Spring 2023

Evaluating eDNA Metabarcoding as a Mic-Roe-Scopic Net to Catch Salmon Pathogens

Noah Burby

EVALUATING eDNA METABARCODING AS A MIC-ROE-SCOPIC NET TO CATCH SALMON PATHOGENS

by

Noah Joseph Burby

A Thesis Submitted in Partial Fulfillment
of the Requirements for a Degree with Honors
(Microbiology)

The Honors College

University of Maine

May 2023

Advisory Committee:

Dr. Erin Grey, Assistant Professor of Aquatic Genetics, School of Biology and Ecology, Advisor
Dr. Benjamin King, Assistant Professor of Bioinformatics, School of Molecular and Biomedical Sciences, Committee Member
Dr. Michael Kinnison, Professor of Evolutionary Applications, School of Biology and Ecology, Committee Member

ABSTRACT

Wild Atlantic salmon in the Gulf of Maine (GOM) is a Distinct Population Segment (DPS) that has been listed since 2000 as endangered by the U.S. Fish and Wildlife Service (USFWS) and the National Oceanic and Atmospheric Administration (NOAA). The current challenge is year-over-year decreases in the number of mature salmon returning to the Penobscot River for reproduction. Early detection of pathogen presence could allow for the identification of infection and the application of corrective measures. Environmental DNA (eDNA) is simply DNA that is collected from environmental samples (e.g., water, air, and soils), which consists of whole microorganisms and genetic material shed from macroorganisms (feces, skin, gametes, etc.). Purifying, testing, sequencing, and analyzing eDNA can help us rapidly identify the presence of these organisms in the sample. This project evaluates current methods' ability to detect salmon parasites from eDNA samples. Using computer-based alignment analysis, I first verified the potential of published primer sets to amplify known pathogen sequences *in silico*. Then, I tested amplification *in vitro* via quantitative PCR (qPCR) assay with gBlocks of target parasite sequences. Finally, I used DNA metabarcoding data from samples collected along the Maine coast to determine whether these pathogens were present. The metabarcoding analysis results will help identify the presence of these pathogens. Continued monitoring using this novel approach will further the goals of protecting the GOM Atlantic salmon DPS to survive in its native habitat.

ACKNOWLEDGEMENTS

The University of Maine recognizes that it is located on Marsh Island in the homeland of the Penobscot Nation, where issues of water and territorial rights, and encroachment upon sacred sites, are ongoing. Penobscot homeland is connected to the other Wabanaki Tribal Nations — the Passamaquoddy, Maliseet, and Mi'kmaq — through kinship, alliances, and diplomacy. The university also recognizes that the Penobscot Nation and the other Wabanaki Tribal Nations are distinct, sovereign, legal, and political entities with their own powers of self-governance and self-determination.

With the waters of the Penobscot River being the subject of this project, it is crucial to acknowledge the unique ethical considerations eDNA science poses. eDNA science is a powerful tool that can provide valuable insights into the river's biodiversity. Still, it is important to consider the potential ethical implications of the collected data. For example, the data could inform decisions about resource management, which could directly impact the local environment and communities. It is important to ensure that any data collected is used responsibly and ethically and that any potential impacts on the environment and people are considered.

The BC (Biocultural) Notice is a visible notification that there are accompanying cultural rights and responsibilities that need further attention for any future sharing and use of this material or data. The BC Notice recognizes the rights of Indigenous peoples to permission the use of information, collections, data and digital sequence information (DSI) generated from the biodiversity or genetic resources associated with traditional lands, waters, and territories.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Atlantic salmon populations in Maine have been in a state of critical decline for some time, despite the efforts of hatcheries to supplement the population and improve habitats. In 2000, the Gulf of Maine (GOM) Distinct Population Segment (DPS) was officially listed as endangered under the Endangered Species Act (ESA), with the National Oceanic and Atmospheric Administration (NOAA) and the United States Fish and Wildlife Service (USFWS) sharing joint listing authority (Flye et al., 2021). The various regulatory bodies, tribal, federal, and state managers, all have their own independent authorities in making decisions related to the recovery of the species, meaning that effective monitoring is critical to the communication and coordination among the different groups and is essential for successful recovery efforts.

Atlantic Salmon, historically found in North American rivers and streams from Ungava Bay in Canada to Long Island Sound in the United States, has dramatically decreased in population over the years. In the State of Maine, where Atlantic Salmon were once found in 34 rivers, only 2000 return annually, a mere 1% of the historic population (Saunders, 2006). This species is anadromous, meaning they occupy freshwater habitats during their juvenile life stages and marine environments during their adult life stages. When reproductively mature, they return to their natal streams to spawn after spending one to three winters at sea. Unfortunately, due to their long migrations and habitat sensitivity, Atlantic Salmon are vulnerable to population declines and habitat disruptions from dams, pollution, pathogens, and over-exploitation (Flye et al., 2021).

An ongoing risk to these salmon is a parasitic infection. In the marine environment, salmon are part of a multi-host infectious disease systems. Reservoir host

species can maintain the parasite and support transmission to other host species. Rainbow trout is an example host species for *Yersinia ruckeri*, a pathogen that affects wild Atlantic salmon, which is of regulatory concern in Maine. Transmission from reservoir hosts is a significant factor in the emergence or re-emergence of infectious diseases and can also lead to wildlife extinction. In the marine environment, the low barriers to parasite dispersal and the large migrations of marine fauna make multi-host parasite transmission more likely (Shea et al., 2020). Fishing and aquaculture can deplete wild fish populations or introduce new reservoir host populations, changing the dynamics of multi-host transmission. Despite the importance of reservoir hosts for parasite spread, disease emergence, and biodiversity conservation, there has been very little empirical assessment of how reservoir hosts influence parasite communities in natural and domestic systems and only for a limited range of parasites (Shea et al., 2020).

Tracking the dynamics of multi-host parasite transmission has been a critical challenge for aquaculture, mainly as domesticated fish can act as a reservoir host for parasites in coastal environments (Krkošek, 2017; Nowak, 2007). Seafood demand and plateaued fisheries have led to the rapid expansion of salmon aquaculture, and the abundance of domesticated salmon now exceeds wild populations in areas where they coexist. Global production of domestic salmon even exceeds wild catch (FAO, 2018; Krkošek, 2017; Lafferty et al., 2015). The benefits of aquaculture development will largely depend on who has the resources to participate in the sector and how it is managed. In Maine, USA, commercial fishermen increasingly use aquaculture to expand and diversify their livelihoods (Stoll et al., 2019). Studies have been conducted to investigate macroparasite transmission between wild and domestic populations in

salmonid aquaculture (Shea et al., 2020). However, few studies explore these pathogens in the environment where they can be shed and acquired by both domestic and wild fish populations. This is a concern for the conservation of wild salmon in Maine, as pathogen transmission from domesticated salmon could have a detrimental effect.

