021 [cited 2024 Jul 5]. Available from: https://digitalcommons.library.umaine.edu/etd/3538

6. Harrison JB, Sunday JM, Rogers SM. Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proc Biol Sci [Internet]. 2019 Nov 20;286(1915):20191409. Available from: http://dx.doi.org/10.1098/rspb.2019.1409

7. Bohmann K, Lynggaard C. Transforming terrestrial biodiversity surveys using airborne eDNA. Trends Ecol Evol [Internet]. 2023 Feb 1 [cited 2024 Jul 5];38(2):119–21. Available from: http://dx.doi.org/10.1016/j.tree.2022.11.006

8. Skelton J, Cauvin A, Hunter ME. Environmental DNA metabarcoding read numbers and their variability predict species abundance, but weakly in non-dominant species. Environmental DNA [Internet]. 2022 Sep 30; Available from: http://dx.doi.org/10.1002/edn3.355

9. Carraro L, Stauffer JB, Altermatt F. How to design optimal eDNA sampling strategies for biomonitoring in river networks. Environ DNA [Internet]. 2021 Jan 1 [cited 2024 Jul 5];3(1):157–72. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/edn3.137

10. Rivera SF, Rimet F, Vasselon V, Vautier M, Domaizon I, Bouchez A. Fish eDNA metabarcoding from aquatic biofilm samples: Methodological aspects. Mol Ecol Resour [Internet]. 2022 May 1 [cited 2024 Jul 5];22(4):1440–53. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.13568

11. Crane LC, Goldstein JS, Thomas DW, Rexroth KS, Watts AW. Effects of life stage on eDNA detection of the invasive European green crab (Carcinus maenas) in estuarine systems. Ecol Indic [Internet]. 2021 May 1;124:107412. Available from: https://www.sciencedirect.com/science/article/pii/S1470160X21000777

12. Elbrecht V, Braukmann TWA, Ivanova NV, Prosser SWJ, Hajibabaei M, Wright M, et al. Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ [Internet]. 2019 Oct 7 [cited 2024 Jul 5];7(e7745):e7745. Available from: https://peerj.com/articles/7745

13. Derycke S, Maes S, Van den Bulcke L, Vanhollebeke J, Wittoeck J, Hillewaert H, et al. Optimisation of metabarcoding for monitoring marine macrobenthos: primer choice and morphological traits determine species detection in bulkDNA and eDNA from the ethanol preservative [Internet]. Authorea Preprints. Authorea, Inc.; 2020 [cited 2024 Jul 5]. Available from: https://www.authorea.com/doi/pdf/10.22541/au.159665093.39993653

14. Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, et al. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. R Soc Open Sci [Internet]. 2015 Jul 26 [cited 2024 Jul 5];2(7):150088. Available from: https://royalsocietypublishing.org/doi/10.1098/rsos.150088

15. Shogren AJ, Tank JL, Andruszkiewicz E, Olds B, Mahon AR, Jerde CL, et al. Controls on eDNA movement in streams: Transport, Retention, and Resuspension. Sci Rep [Internet]. 2017 Jul 11;7(1):5065. Available from: http://dx.doi.org/10.1038/s41598-017-05223-1

16. Shogren AJ, Tank JL, Egan SP, Bolster D, Riis T. Riverine distribution of mussel environmental DNA reflects a balance among density, transport, and removal processes. Freshw Biol [Internet]. 2019 Aug;64(8):1467–79. Available from: https://onlinelibrary.wiley.com/doi/10.1111/fwb.13319

17. Baudry T, Laffitte M, Noizat C, Delaunay C, Ricou G, Vasselon V, et al. Influence of distance from source population and seasonality in eDNA detection of white-clawed crayfish, through qPCR and ddPCR assays. Environ DNA [Internet]. 2023 Jul 1 [cited 2024 Jul 5];5(4):733–49. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/edn3.435

18. Leese F, Sander M, Buchner D, Elbrecht V, Haase P, Zizka VMA. Improved freshwater macroinvertebrate detection from environmental DNA through minimized nontarget amplification. Environ DNA [Internet]. 2021 Jan;3(1):261–76. Available from: https://onlinelibrary.wiley.com/doi/10.1002/edn3.177

19. Shogren AJ, Tank JL, Hanrahan BR, Bolster D. Controls on fine particle retention in experimental streams. Freshw Sci [Internet]. 2020 Mar 1;39(1):28–38. Available from: https://doi.org/10.1086/707396

20. Stoeckle MY, Ausubel JH, Coogan M. 12S gene metabarcoding with DNA standard quantifies marine bony fish environmental DNA , identifies threshold for reproducible detection, and overcomes distortion due to amplification of non-fish DNA. Environ DNA

[Internet]. 2022 Dec 8; Available from: https://onlinelibrary.wiley.com/doi/10.1002/edn3.376

21. Vanderpool DD, Wilcox TM, Young MK, Pilgrim KL, Schwartz MK. Simultaneous species detection and discovery with environmental DNA metabarcoding: A freshwater mollusk case study. Ecol Evol [Internet]. 2024 Feb 1 [cited 2024 Jul 5];14(2):e11020. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.11020

22. Wilson JJ. DNA barcodes for insects. Methods Mol Biol [Internet]. 2012 [cited 2024 Jul 5];858:17–46. Available from: https://link.springer.com/protocol/10.1007/978-1-61779-591-6_3

23. Toczydlowski RH, Liggins L, Gaither MR, Anderson TJ, Barton RL, Berg JT, et al. Poor data stewardship will hinder global genetic diversity surveillance. Proc Natl Acad Sci U S A [Internet]. 2021 Aug 24;118(34). Available from: http://dx.doi.org/10.1073/pnas.2107934118

24. Elbrecht V, Leese F. Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. Front Environ Sci Eng China [Internet]. 2017 Apr 10;5. Available from: http://journal.frontiersin.org/article/10.3389/fenvs.2017.00011/full

25. Braukmann TWA, Ivanova NV, Prosser SWJ, Elbrecht V, Steinke D, Ratnasingham S, et al. Metabarcoding a diverse arthropod mock community. Mol Ecol Resour [Internet]. 2019 May 1 [cited 2024 Jul 5];19(3):711–27. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.13008

26. Davis BY. Working species list and sources for Maine eDNA reference library [Internet]. Zenodo; 2023 [cited 2024 Jul 5]. Available from: https://zenodo.org/records/11225486

27. Download [Internet]. [cited 2024 Jul 5]. Available from: https://www.gbif.org/occurrence/download/0138130-220831081235567

28. Mächler E, Little CJ, Wüthrich R, Alther R, Fronhofer EA, Gounand I, et al. Assessing different components of diversity across a river network using eDNA. Environ DNA [Internet]. 2019 Sep 1 [cited 2024 Jul 5];1(3):290–301. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/edn3.33

29. Shields FD, Knight SS, Stofleth JM. Large wood addition for aquatic habitat rehabilitation in an incised, sand-bed stream, Little Topashaw Creek, Mississippi. River Res Appl [Internet]. 2006 Sep;22(7):803–17. Available from: https://onlinelibrary.wiley.com/doi/10.1002/rra.937

30. Draft 2024 Integrated Water Quality Monitoring and Assessment Report [Internet]. Maine Department of Environmental Protection. [cited 2024 Jul 5]. Available from: https://content.govdelivery.com/accounts/MEDEP/bulletins/3a37c14

31. Parr TB, Cronan CS, Danielson TJ, Tsomides L, Simon KS. Aligning indicators of community composition and biogeochemical function in stream monitoring and ecological assessments. Ecol Indic [Internet]. 2016 Jan 1;60:970–9. Available from: https://www.sciencedirect.com/science/article/pii/S1470160X15004811

32. Weil KK, Cronan CS, Lilieholm RJ, Danielson TJ, Tsomides L. A statistical analysis of watershed spatial characteristics that affect stream responses to urbanization in Maine, USA. Appl Geogr [Internet]. 2019 Apr 1 [cited 2024 Jul 5];105:37–46. Available from: http://dx.doi.org/10.1016/j.apgeog.2019.02.004

33. Stream Explorers [Internet]. Maine Audubon. 2020 [cited 2024 Jul 5]. Available from: https://maineaudubon.org/projects/stream-explorers/

34. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods [Internet]. 2016 Jul 23 [cited 2024 Jul 5];13(7):581–3. Available from: https://www.nature.com/articles/nmeth.3869

35. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One [Internet]. 2013 Apr 22 [cited 2024 Jul 5];8(4):e61217. Available from: https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0061217&type=printable

36. Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J (2024). *vegan: Community Ecology Package*. R package version 2.6-7, https://github.com/vegandevs/vegan, https://vegandevs.github.io/vegan/.