In this project, I investigated 21 pathogens from Shea et al. (2020) and 6 from the Maine Department of Marine Resources (MDMR, 2021) with three bacteria that overlapped (*R. salmoninarum, A. salmonicida, and Y. ruckeri*) for a total of 24 pathogens. An emphasis was placed on the MDMR pathogens as they pose a current and ongoing risk in the GOM. The diseases and causative pathogens are Bacterial Kidney Disease (*Renibacterium salmoninarum*), Furunculosis (*Aeromonas salmonicida*), Enteric Redmouth (*Yersinia ruckeri*), Whirling Disease (*Myxobolus cerebralis*), Ceratomyxosis (*Ceratomyxa shasta*), Ceratomyxosis (*Ceratomyxa shasta*) (MDMR, 2021). The pathogens from Shea et al. 2020 were studied in the Pacific Northwest of the United States and Canada but pose an existential threat to the GOM Atlantic Salmon DPS due to migration due to global trade and climate change.

Current quantitative testing for pathogens described in chapter 24, “Importation of Live Marine Organisms” by the Maine Department of Marine Resources (MDMR), 2021, occurs only at hatchery facilities. Two methods are employed; Reproductive fluids shall be sampled at the 100% level or lethal sampling at a 10% assumed prevalence up to a maximum of 30 fish. For a 2% assumed prevalence level, reproductive fluids can be collected by trained facility personnel under the inspector's direction using a specimen chain of custody form. The sampling must be done in a safe and humane manner

(MDMR, 2021). These samples are tested for pathogen presence using RT-qPCR and indirect fluorescent antibody testing.

Environmental DNA (eDNA) metabarcoding offers a way to detect the presence of parasites of concern known to the Maine DMR and other pathogens that could migrate to the GOM that are not yet on MDMR's radar. eDNA has been defined as the DNA that can be extracted from bulk environmental samples, such as soil, air, or water, without first isolating any target organisms (Seymour, 2019). eDNA contains the cellular DNA from living organisms and extracellular DNA originated by cell lysis and extrusion. The biological functions of eDNA include acting as a structural component in biofilms, a nutrient source, and horizontal gene transfer. eDNA-based monitoring is advantageous compared to traditional survey methods in terms of cost, non-invasiveness, and higher efficiency in detecting organisms, including rare and invasive species (Amarasiri et al., 2021; Wang et al., 2019).

Detecting pathogens, parasites, and other symbionts in environmental samples through eDNA is becoming increasingly important to gain information about their presence and activity. eDNA metabarcoding is a method of taxonomic identification used to identify multiple species in a mixed sample of community DNA or eDNA. This method involves PCR-amplifying and sequencing the sample on a high-throughput platform such as Illumina or Ion Torrent. This process allows for identifying multiple species from a single sample, providing a powerful tool for taxonomic identification (Amarasiri et al., 2021; Bass et al., 2023). This has great potential to be used as a proxy for infection of host organisms in connected habitats, pathogen monitoring and

surveillance, and early warning systems for disease (Bass et al., 2023). This is a very useful tool for understanding the spread of diseases and helping to prevent them.

The development of molecular markers called primers is invaluable for detecting aquatic taxa eDNA. For accurate detection of target species or assemblages, the primers must be capable of ensuring precision. In the past, specific primers were designed for a single species, while universal primers were created for community assemblages. Recently, researchers have gone a step further and developed universal primers for 12S/16S/18S rDNA, Cytb, COI, and even mitogenome sequences to detect aquatic communities (Huang et al., 2022). This advancement has greatly improved the accuracy of aquatic taxa eDNA detection.

In this study, we assessed the effectiveness of existing eDNA metabarcoding analysis using 16S and 18S universal primers in the Gulf of Maine (GOM) for amplifying and detecting pathogens of concern to the Atlantic salmon Distinct Population Segment (DPS) (Comeau et al., 2011; Parada et al., 2016; Walters et al., 2015). We want to better understand this method's potential to detect and monitor the presence of these pathogens in the GOM. By doing so, we hope to provide valuable insights that can be used to inform management and conservation efforts for this important species.

METHODS AND MATERIALS

Maine-eDNA Sample Collection and Metabarcoding

Environmental DNA sample collection occurs monthly at index sites, preset estuarine locations along the Maine coast, one hour after high tide during slack tide. Index sites are divided among four regional institutions in the Maine-eDNA project: the University of Southern Maine/Gulf of Maine Research Institute, Bigelow Laboratory for Ocean Sciences, the University of Maine Orono/Maine Center for Genetics Center and the Environment, and the University of Maine Machias. A 5 L Niskin is used to sample at 1 m above the bottom of the estuary, the middle of the water column, and the surface. The three 5 L samples are dispensed into 5 L amber Nalgene bottles. Sample data, including GPS coordinates, time, air temperature, turbidity, and depth, are collected in ArcGIS. Bottles are placed in coolers with ice to slow DNA denaturation. Samples are then rushed to the regional lab for immediate filtration.

The filtration of these samples utilizes three types of filters: Nitex, glass fiber, and Millipore. The three types have progressively smaller pore sizes for various downstream analyses. The 5 μm size of the glass fiber is the standard for DNA collection. Filtering begins with 2 L of the 5 L sample passed through the Nitex filter via gravity filtration. Then 2 L of the 5 L sample is filtered through the glass fiber filter via vacuum filtration. Lastly, 0.5 L of the 5 L is passed through the Millipore filter via vacuum filtration. The remaining 0.5 L balances the user error, and the leftover sample water is discarded. Filters are rolled, placed in cryogenic microcentrifuge tubes, and frozen or directly extracted.

DNA extraction for Next Generation Sequencing (NGS) uses glass fiber filters through which 2 L of the environmental sample has been filtered. This protocol uses the Qiagen DNeasy PowerWater Toolkit protocol. Once the DNA is in its final elution, it is purified to remove tannins, nucleases, and PCR inhibitors using the Zymo Research One-Step PCR Inhibitor Removal Kit. At this point, the DNA can either be frozen for later or amplified by PCR.

Amplification of the DNA is achieved using 16S Parada or 18S Comeau universal primers with NGS adapter tails, and a probe binds to the template DNA between primers. The probe is fluorescently tagged for detection in RT-qPCR. Once amplified, samples are run on an ethidium bromide gel electrophoresis to ensure only the target DNA is sequenced downstream. The gel is cut to excise only the target band. The DNA is extracted from the gel using the Qiagen QIAquick Gel Extraction Kit.

The DNA library is then prepared for Illumina MiSeq sequencing using the Illumina Nextera XT DNA Library Preparation Kit. Once prepared, it is loaded into a cartridge with other sequencing reagents. The cartridge is loaded into the MiSeq, and the chip is removed from the ethanol it is stored in, rinsed with DI water, dried without touching the channels, and inserted into the MiSeq. The sequencing program is initiated and allowed to run to completion.

Target Pathogen Analysis

A list of pathogens was created to determine if the universal primer eDNA sequencing would capture salmon pathogens. First consisting of target pathogens from the inspiratory paperShea et al. (2020), this list was expanded to include pathogens of

regulatory concern for the Maine DMR. Pathogens were then divided into 16S (Bacteria) and 18S (Eukaryotic Parasites) as shown in Table 1.