37. Genus: Corydalus [Internet]. [cited 2024 Jul 5]. Available from: https://www.insectidentification.org/taxonomy/genus.php?by-genus=Corydalus

38. Are WW. Giving Our Rivers Room to Move: A New Strategy and Contribution to Protecting Vermont's Communities and Ensuring Clean Water [Internet]. Lake Champlain Basin Program. 2016 [cited 2024 Jul 5]. Available from: https://www.lcbp.org/publications/giving-rivers-room-move-new-strategy-contribution-protecting-vermonts-communities-ensuring-clean-water/

39. Wohl E. A legacy of absence: Wood removal in US rivers. Prog Phys Geogr [Internet]. 2014 Oct 15 [cited 2024 Jul 5];38(5):637–63. Available from: https://journals.sagepub.com/doi/10.1177/0309133314548091

40. Kratzer JF. Response of Brook Trout biomass to strategic wood additions in the East Branch Nulhegan River watershed, Vermont. N Am J Fish Manag [Internet]. 2018 Dec 1 [cited 2024 Jul 5];38(6):1415–22. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/nafm.10241

41. Cordova JM, Rosi-Marshall EJ, Yamamuro AM, Lamberti GA. Quantity, controls and functions of large woody debris in Midwestern USA streams. River Res Appl [Internet]. 2007 Jan 1 [cited 2024 Jul 5];23(1):21–33. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/rra.963

42. Strategic Wood Addition to Improve Stream Habitat [Internet]. [cited 2024 Jul 5]. Available from: https://vtfishandwildlife.com/conserve/aquatic-habitat-conservation/strategic-wood-addition-to-improve-stream-habitat

43. About – Appalachian Mountain Club [Internet]. Appalachian Mountain Club (AMC). Appalachian Mountain Club; 2022 [cited 2024 Jul 5]. Available from: https://www.outdoors.org/about/

44. Benhadi-Marín J. A conceptual framework to deal with outliers in ecology. Biodivers Conserv [Internet]. 2018 Oct 1 [cited 2024 Jul 5];27(12):3295–300. Available from: https://link.springer.com/article/10.1007/s10531-018-1602-2

45. McGill BJ, Etienne RS, Gray JS, Alonso D, Anderson MJ, Benecha HK, et al. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecol Lett [Internet]. 2007 Oct;10(10):995–1015. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1461-0248.2007.01094.x

46. Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL, et al. Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist: Roadmap for beta diversity. Ecol Lett [Internet]. 2011 Jan 1 [cited 2024 Jul 5];14(1):19–28. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2010.01552.x

47. Davis BY. mednaTaxaRef: Development of an R toolset to generate reference databases for use in Maine-eDNA sequence analyses based on merging existing functionalities [Internet]. Github; [cited 2024 Jul 5]. Available from: https://github.com/bethydavis/mednaTaxaRef

48. Borstein SR, O'Meara BC. AnnotationBustR: An R package to extract subsequences from GenBank annotations [Internet]. PeerJ. 2017 [cited 2024 Jul 11]. Available from: https://www.researchgate.net/publication/344989768_AnnotationBustR_An_R_package_to_extract_subsequences_from_GenBank_annotations

49. Chamberlain SA, Szöcs E. taxize: taxonomic search and retrieval in R. F1000Res [Internet]. 2013 Sep 18 [cited 2024 Jul 11];2(191):191. Available from: https://f1000research.com/articles/2-191/pdf

50. Macher J-N, Vivancos A, Piggott JJ, Centeno FC, Matthaei CD, Leese F. Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate cytochrome c oxidase I primers. Mol Ecol Resour [Internet]. 2018 Nov 1 [cited 2024 Jul 29];18(6):1456–68. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.12940

51. Hajibabaei M, Porter TM, Robinson CV, Baird DJ, Shokralla S, Wright MTG. Watered-down biodiversity? A comparison of metabarcoding results from DNA extracted from matched water and bulk tissue biomonitoring samples. PLoS One [Internet]. 2019 Dec 12;14(12):e0225409. Available from: http://dx.doi.org/10.1371/journal.pone.0225409

52. Gleason JE, Elbrecht V, Braukmann TWA, Hanner RH, Cottenie K. Assessment of stream macroinvertebrate communities with eDNA is not congruent with tissue-based metabarcoding. Mol Ecol [Internet]. 2021 Jul 1 [cited 2024 Jul 29];30(13):3239–51. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15597

53. Watts C, Dopheide A, Holdaway R, Davis C, Wood J, Thornburrow D, et al. DNA metabarcoding as a tool for invertebrate community monitoring: a case study comparison with conventional techniques: Monitoring invertebrates using DNA metabarcoding. Aust Entomol [Internet]. 2019 Aug 1 [cited 2024 Jul 29];58(3):675–86. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/aen.12384

APPENDIX

Additional Files:

The following files are available as additional uploads through DigitalCommons:

- ExtractionProtocol.odt - The extraction protocol used for all eDNA samples
- RefLibSummary_12S.csv - A summary of species names and the number of associated target sequences found for the 12S reference library
- RefLibSummary_COI.csv - A summary of species names and the number of associated target sequences found for the COI reference library
- RefLib_SharedSp.csv - A list of species with sequences found in both the 12S and COI reference libraries
- AMC12S_WorkflowVerification.csv - A spreadsheet of the read numbers after each dada2 processing step for AMC 12S samples
- AMCCOI_WorkflowVerification.csv - A spreadsheet of the read numbers after each dada2 processing step for AMC COI samples

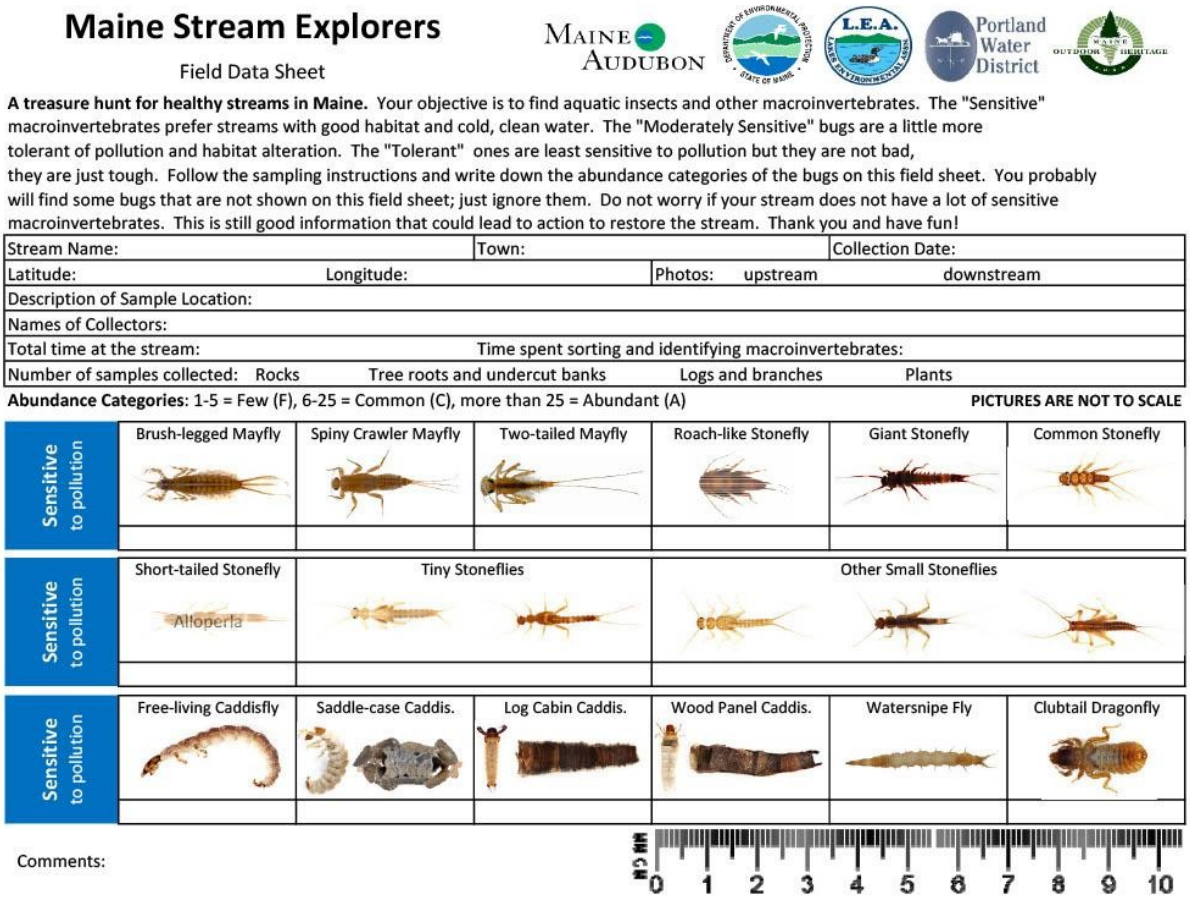

Figure A.1 Stream Explorers field data sheet - top. Image of the top side of the Maine Stream Explorers survey field data sheet, depicting both site metadata to collect and examples of target organisms for comparison to collected specimens.