16S/18S	Pathogen Name	Disease Caused	Reference Source
16S	*Flavobacterium psychrophilum*	Bacterial Coldwater Disease	Shea et al., 2020
16S	*Vibrio anguillarum*	Vibriosis	Shea et al., 2020
16S	*Vibrio salmonicida*	Cold-Water Vibriosis	Shea et al., 2020
16S	*Tenacibaculum maritimum*	Tenacibaculosis	Shea et al., 2020
16S	*Yersinia ruckeri*	Enteric Redmouth	Shea et al., 2020 & MDMR, 2021
16S	*Moritella viscosa*	Winter Ulcers	Shea et al., 2020
16S	*Candidatus Branchiomonas cysticola*	Proliferative Gill Disease	Shea et al., 2020
16S	*Renibacterium salmoninarum*	Bacterial Kidney Disease	Shea et al., 2020 & MDMR, 2021
16S	*Aeromonas salmonicida*	Furunculosis	Shea et al., 2020 & MDMR, 2021
16S	*Candidatus Branchiomonas cysticola*	Epitheliocystis	Shea et al., 2020
16S	*Piscirickettsia salmonis*	Piscirickettsiosis	Shea et al., 2020
16S	*Piscichlamydia salmonis*	Epitheliocystis	Shea et al., 2020
18S	*Ichthyophonus hoferi*	Ichthyophoniasis	Shea et al., 2020
18S	*Kudoa thyrsites*	Obligate Parasite	Shea et al., 2020
18S	*Sphaerothecum destruens*	Parasite (the Rosette Agent)	Shea et al., 2020
18S	*Neoparamoeba perurans*	Amoebic Gill Disease	Shea et al., 2020
18S	*Parvicapsula minibicornis*	Parasitic Kidney Disease	Shea et al., 2020
16S/18S	Pathogen Name	Disease Caused	Reference Source

18S	*Nucleospora salmonis*	Parasitic Kidney Disease	Shea et al., 2020
18S	*Nanophyetus salmincola*	Obligate Parasite	Shea et al., 2020
18S	*Myxobolus arcticus*	Whirling Disease	Shea et al., 2020
18S	*Loma salmonae*	Microsporidial Gill Disease	Shea et al., 2020
18S	*Myxobolus cerebralis*	Whirling Disease	MDMR, 2021
18S	*Ceratomyxa shasta*	Ceratomyxosis	MDMR, 2021
18S	PKX unclassified myxozoan	Proliferative Kidney Disease	MDMR, 2021

Table 1. A compiled list of Salmon Pathogens likely amplified by 16S Parada or 18S Comeau universal primers with the diseases caused and the source article. This list includes pathogens with no mismatches and those with some that are hypothesized to work.

In silico analysis

To assess the feasibility of the Maine-eDNA Comeau 18S primers amplifying the DNA of these pathogens, a few alignments of the 18S sequences were run using *I. hoferi* and *K. thyrsites* as templates using the automatic MAFFT program. The result of these test alignments showed regions of promise, and with that, the primers 572F and 1009R from the Comeau set were added to Benchling and the primer attachment tool was used. Successful alignment required changing the Tm parameters from SantaLucia 1998 to Modified Breslauer 1986 (Phusion, Phire, DyNAzyme). After, all primers previously attached to each individual sequence using the other Tm protocol were deleted. Next, the primers tool was activated, “attach existing” was selected, and used primers in the 18S Benchling folder (572F and 1009R), with a nucleotide type of DNA, with at least 10 matching bases, separated by ≤ 2 mismatches, no more than 10 mismatches, Tm above 48ºC and under 100ºC, leaving the two checkboxes empty, and 3' position by default 1-end #. I then selected “find binding sites” and selected F/R sites with no mismatches or

the least amount for both Shea et al., 2020 and MDMR, 2021 pathogens. Then the F/R primers were linked, *in silico* PCR data was collected, and the data recorded.

To assess the feasibility of the Maine-eDNA Parada 16S primers amplifying the DNA of these pathogens, the 16S bacteria sequences were added to the Benchling folder labeled 16S (Bacteria) using the import sequence feature on Benchling. Next, the Parada primers were added to the 16S (Bacteria) folder for use in the alignments to come. Then, going through each individual genome, the Parada primer set (515F and 926R) was applied. The Tm parameters were also from SantaLucia 1998 to Modified Breslauer 1986 (Phusion, Phire, DyNAzyme). Then, going to the primers tool, "attach existing" was selected, and used the primers in the 16S (Bacteria) Benchling folder (515F and 926R), with a nucleotide type of DNA, with at least 10 matching bases, separated by ≤ 2 mismatches, no more than 10 mismatches, Tm above 48ºC and under 100ºC, leaving the two checkboxes empty, and 3' position by default 1-end #. Next, "find binding sites" was selected and F/R sites with no mismatches or the least amount were identified for both Shea et al., 2020 and MDMR, 2021 pathogens. Then, F/R primers were linked, in silico PCR data collected, and the data recorded.

<u>*In vitro* analysis</u>

To assess the results of the *in silico* analysis using Benchling we needed to see if the primers would amplify the DNA segments in PCR. We ordered synthetic genes (gBlocks) of 7 pathogens: 4 of the 16S pathogens (*Yersinia ruckeri* [BR], *Candidatus Syngnamydia salmonis* [CSS], *Piscirickettsia salmonis* [PS], & *Piscichlamydia salmonis* [PY]) and 3 of the 18S pathogens (*Nanophyetus salmincola* [NC], *Ceratomyxa shasta* [CS], & PKX unclassified myxozoan [PKD]). The dried gBlocks were spun down at max

rpm for 3 min to prevent loss of DNA and hood contamination. gBlocks were then resuspended to $1e^{10}$ copies/µL with TE Buffer using the equation: fmole*1.0E-15*(6.022E23 copies)1.0E10copies/µL=amount TE Buffer (µL) from the UMaine Guide for gBlock Resuspension and Serial Dilution. Then, using the guide, the gBlocks were diluted 1:5 to 31250, 6250, 1250, 250, 50, & 10 copies/µL.

The next step of *in silico* analysis was to run an SYBER dye qPCR of the gBlocks using the 18S and 16S Nextera primers. These 20 µL reactions used 10 µL Bio-Rad SsoFast EvaGreen 2x Supermix, 8 µL nuclease-free PCR water, 0.5 µL forward/sense primer, 0.5 µL reverse/antisense primer, and 1 µL of template DNA. These reactions were loaded onto a Bio-Rad CFX thermocycler, and the thermal programs described in Table 2 were applied.

16S qPCR Thermal Program			18S qPCR Thermal Program		
Cycling Step	Temperature	Time	Cycling Step	Temperature	Time
Enzyme Activation	98ºC	2 min	Enzyme Activation	98ºC	2 min
Denaturation	98ºC	10 sec	Denaturation	98ºC	10 sec
Annealing	55ºC	30 sec	Annealing	50ºC	20 sec
Extension	60ºC	30 sec	Extension	60ºC	20 sec
Repeat Denaturation-Extention 50 times			Repeat Denaturation-Extention 50 times		
Melt Curve	65-95ºC	5 sec/step	Melt Curve	65-95ºC	5 sec/step

Table 2. qPCR thermal programs for 16S and 18S gBlock Nextera verification assay.

After qPCR, Qc and melt curve data were collected for analysis. Then, the qPCR product was run on an SYBER gel to verify the length of amplified segments. The gel

was imaged on an SYBER ultraviolet (UV) light table (Figures 1 & 2). Next, to determine the effect of the Nextera tails in the amplification process, the gBlocks were amplified using Comeau/Parada primers without the tails. The previously used supermix had run out, so Bio-Rad SsoAdvanced Universal SYBER Green 2x Supermix was used for this test. These 20 µL reactions used 10 µL Bio-Rad SsoAdvanced Universal SYBER Green 2x Supermix, 8 µL nuclease-free PCR water, 0.5 µL forward/sense primer, 0.5 µL reverse/antisense primer, and 1 µL of template DNA. These reactions were loaded onto a Bio-Rad CFX thermocycler, and the thermal programs described in Table 3 were applied.

16S/18S qPCR Thermal Program		
Cycling Step	Temperature	Time
Enzyme Activation	98ºC	2:45 min
Denaturation	98ºC	15 sec
Annealing/ Extension	60ºC	45 sec
Repeat Denaturation-Extention 39 times		
Melt Curve	65-95ºC	5 sec/step

Table 3. qPCR thermal program for 16S and 18S gBlock Non-Nextera control assay.

In vivo analysis

After completing *in vitro* assays, 18S Metabarcoding data from the Hampden, Bucksport, and Chemo Pond index sites were analyzed. The samples included: Hampden March 2021 (eLP_E01_21w_0213), Hampden March 2021 (eLP_E01_21w_0215), Bucksport May 2021 (eLP_E04_21w_0224), Bucksport June 2021 (eLP_E04_21w_0228), Chemo Pond Control May 2021 (eLP_L02_21w_0214), Chemo Pond May 2021(eLP_L02_21w_0215), Chytrid Mock, and Negative Control. The raw

metabarcoding data was uploaded to Galaxy, a web-based computing software. Then, the amplicons of the salmon pathogens were also uploaded to the Galaxy program, and BLASTn was run against the index site data.

RESULTS

In silico analysis

The 16S *in silico* analysis of bacterial pathogens (Table 3) resulted in eleven pathogens with no mismatches with the primers and three with mismatches away from the 3' end. *Candidatus Syngnamydia salmonis* (CSS) has two mismatches, one in the foreward primer and one in the reverse primer. This is also the case for *Piscichlamydia salmonis* (PY), but *Piscirickettsia salmonis* (PS) has only one mismatch in the reverse primer. All amplicons are between 407-414 bp. The position on the 16S gene was consistent around +/500 and -/900. GC content was between 47.16% and 56.76%. The absolute difference in primer melting temperatures was 14.3ºC.

The 18S *in silico* analysis of eukaryotic parasites (Table 4) resulted in seven pathogens with no mismatches with the primers and five with mismatches away from the 3' end. *Myxobolus arcticus* has two mismatches, one in the forward primer and one in the reverse primer. This is also the case for *Loma salmonae*. PKX unclassified myxozoan (PKD) has only one mismatch in the reverse primer, *Nanophyetus salmincola* (NC) has one in the forward primer. Meanwhile, *Nucleospora salmonis* has four mismatches, two in both the forward and reverse primers. All amplicons are between 274-606 bp. The position on the 18S gene was inconsistent, around +/550 and -/1000. GC content was between 33.50% and 50.55%. The absolute difference in primer melting temperatures was 14.3ºC.

In vitro analysis

The 16S *in vitro* SYBER qPCR was completed twice to confirm the initial test results. These results show robust amplification of *Candidatus Syngnamydia salmonis*

(CSS), *Piscirickettsia salmonis* (PS), *Piscichlamydia salmonis* (PY), and *Yersinia ruckeri* (BR) at concentrations 1250, 250, 50, and 10 copies/µL. Results were difficult to distinguish at the concentrations of 31250 and 6250 copies/µL. This test also showed a 3D structure created by dimerization as the no template controls have robust amplification, a melt curve, and a peak at 85.50ºC.

The 16S PCR products were run on an SYBER agar gel electrophoresis rig (Figure 1). Loading from left to right: ladder, CSS 31250 copies/µL, CSS 6250 copies/µL, CSS 50 copies/µL, PS 31250 copies/µL, PS 6250 copies/µL, PS 50 copies/µL, PY 31250 copies/µL, PY 6250 copies/µL, PY 50 copies/µL, 16S no template control (ntc), BR 31250 copies/µL, BR 6250 copies/µL, BR 1250 copies/µL, BR 250 copies/µL, BR 50 copies/µL, BR 10 copies/µL, blank, blank, and ladder. This gel was run at 130 V for 25 min. Bands were identified in the 50 copies/µL channels of CSS, PS, and PY. Also, BR 6250 copies/µL, BR 1250 copies/µL, BR 250 copies/µL, BR 50 copies/µL, and BR 10 copies/µL around 400 bp. Another result of this test was identifying a primer dimer band towards the bottom of the gel.

18S *in vitro* SYBER qPCR was completed twice to confirm the initial test results. These results show robust amplification of *Ceratomyxa shasta* (CS), PKX unclassified myxozoan (PKD), and *Nanophyetus salmincola* (NC) at concentrations 6250, 1250, 250, 50, and 10 copies/µL. Results were difficult to distinguish at the concentration of 31250 copies/µL. This test also showed a 3D structure created by dimerization as the no template controls have robust amplification, a melt curve, and a peak at 80.67ºC.

The 18S PCR products were run on an SYBER agar gel electrophoresis rig (Figure 2). Loading from left to right: ladder, CS 31250 copies/µL, CS 6250 copies/µL,

CS 50 copies/µL, PKD 31250 copies/µL, PKD 6250 copies/µL, PKD 50 copies/µL, NC 31250 copies/µL, NC 6250 copies/µL, NC 50 copies/µL, 18S no template control (ntc), ladder. This gel was run at 130 V for 25 min. Bands were identified in the 31250 and 6250 copies/µL and faintly at 50 copies/µL. This test also identified a primer dimer band towards the bottom of the gel.

The Cq data from the two 16S Nextera PCR were plotted in Figure 3; the Cq data from the two 18S Nextera PCR were plotted in Figure 4; the Cq data from the non-Nextera 16S PCR were plotted in Figure 5; the Cq data from the non-Nextera 18S PCR were plotted in Figure 6. The R^2 values from the duplicate Nextera PCRs were averaged, and the slopes from the linear trendlines were used in the PCR efficiency equation: -1+10(-1/slope) = PCR Efficiency. These data were compiled in Table 6.

In vitro analysis

The results of the BLASTn of 18S pathogen amplicons against 18S metabarcoding data from select index sites found no hits. This indicates that none of the 18S pathogens were present in detectable quantities at the times and locations where the samples were collected.