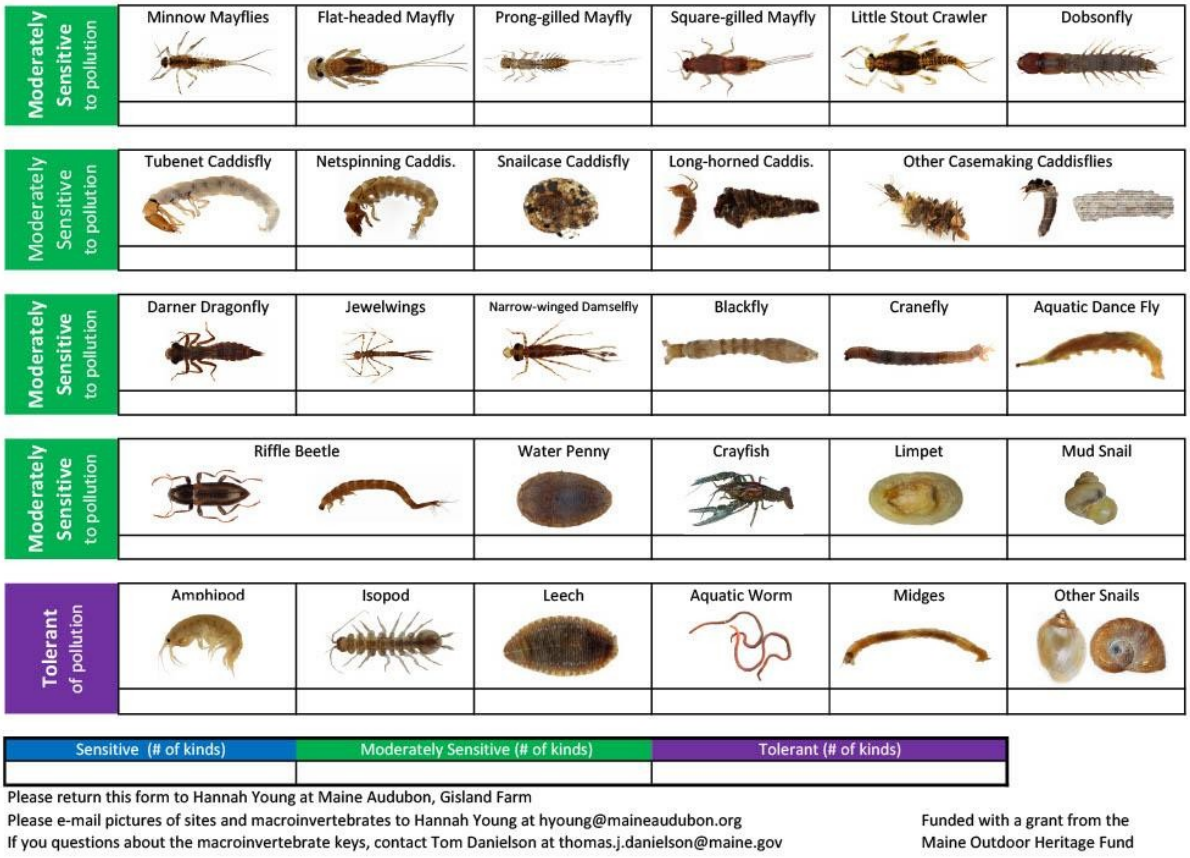

Figure A.2 Stream Explorers field data sheet - reverse. Image of the reverse side of the Maine Stream Explorers survey field data sheet, depicting both site metadata to collect and examples of target organisms for comparison to collected specimens.

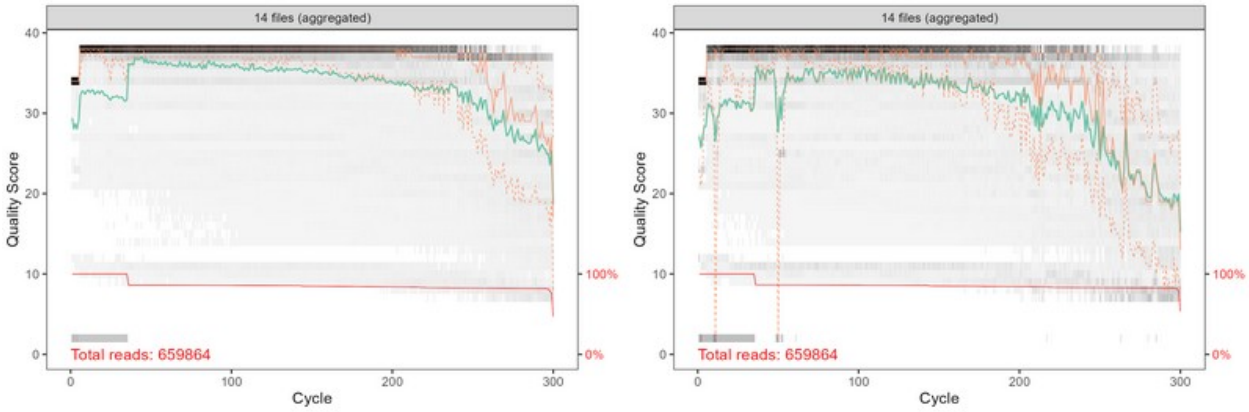

Figure A.3 Aggregate quality profile plots for PNW raw reads. Visualization of the aggregate quality scores of the raw forward reads (left) and raw reverse reads (right) from all PNW samples.

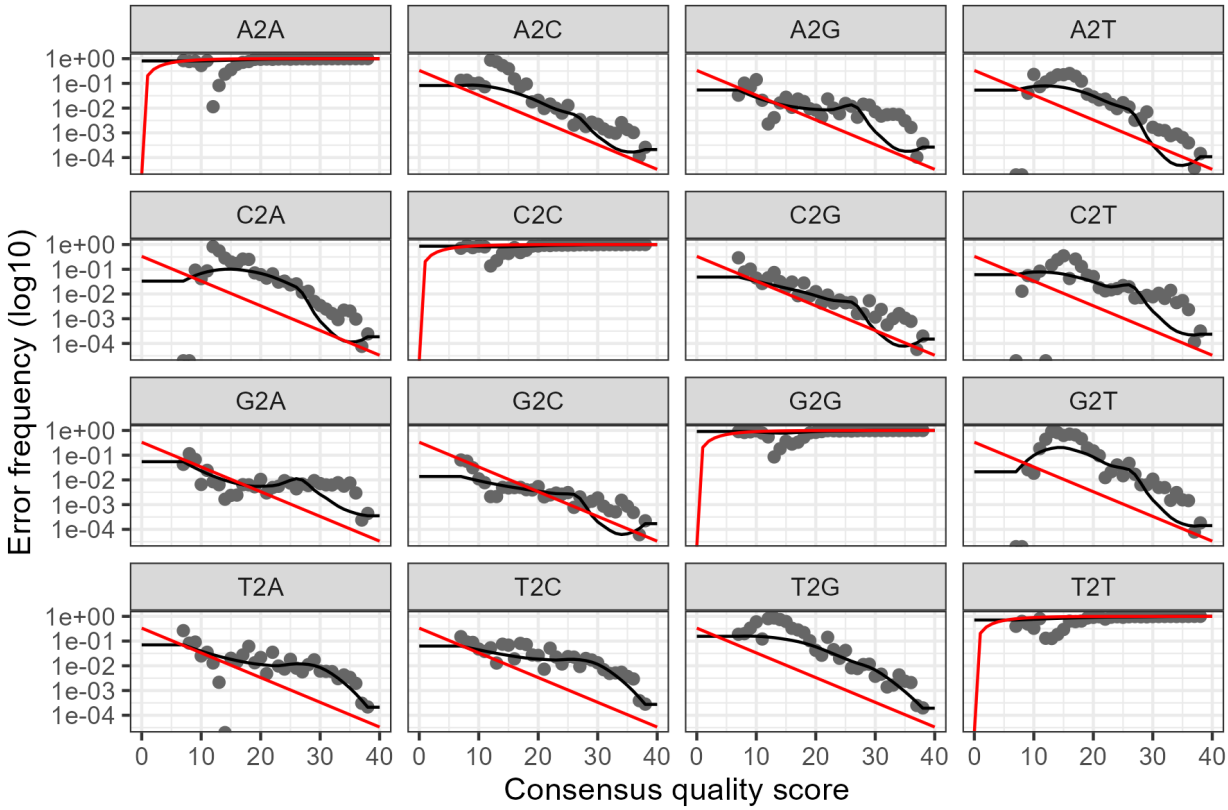

Figure A.4 Error rates for PNW 12S forward reads. Generated error rate visualization for PNW 12S forward reads after filtering, trimming, and dereplicating reads.

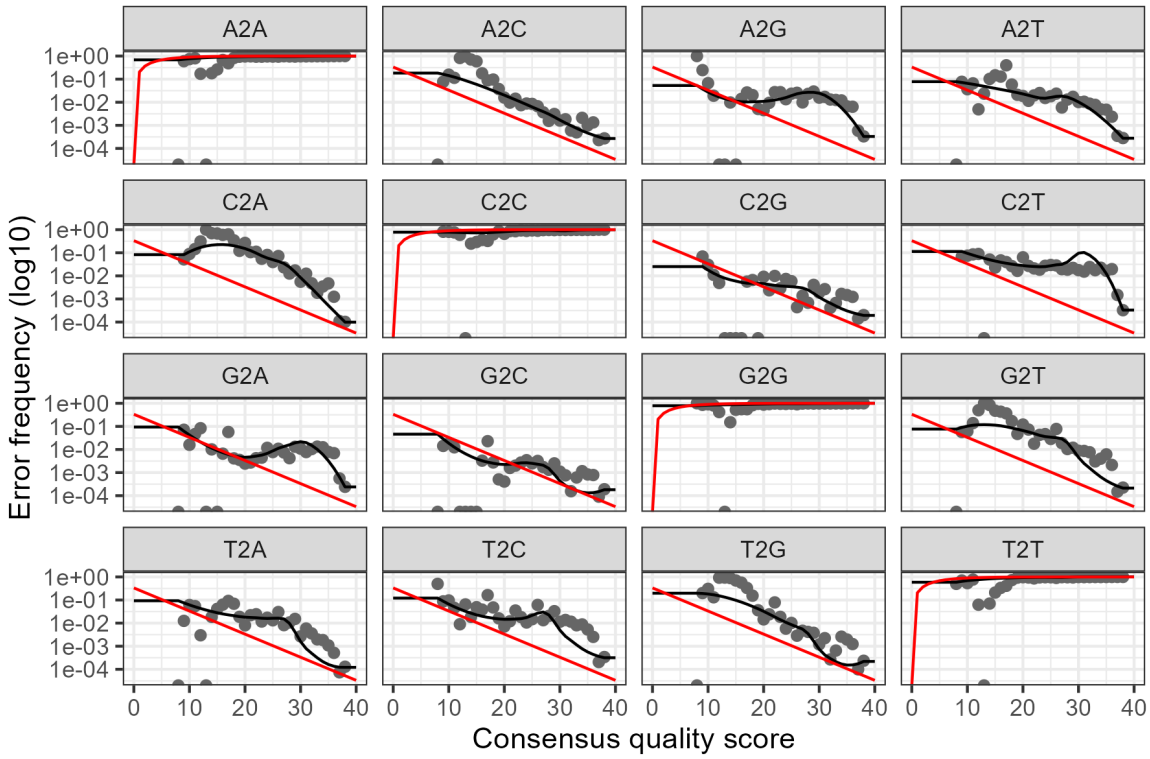

Figure A.5 Error rates for PNW 12S reverse reads. Generated error rate visualization for PNW 12S reverse reads after filtering, trimming, and dereplicating reads.

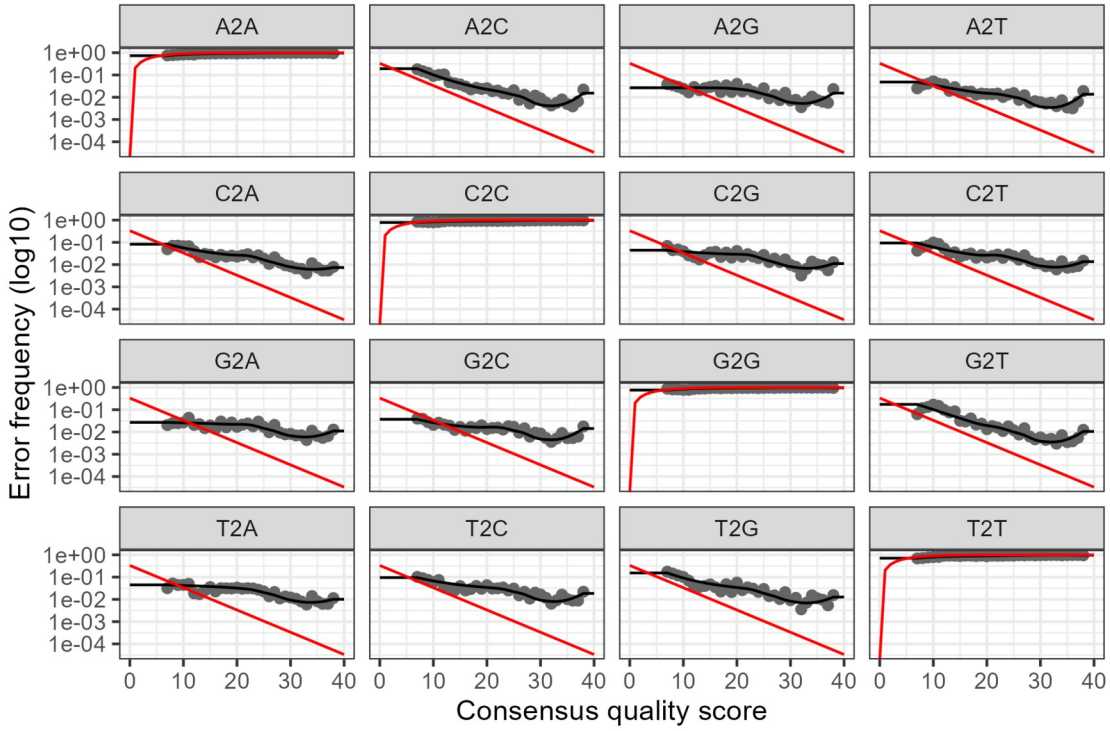

Figure A.6 Error rates for PNW COI forward reads. Generated error rate visualization for PNW COI forward reads after filtering, trimming, and dereplicating reads.

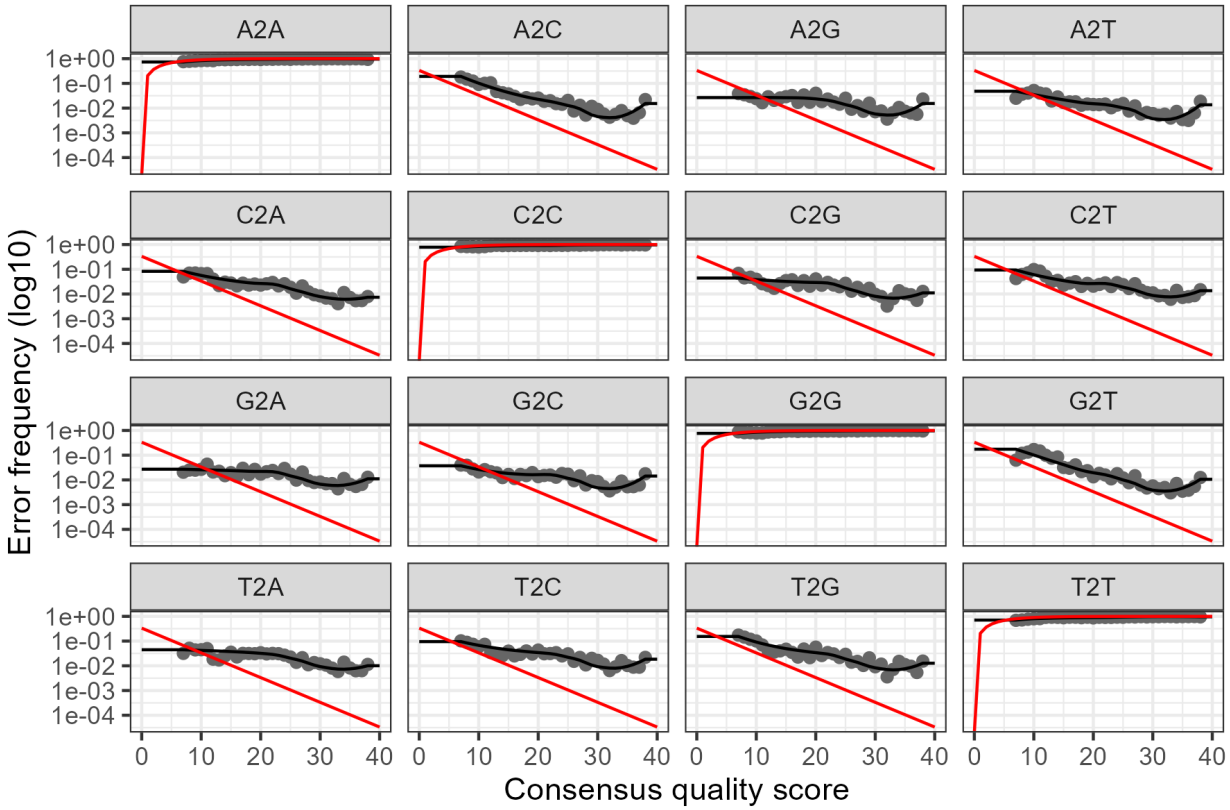

Figure A.7 Error rates for PNW COI reverse reads. Generated error rate visualization for PNW COI reverse reads after filtering, trimming, and dereplicating reads.

Table A.1 Workflow verification for cleaning PNW 12S reads. A summary of the number of reads dropped at each stage of dada2 cleaning for the PNW samples amplified with the 12S primer.

sample	input	filtered	denoisedF	denoisedR	merged	nonchim
PNW22_BOWD	40316	37227	37112	37122	36667	35862
PNW22_BOWU	57107	52234	52121	52132	51624	50069
PNW22_C914	78352	72181	71935	71901	71291	70313
PNW22_LUND	56718	52213	51701	51691	51192	50651
PNW22_LUNU	43325	39977	39716	39613	39318	38372
PNW22_MATD	86165	81272	81265	81251	81127	80338
PNW22_MATU	70041	66566	66562	66486	66268	65437
PNW23_BOWD	96981	88139	88007	87970	87559	84986
PNW23_BOWU	6	6	2	6	0	0
PNW23_C523	87249	81249	81244	81224	80742	80474
PNW23_LUND	76915	72535	72480	72459	72342	72004
PNW23_LUNU	61464	56494	56186	56184	55781	55092
PNW23_MATD	248254	229644	229302	229376	227959	223596
PNW23_MATU	53418	50700	50423	50547	49997	49793

Table A.2 Workflow verification for cleaning PNW COI reads. A summary of the number of reads dropped at each stage of dada2 cleaning for the PNW samples amplified with the COI primer.

sample	input	filtered	denoisedF	denoisedR	merged	nonchim
PNW22_BOWD	37218	21742	16331	18856	2619	2584
PNW22_BOWU	18958	11659	7534	9594	3475	3463
PNW22_C914	1	1	1	1	0	0
PNW22_LUND	47449	29611	25415	26673	10808	10097
PNW22_LUNU	38438	22087	20560	21309	16771	16666
PNW22_MATD	30188	20270	17376	18546	10149	10088
PNW22_MATU	44211	28675	25278	27480	19039	18974
PNW23_BOWD	42119	22781	11945	19556	3931	3913
PNW23_BOWU	102178	65371	57937	62604	20751	20571
PNW23_C523	29683	18004	15695	15169	8609	8609
PNW23_LUND	27907	18515	15798	17275	9281	9248
PNW23_LUNU	148889	89459	76674	79177	33344	30655
PNW23_MATD	35468	7467	6979	7338	5868	5811
PNW23_MATU	57157	33844	26555	30337	15032	14703