Pathogen	515F Position	515F Melting Temp (ºC)	926R Position	926R Melting Temp (ºC)	GC Content (%)	Product Length (bp)	Tm Difference (ºC)	Mismatches
Flavobacterium psychrophilum	+/ 492	72.9	-/ 861	58.5	47.42	407	-14.3	0
Vibrio anguillarum	+/ 506	72.9	-/ 880	58.5	53.4	412	-14.3	0

Vibrio salmonicida	+/ 533	72.9	-/ 907	58.5	53.88	412	-14.3	0
Tenacibaculum maritimum	+/ 500	72.9	-/ 869	58.5	47.17	407	-14.3	0
Yersinia ruckeri	+/ 513	72.9	-/ 886	58.5	54.26	411	-14.3	0
Candidatus Syngnamydia salmonis	+/ 528	72.9	-/ 903	58.5	47.94	413	-14.3	2: 1F/1R
Piscirickettsia salmonis	+/ 526	72.9	-/ 902	58.5	51.27	414	-14.3	1: 1R
Moritella viscosa	+/ 529	72.9	-/ 903	58.5	52.43	412	-14.3	0
Candidatus Branchiomonas cysticola	+/ 536	72.9	-/ 908	58.5	50.98	410	-14.3	0
Piscichlamydia salmonis	+/ 504	72.9	-/ 878	58.5	50.73	412	-14.3	2: 1F/1R
Renibacterium salmoninarum	+/ 1145	72.9	-/ 1521	58.5	56.76	414	-14.3	0
Aeromonas salmonicida	+/ 505	72.9	-/ 879	58.5	54.37	412	-14.3	0
Renibacterium salmoninarum	+/ 488	72.9	-/ 864	58.5	56.76	414	-14.3	0
Aeromonas salmonicida	-/ 949	72.9	+/ 575	58.5	54.61	412	14.3	0

Table 4. 16S Pathogen-Primer alignment. Salmon pathogen DNA from GenBank was added to Benchling and assessed for the likelihood of 16S Nextera Primers using the Modified Breslauer 1986 (Phusion, Phire, DyNAzyme) parameters. This produced the binding sites, melting temperatures, the difference of temperatures, and mismatches of

both forward and reverse primers. Also, the GC content and length of the amplicon. Different reference sequences were used for *Renibacterium salmoninarum* and *Aeromonas salmonicida* for analyzing Shea et al. pathogens and MDMR pathogens, resulting in 2 entries.

Pathogen	572F Position	572F Melting Temp (ºC)	1009R Position	1009R Melting Temp (ºC)	GC Content (%)	Product Length (bp)	Tm Difference (ºC)	Mismatches
Ichthyophonus hoferi	+/ 591	62.6	-/ 992	55.6	40.18	438	-7.1	0
Kudoa thyrsites	+/ 549	62.6	-/ 941	55.6	45.22	429	-7.06	0
Nucleospora salmonis	+/ 404	62.6	-/ 627	54.7	50.39	259	-7.9	4: 2F/2R
Sphaerothecum destruens	+/ 575	62.6	-/ 977	55.6	41.69	439	-7.1	0
Nanophyetus salmincola	+/ 522	62.6	-/ 1003	55.6	49.42	518	-7.1	1: 1F
Neoparamoeba perurans	+/ 623	62.6	-/ 1192	55.6	33.50	606	-7.1	0
Myxobolus arcticus	+/ 581	62.6	-/ 1132	54.7	45.32	588	-7.9	2: 1F/1R
Parvicapsula minibicornis	+/ 619	62.6	-/ 1124	55.6	44.46	542	-7.1	0
Loma salmonae	+/ 447	62.6	-/ 685	55.6	50.55	275	-7.1	2: 1F/1R
Myxobolus cerebralis	+/ 583	62.6	-/ 1117	55.6	45.53	571	-7.1	0
Ceratomyxa shasta	+/ 572	62.6	-/ 990	55.6	38.68	455	-7.1	0
PKX unclassified myxozoan	+/ 559	62.6	-/ 1026	55.6	44.64	504	-7.1	1: 1R

Table 5. 18S Pathogen-Primer alignment. Salmon pathogen DNA from GenBank was added to Benchling and assessed for the likelihood of 18S Nextera Primers using the Modified Breslauer 1986 (Phusion, Phire, DyNAzyme) parameters. This produced the binding sites, melting temperatures, the difference of temperatures, and mismatches of both forward and reverse primers. Also, the GC content and length of the amplicon.

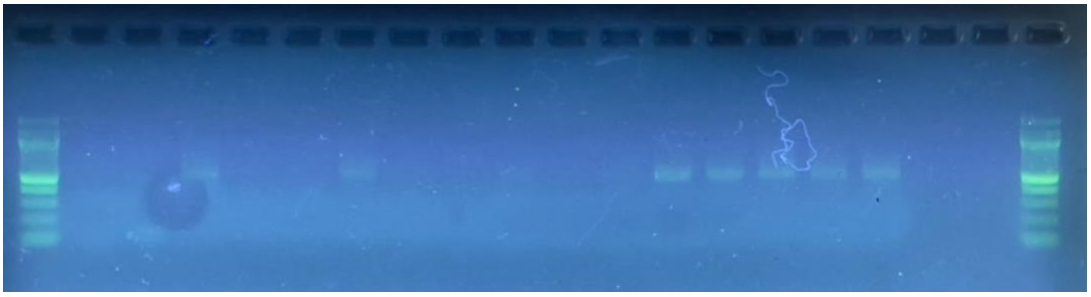

Figure 1. 16S PCR product gel electrophoresis. Loading from left to right: ladder, *Candidatus Syngnamydia salmonis* (CSS) 31250 copies/µL, CSS 6250 copies/µL, CSS 50 copies/µL, *Piscirickettsia salmonis* (PS) 31250 copies/µL, PS 6250 copies/µL, PS 50 copies/µL, *Piscichlamydia salmonis* (PY) 31250 copies/µL, PY 6250 copies/µL, PY 50 copies/µL, 16S no template control (ntc), *Yersinia ruckeri* (BR) 31250 copies/µL, BR 6250 copies/µL, BR 1250 copies/µL, BR 250 copies/µL, BR 50 copies/µL, BR 10 copies/µL, blank, blank, and ladder. This gel was run at 130 V for 25 min.