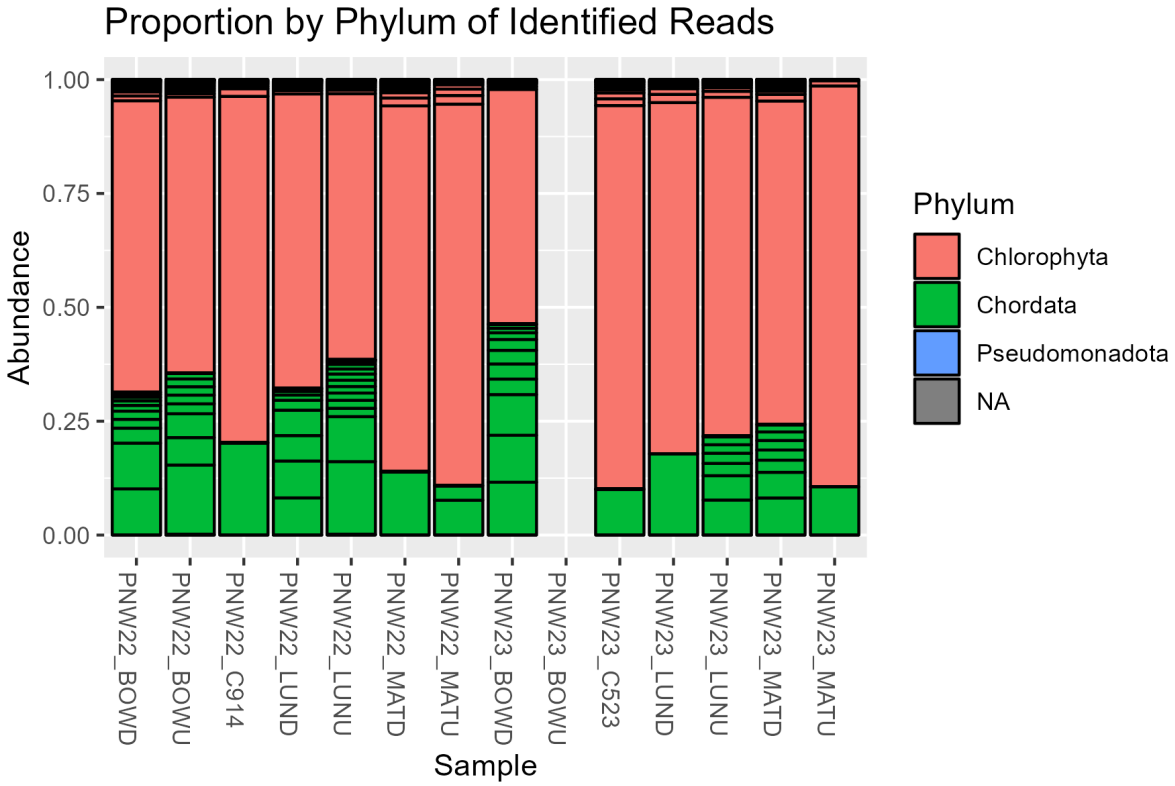

Figure A.8 Phylum identifications of taxa detected in PNW 12S samples. Proportional depiction of taxonomic classes detected in PNW 12S samples before cleaning procedures and after the removal of contaminants and non-target taxa. Abundance on the y-axis refers to the number of ASVs identified to each class.

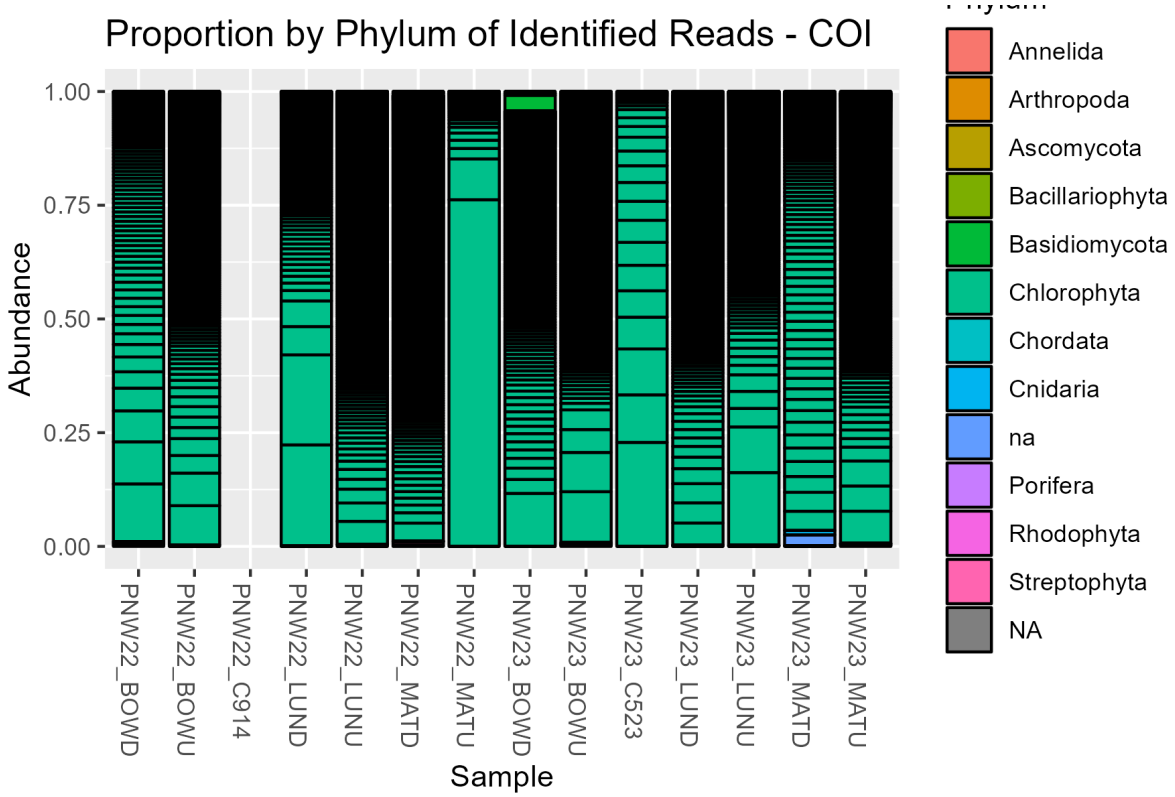

Figure A.9 Phylum identifications of taxa detected in PNW COI samples. Proportional depiction of taxonomic classes detected in PNW COI samples after cleaning procedures and before the removal of contaminants and non-target taxa. Abundance on the y-axis refers to the number of ASVs identified to each class.

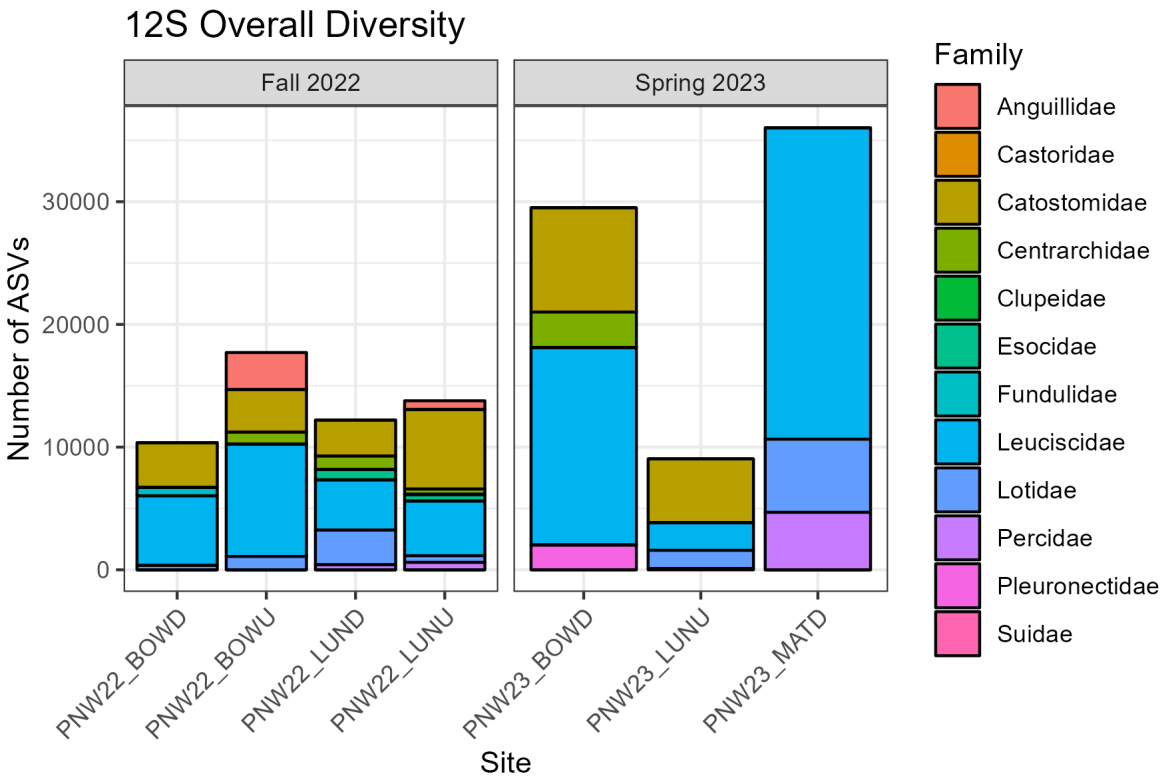

Figure A.10 Diversity and ASV abundance of families detected in PNW 12S samples. Stacked bar plots of the number of ASVs found in each cleaned 12S sample as assigned to family rank, divided by sampling season.

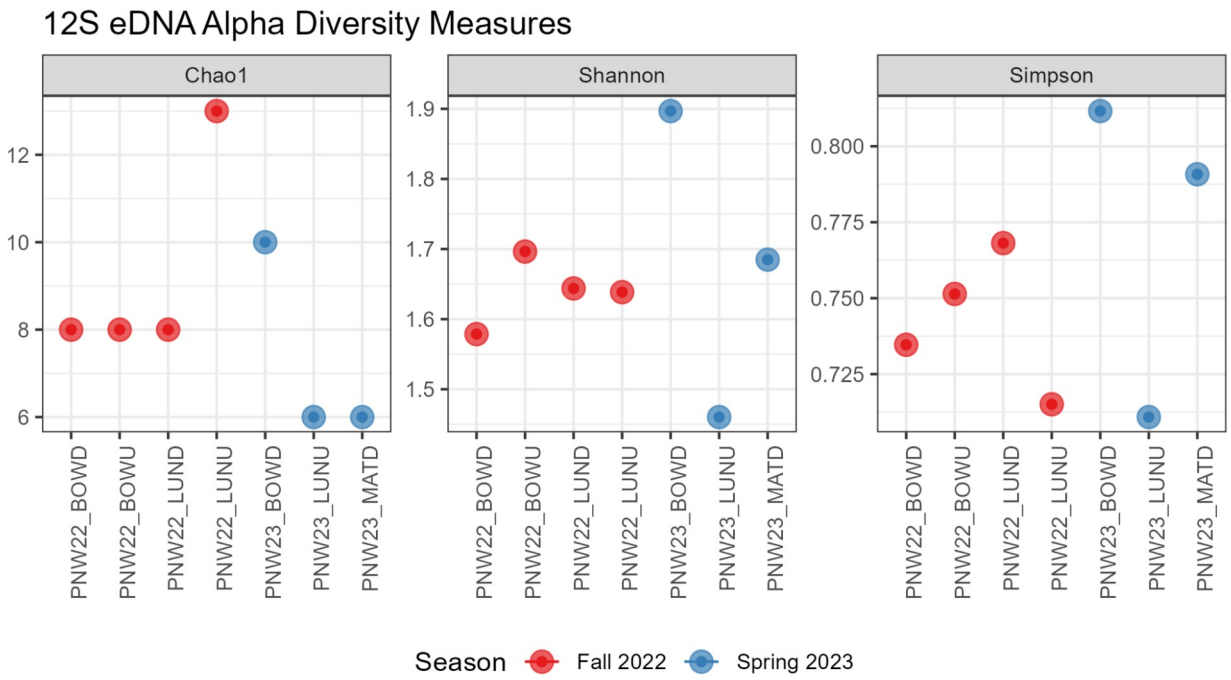

Figure A.11 Plot of alpha diversity metrics conducted on PNW 12S samples. Depiction of Chao1, Shannon, and Simpson diversity values for cleaned PNW 12S samples, colored by sampling season.

Table A.3 Total ASVs counted for each clean PNW sample. Total ASVscounted in each primer dataset for PNW samples after removal of contaminants and non-target taxa. Samples with no remaining ASVs for either primer dataset are not included.

Sample	12S	COI
PNW22_BOWD	10366	NA
PNW22_BOWU	17709	2
PNW22_MATD	NA	31
PNW22_LUND	12211	NA
PNW22_LUNU	13778	NA
PNW23_BOWD	29511	16
PNW23_BOWU	NA	79
PNW23_MATD	36027	55
PNW23_MATU	NA	37
PNW23_LUND	NA	NA
PNW23_LUNU	9040	NA

Table A.4 Number of organismal categories found at each Stream Explorers site. A sum of the number of organismal categories per site where at least one individual was found using the Stream Explorers survey.

Site	Year	Number of Counted Categories
Lunksoos	2022	16
Lunksoos	2023	22
Bowlin	2022	21
Bowlin	2023	24
Matagamon	2022	19
Matagamon	2023	24

Table A.5 Calculated alpha diversity metrics for eDNA samples. Values found for each alpha diversity test for each eDNA sample, calculated with the phyloseq package.

Sample	Chao1	Shannon	Simpson
PNW22_BOWD	8	1.58	0.73
PNW22_BOWU	8	1.7	0.75
PNW22_LUND	8	1.64	0.77
PNW22_LUNU	13	1.64	0.72
PNW23_BOWD	10	1.9	0.81
PNW23_LUNU	6	1.46	0.71
PNW23_MATD	6	1.68	0.79

Table 2.4 Individual and ASV counts for taxa observed through Stream Explorers and eDNA surveys. The number of counted individuals (for Stream Explorers) and ASVs (for eDNA) that could be identified to the overlapping taxonomic families. [d]ASV numbers came from 2 separate species identified to the family.

Site	Ephemerellidae	Heptageniidae	Perlidae	Siphlonuridae	Tipulidae
Lunksoos 2022	0	15	13	25	0
Bowlin 2022	0	13	13	27	0
Matagamon 2022	0	35	41	87	0
PNW22_MATD	19[d]	2	0	3	0
Lunksoos 2023	23	17	6	2	1
PNW23_LUND	15[d]	0	7	0	0
Bowlin 2023	10	33	10	2	0
PNW23_BOWD	0	0	0	0	16
PNW23_BOWU	10	0	0	0	0
Matagamon 2023	93	47	39	61	0
PNW22_MATD	19[d]	2	0	3	0
PNW23_MATU	20	10[d]	0	0	0

Table A.6 PERMANOVA results for PNW eDNA samples based on presence-absence. F-value results of PERMANOVA tests using calculated Jaccard distances.

Variable Test	12S	COI
Distance ~ Placement	0.51	1
Distance ~FilterAmount	0.25	0.74
Distance ~ Season	0.74	0.52
Distance ~ Reach	0.01	1

Table A.7 PERMANOVA results for PNW eDNA samples based on ASV count. F-value results of PERMANOVA tests using calculated Bray-Curtis distances.

Test	12S	COI
Distance ~ Placement	0.49	1
Distance ~FilterAmount	0.26	0.72
Distance ~ Season	0.79	0.52
Distance ~ Reach	0.02	1

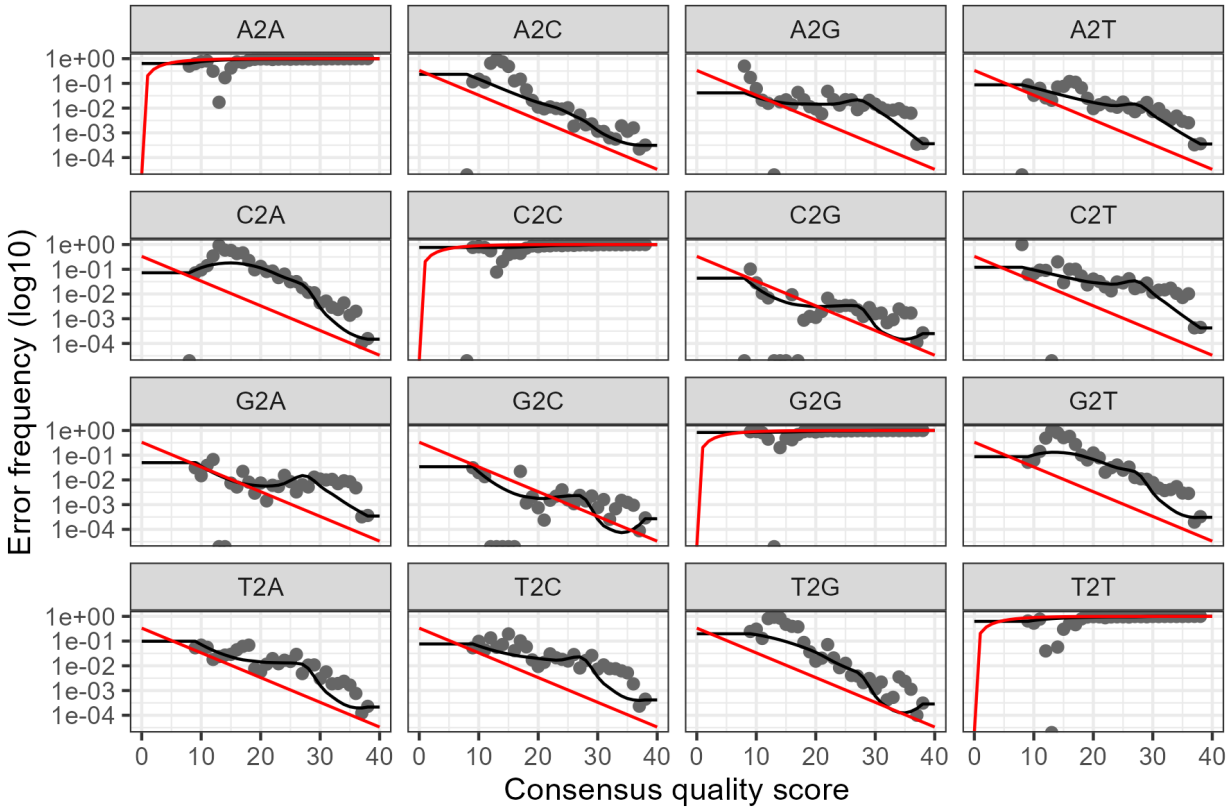

Figure A.12 Error rates for AMC 12S forward reads. Generated error rate visualization for AMC 12S forward reads after filtering, trimming, and dereplicating reads.