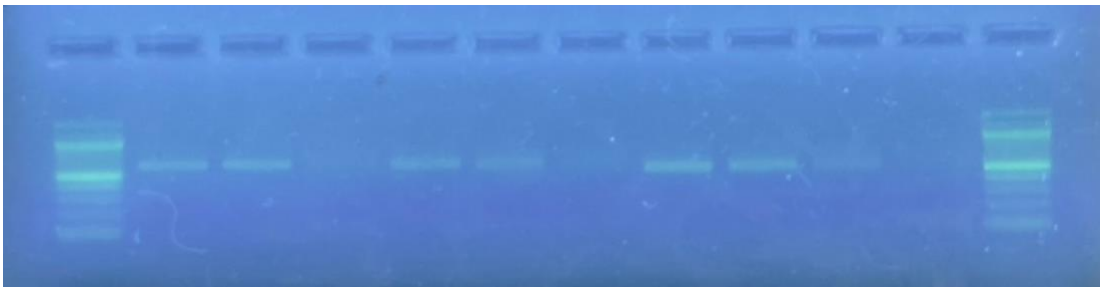

Figure 2. 18S PCR product gel electrophoresis. Loading from left to right: ladder, *Ceratomyxa shasta* (CS) 31250 copies/µL, CS 6250 copies/µL, CS 50 copies/µL, PKX unclassified myxozoan (PKD) 31250 copies/µL, PKD 6250 copies/µL, PKD 50 copies/µL, *Nanophyetus salmincola* (NC) 31250 copies/µL, NC 6250 copies/µL, NC 50 copies/µL, 18 no template control (ntc), ladder. This gel was run at 130 V for 25 min.

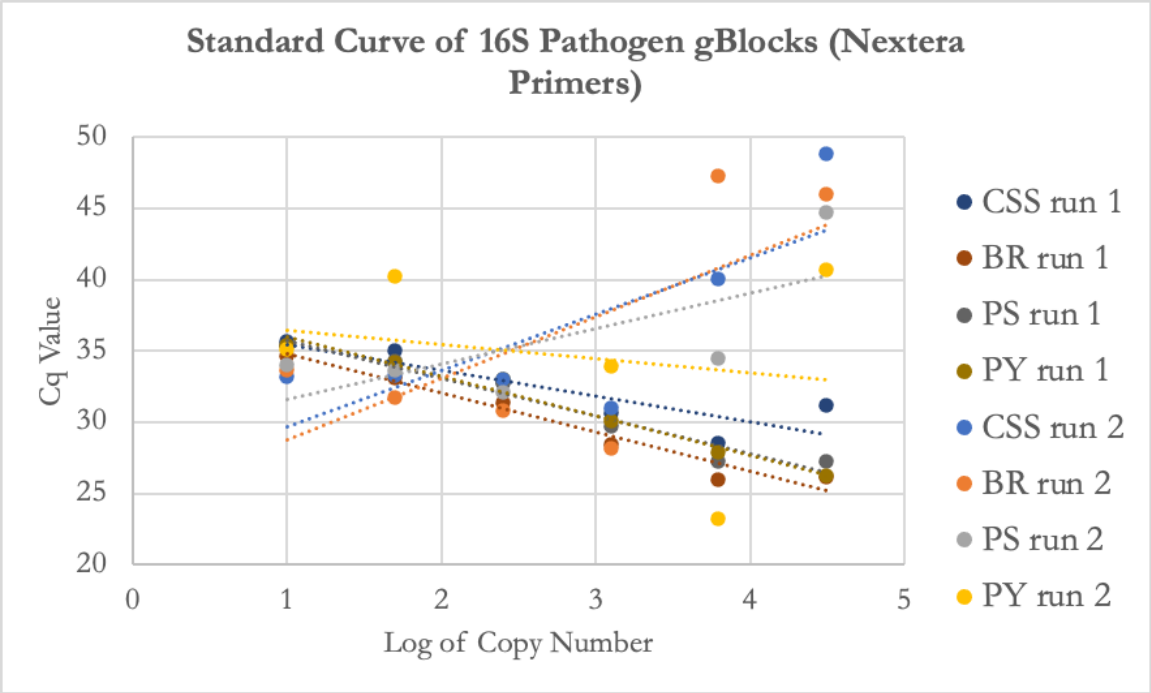

Figure 3. Standard Curve of 16S Pathogens with Nextera tails. gBlocks of pathogen amplicons were diluted to 31250, 6250, 1250, 250, 50, and 10 copies/µL and amplified by SYBER Green qPCR. This test was run in duplicate. Cq values were plotted against the log of the copy number. Linear trendlines produced R^2 values, and the slopes were used to calculate the PCR efficiencies.

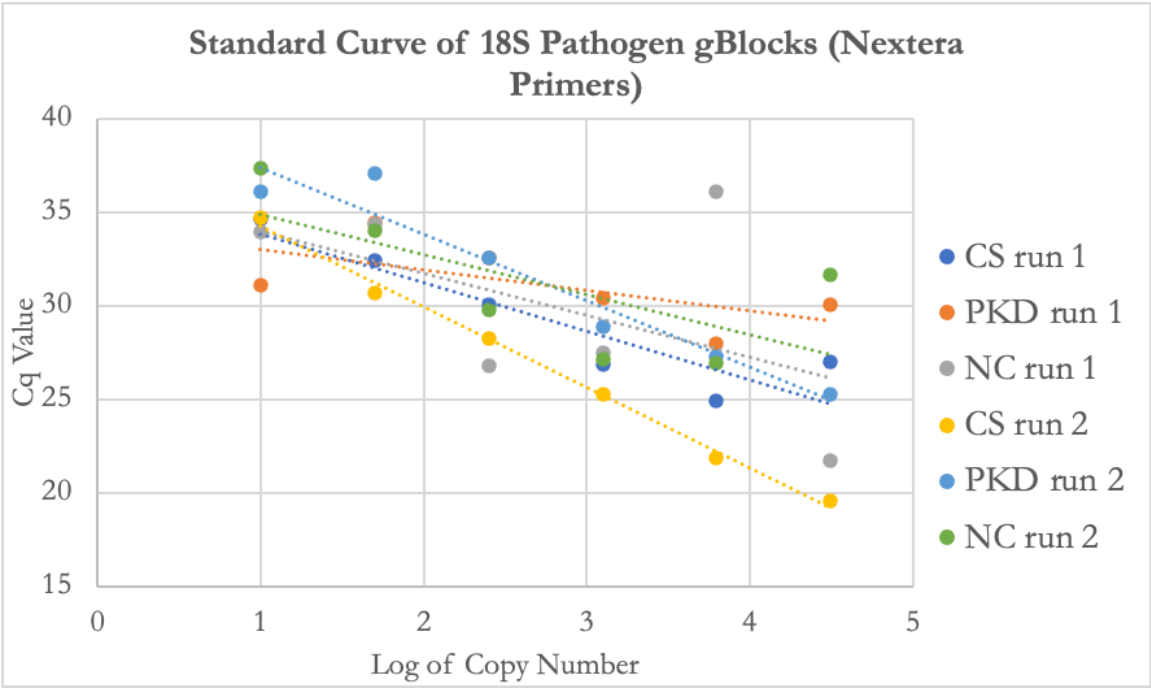

Figure 4. Standard Curve of 18S Pathogens with Nextera tails. gBlocks of pathogen amplicons were diluted to 31250, 6250, 1250, 250, 50, and 10 copies/µL and amplified by SYBER Green qPCR. This test was run in duplicate. Cq values were plotted against

the log of the copy number. Linear trendlines produced R^2 values, and the slopes were used to calculate the PCR efficiencies.

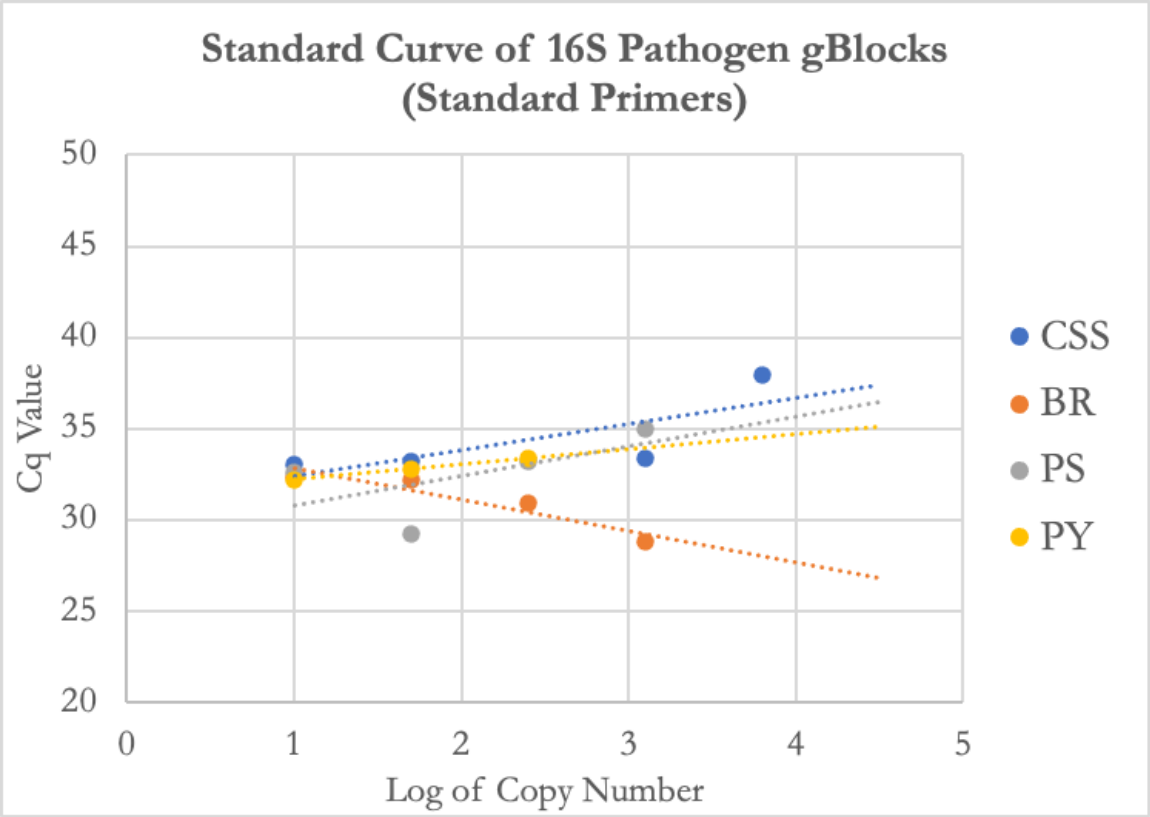

Figure 5. Standard Curve of 16S Pathogens without Nextera tails. gBlocks of pathogen amplicons were diluted to 31250, 6250, 1250, 250, 50, and 10 copies/µL and amplified by SYBER Green qPCR. Cq values were plotted against the log of the copy number. Linear trendlines produced R^2 values, and the slopes were used to calculate the PCR efficiencies.

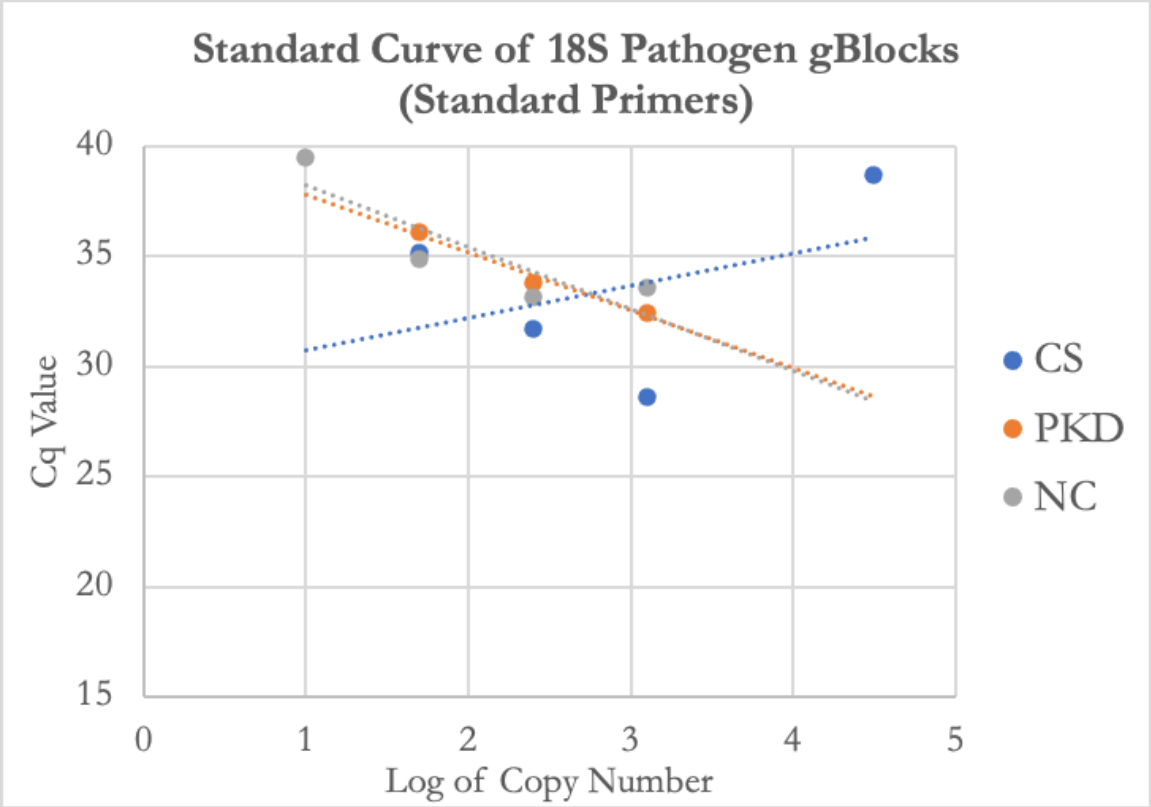

Figure 6. Standard Curve of 18S Pathogens without Nextera tails. gBlocks of pathogen amplicons were diluted to 31250, 6250, 1250, 250, 50, and 10 copies/µL and amplified by SYBER Green qPCR. Cq values were plotted against the log of the copy number. Linear trendlines produced R^2 values, and the slopes were used to calculate the PCR efficiencies.