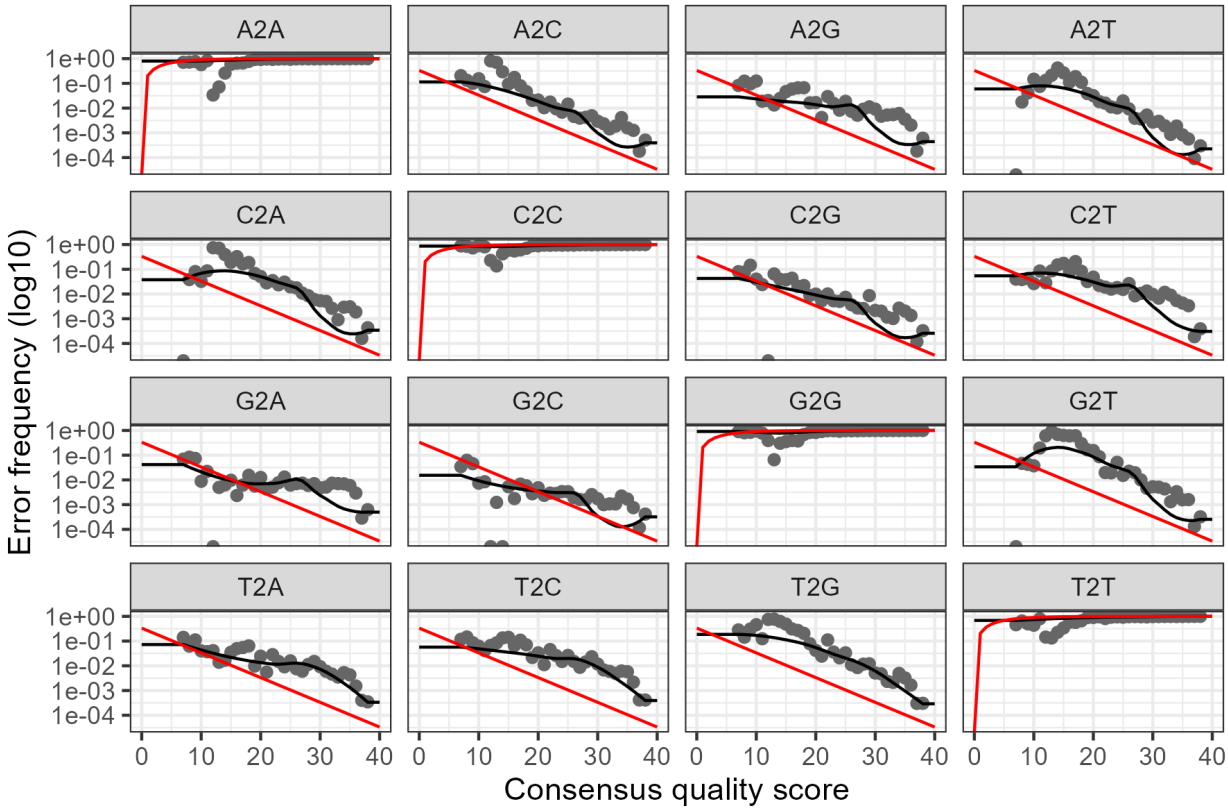

Figure A.13 Error rates for AMC12S reverse reads. Generated error rate visualization for AMC 12S reverse reads after filtering, trimming, and dereplicating reads.

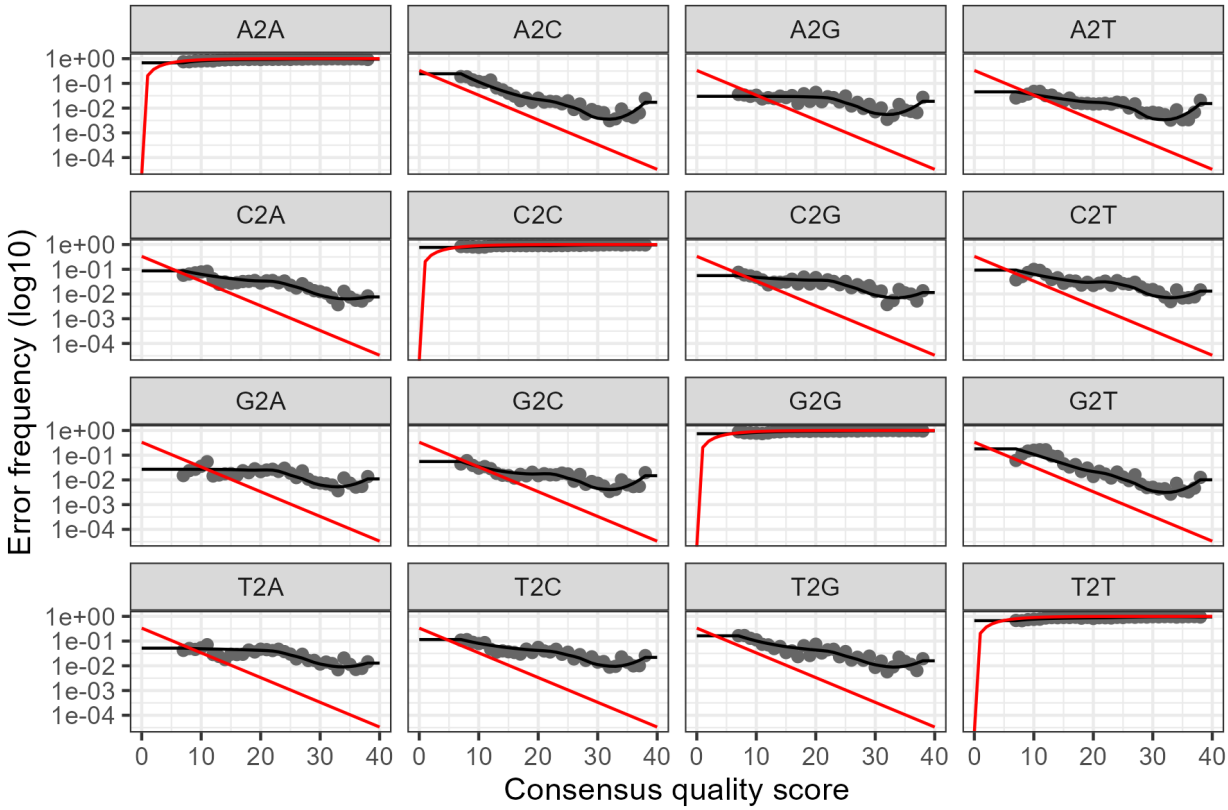

Figure A.14 Error rates for AMC COI forward reads. Generated error rate visualization for AMC COI forward reads after filtering, trimming, and dereplicating reads.

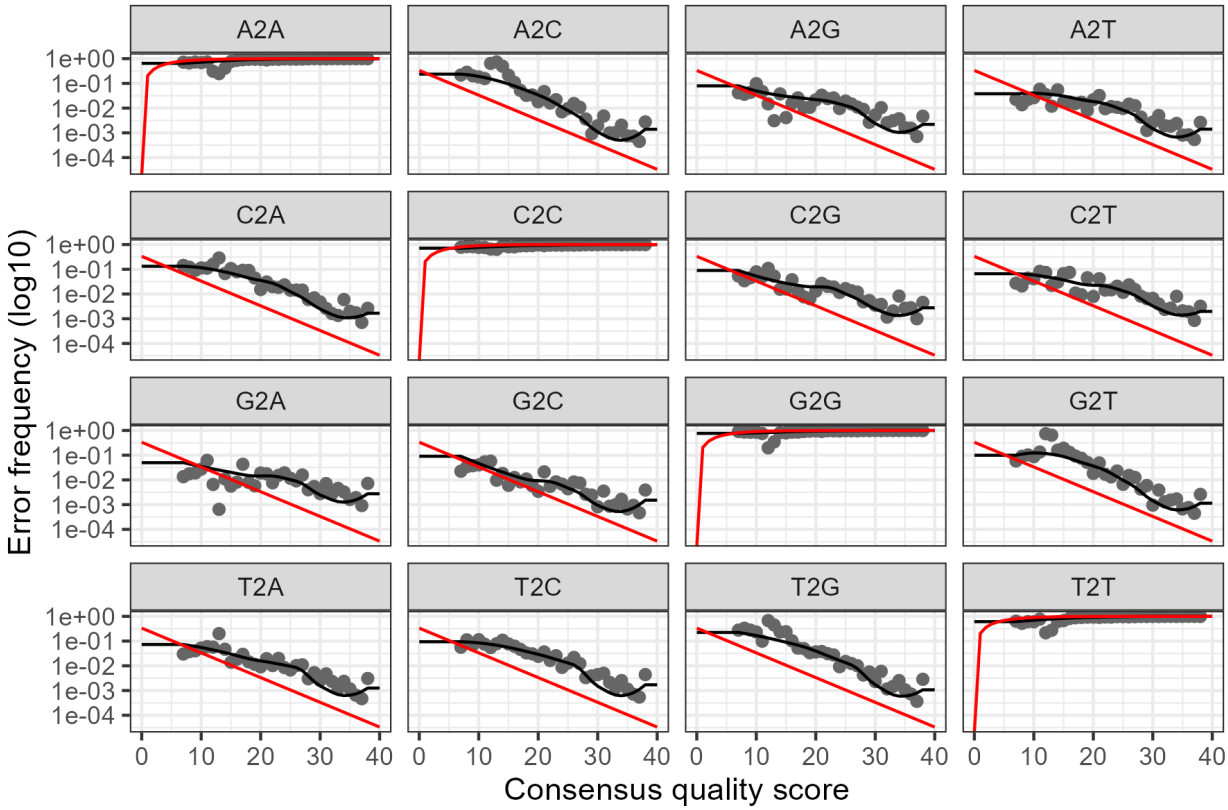

Figure A.15 Error rates for AMC COI reverse reads. Generated error rate visualization for AMC COI reverse reads after filtering, trimming, and dereplicating reads.

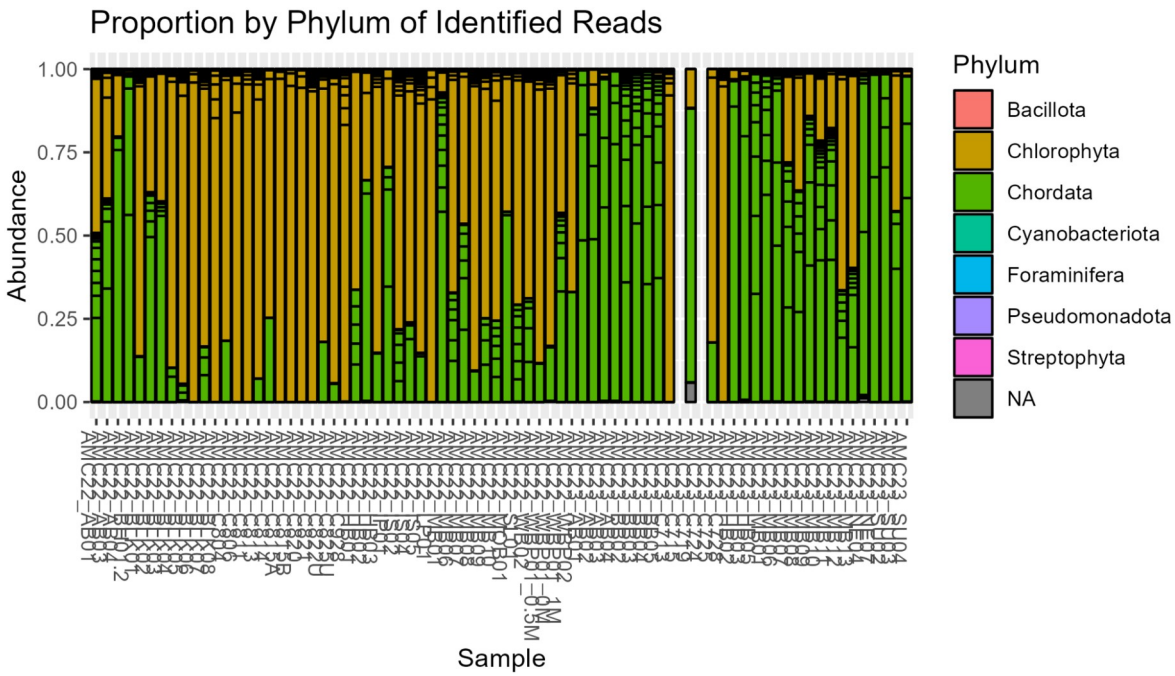

Figure A.16 Phylum identifications of taxa detected in AMC 12S samples. Proportional depiction of taxonomic classes detected in AMC 12S samples after cleaning reads and before the removal of contaminants and non-target taxa. Abundance on the y-axis refers to the number of ASVs identified to each class.

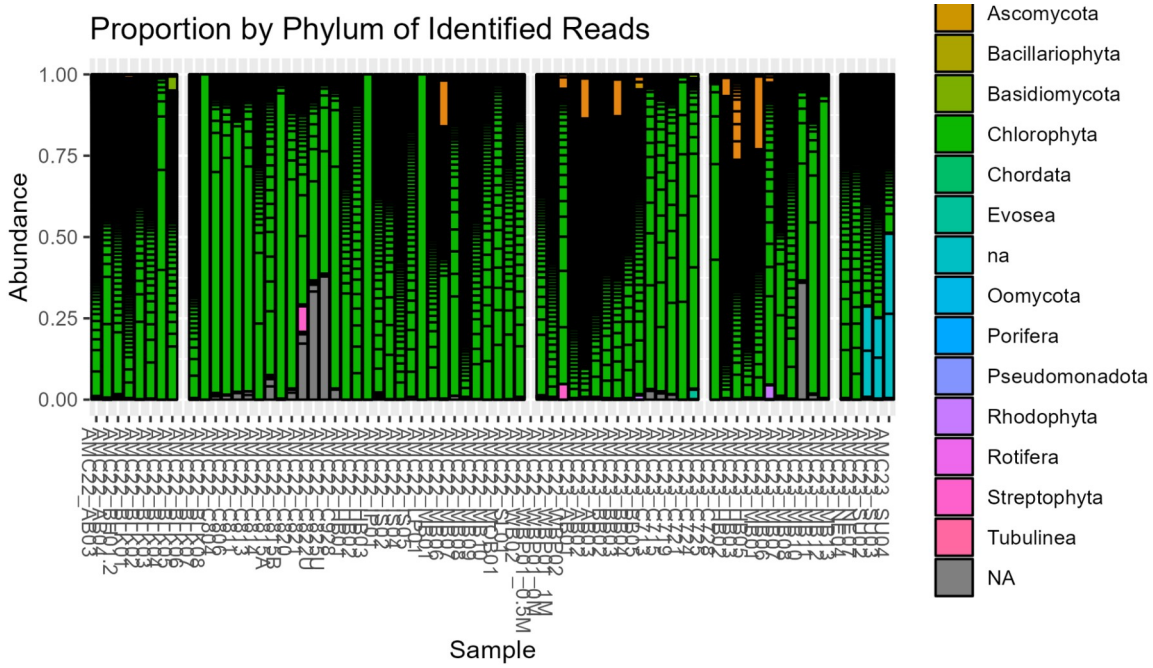

Figure A.17 Phylum identifications of taxa detected in AMC COI samples. Proportional depiction of taxonomic classes detected in AMC COI samples after cleaning reads and before removal of contaminants and non-target taxa. Abundance on the y-axis refers to the number of ASVs identified to each class.

Table A.11 Calculated alpha diversity metrics for AMC 12S eDNA samples. Values found for each alpha diversity test for each AMC eDNA sample using the 12S primer, calculated with the phyloseq package.

Sample	Chao1	Shannon	Simpson
AMC22_BLK01	1	0	0
AMC22_BLK02	7	0.72	0.32
AMC22_BLK03	6	0.26	0.09
AMC22_HB02	4	0.23	0.11
AMC22_MB01	18	1.32	0.54
AMC22_AB01	14	1.58	0.67
AMC22_AB03	7	1	0.56
AMC22_AB04	5	0.21	0.09
AMC22_BLK04	3	0.1	0.04
AMC22_BLK05	3	0.58	0.3
AMC22_BLK08	1	0	0
AMC22_HB01	5	1.56	0.78
AMC22_MB06	6	1.36	0.72
AMC22_MB07	7	1.52	0.73
AMC22_MB08	1	0	0
AMC22_MB09	4	0.89	0.52
AMC22_MB10	7	1.56	0.78
AMC22_BLK07	2	0.66	0.47
AMC23_ABO1	6	1.15	0.64
AMC23_BB04	11	1.68	0.76
AMC23_HB02	4	0.31	0.15
AMC23_HB03	10	0.53	0.3
AMC23_SU04	4	0.92	0.54
AMC23_AB04	4	0.61	0.32
AMC23_MB07	11	1.39	0.7
AMC23_MB08	10	1.33	0.67
AMC23_NE07	6	0.67	0.44
AMC23_SU02	5	0.91	0.46
AMC23_SU03	4	0.78	0.45
AMC23_AB02	8	1.07	0.6
AMC23_AB03	4	0.98	0.56
AMC23_MB06	10	1.22	0.65
AMC23_MB09	13	1.38	0.67
AMC23_MB10	16	1.64	0.67
AMC23_MB11	19	1.6	0.68
AMC23_MB12	7	1.24	0.6
AMC23_MB13	9	1.33	0.66
AMC23_NE04	7	0.84	0.53
AMC23_BB02	19	1.93	0.79
AMC23_BB03	14	1.43	0.63
AMC23_BB05	17	1.81	0.77
AMC23_HB05	15	1.7	0.78
AMC23_MB01	9	1.16	0.51

Table A.12 Calculated alpha diversity metrics for AMC COI eDNA samples. Values found for each alpha diversity test for each AMC eDNA sample using the COI primer, calculated with the phyloseq package.

Sample	Chao1	Shannon	Simpson
AMC22_BLK01	5	0.71	0.33
AMC22_BLK02	6	1.31	0.61
AMC22_BLK03	3	0.87	0.5
AMC22_MB01	2	0.65	0.46
AMC22_AB03	13	2.24	0.86
AMC22_AB04	11	1.95	0.82
AMC22_BLK07	1	0	0
AMC22_MB06	9	0.53	0.22
AMC22_MB08	2	0.68	0.49
AMC22_MB09	1	0	0
AMC23_ABO1	4	1.19	0.64
AMC23_BB04	15	2.49	0.9
AMC23_HB02	24	0.93	0.3
AMC23_HB03	23	2.29	0.86
AMC23_MB07	9	1.83	0.79
AMC23_SU03	3	0.96	0.57
AMC23_AB02	26	0.59	0.2
AMC23_AB03	5	1.52	0.77
AMC23_MB06	2	0.65	0.45
AMC23_NE04	1	0	0
AMC23_BB02	9	2.05	0.85
AMC23_BB03	21	0.68	0.23
AMC23_BB05	4	1.02	0.53
AMC23_HB05	20	1.82743729082173	0.67
AMC23_MB01	3	0.14	0.05

BIOGRAPHY OF THE AUTHOR

Beth Yima Davis was born in Rochester, New York in November 1998. She was raised primarily in Swansea, SC, and ultimately graduated from Canby High School in Canby, OR in 2016. She attended Oregon State University and graduated in 2020 with a Bachelor's degree in Integrative Biology, with a focus on Marine Biology. She enrolled in the Ecology and Environmental Sciences graduate program at the University of Maine in the fall of 2021. After receiving her degree, Beth will be pursuing her Ph.D. in Zoology at the University of Hawai'i at Manoa with Dr. Andy Rominger. Beth is a candidate for the Master of Science degree in Ecology and Environmental Sciences from the University of Maine in August 2024.