Species	Primer	R^2	PCR Efficiency
(CSS) *Candidatus Syngnamydia salmonis*	16S with tail	0.6609	0.9223
(CSS) *Candidatus Syngnamydia salmonis*	16S no tail	0.5823	-0.8019
(BR) *Yersinia ruckeri*	16S with tail	0.7136	0.4108
(BR) *Yersinia ruckeri*	16S no tail	0.8821	2.8237
(PS) *Piscirickettsia salmonis*	16S with tail	0.7286	0.277
(PS) *Piscirickettsia salmonis*	16S no tail	0.3623	-0.7618
(PY) *Piscichlamydia salmonis*	16S with tail	0.5103	1.0071
(PY) *Piscichlamydia salmonis*	16S no tail	0.9971	-0.9362
(CS) *Ceratomyxa shasta*	18S with tail	0.918	1.0668
(CS) *Ceratomyxa shasta*	18S no tail	0.1572	-0.7969
(PKD) PKX unclassified myxozoan	18S with tail	0.6754	4.0056
(PKD) PKX unclassified myxozoans	18S no tail	0.9792	1.4213
(NC) *Nanophyetus salmincola*	18S with tail	0.375	1.8534
(NC) *Nanophyetus salmincola*	18S no tail	0.7514	1.2818

Table 6. PCR efficiencies comparing Nextera tail primers to primers without the tails. gBlocks of pathogen amplicons were diluted to 31250, 6250, 1250, 250, 50, and 10 copies/µL and amplified by SYBER Green qPCR. The assay with tails was run in duplicate. Cq values were plotted against the log of the copy number. Linear trendlines produced R^2 values (duplicates were averaged), and the slopes were used to calculate the PCR efficiencies (-1 + 10^(-1/slope)).

DISCUSSION & FUTURE DIRECTIONS

Overall, this study found good potential for eDNA metabarcoding to be useful for salmon parasite screening. From the *in silico* analysis I found that most of the pathogens I was analyzing had no mismatches with primers indicating a high likelihood of amplification. The *in vitro* analysis then showed amplification at low concentrations with acceptable efficiency with the Nextera primers. Lastly, the *in vivo* analysis shows that its possible to search for these pathogens in metabarcoding data.

During our initial research, we sought out software for pathogen sequence analysis. Our testing of Mega 11 revealed that it needed to meet the requirements for our project, specifically lacking degenerate primer functionality and in silico PCR tools. My prior experience with Benchling in the University of Maine eDNA Coordinated Operating Research Entity (CORE) Laboratory for primer design made it the next avenue to pursue for this project. Benchling has proved to be a powerful tool for degenerate primer analysis and web-based nucleic acid assays.

In my analysis in Benchling, I attempted to apply the 16S Parada and 18S Comeau primers to viruses from Shea et al. and the MDMR publication. More pathogens could be assessed in the future. After addressing issues with the parameters that prevented proper alignment, I was then able to apply the primers successfully. The default SantaLucia, 1998 PCR parameters on Benchling needed to be switched to Modified Breslauer 1986 (Phusion, Phire, DyNAzyme). I then selected forward and reverse sites with no mismatches or the least amount. Then I linked the F/R primers, took in silico PCR data, and recorded it in Table 4 for 16S and Table 5 for 18S.

After completing the *in silico* analysis, I predict that pathogen sequences with more than one mismatch will not be detectable with the Maine-eDNA metabarcoding assays. I then tested amplification *in vitro* via quantitative PCR (qPCR) assay with gBlocks, synthetic genes, of target parasites. The pathogen amplicons I ordered were *Yersinia ruckeri* (16S, 0 mismatches), *Candidatus Syngnamydia salmonis* (16S, 2 mismatches: 1F/1R), *Piscirickettsia salmonis* (16S, 1 mismatch: 1R), *Piscichlamydia salmonis* (16S, 2 mismatches: 1F/1R), *Nanophyetus salmincola* (18S, 1 mismatch: 1F), *Ceratomyxa shasta* (18S, 0 mismatches), and PKX unclassified myxozoan (18S, 1 mismatch: 1R). I ordered fewer gBlocks than I initially proposed but reasoned that this set would help test my hypothesis on the robust amplification. This set also included three of the six pathogens from MDMR that passed the *in silico* analysis.

The qPCR results showed inconsistent amplification. This could be explained by the mismatches in the primers and the Nextera tails. Another consideration is that the reference sequences utilized were not all from this region due to availability on GenBank. I used Parada and Comeau primers with Nextera tails as they are used by the Maine-eDNA Project, where future metabarcoding data will likely originate. We saw abnormal curves at high copy numbers but not low ones. I am unsure about this cause and would like to study it further. I hypothesize that the Nextera tails produce a 2D or 3D structure, as our melt curves showed a unique curve in wells with no template control. This is an area for further research to confirm the effects of the tails on amplification. I attempted a PCR test of the same gBlocks using primers without the tails and got worse overall results.

To quantify the qPCR results, I performed standard curves using the Cq values and the log of the gBlock copy number (Figures 3-6). I then added a linear trendline and recorded the slope of the line and the R^2 value of the data points. I used the slope to calculate the efficiencies. Ideal, optimized PCRs have a slope of -3 and an efficiency between 0.9 and 1.1. Looking at Table 6 it is evident that these assays could be optimized. I also saw in the amplification curves that amplification does occur, even with mismatches, for all seven pathogens. With verification that the genes of these pathogens amplify, I expect pathogen sequences with a maximum of 2 mismatches will be detectable with the Maine-eDNA metabarcoding assays.

I used 18S DNA metabarcoding data from index sites (estuarine locations along the Maine coast where samples are routinely collected for Maine-eDNA) to determine whether these pathogens were present. The metabarcoding analysis results will help identify the presence of these pathogens. It is also important to note that these methods cannot quantify pathogen DNA abundance reliably. I do not expect to identify these pathogens in the data from the metabarcoding assays. We were not able to identify any of the pathogens in the metabarcoding data. Continued monitoring using this novel approach will further the goals of protecting the GOM Atlantic salmon DPS to survive in its native habitat. With the ongoing efforts of Maine-eDNA it will only require computational assessment of output data using verified amplicons of pathogen DNA.

Future Work will include further verification of the Comeau and Parada Nextera primers amplifying salmon pathogens. Proving these amplifications will allow the pathogens to be added to metabarcoding screens. qPCR assays could be optimized, including higher quality taq or switching to a probe in place of SYBER green.

Metabarcoding assessment is the largest area of interest as more data will allow for more monitoring, and 16S data will allow me to test for the bacterial pathogens I investigated in this study.

Additional considerations and complications include one of the most influential factors which affect the likelihood of species detection with molecular methods is the concentration of its genetic material in the environment. This concentration is directly linked to the rate of eDNA shedding, which is determined by the organisms' type, size, life stage, and abundance. As such, the concentration of genetic material in the environment can significantly impact the success of molecular species detection (Wood et al., 2020).

Another challenge in pathogen monitoring via eDNA metabarcoding involves bacteria uptaking free DNA in the marine environment. Although eDNA is known to degrade quickly in water, it has been found in streams up to 10 km downstream from its source (Nukazawa et al., 2020). This has raised concerns that environmental bacteria could uptake free DNA, transfer their genetic traits, and amplify, thus becoming a source of the measured eDNA rather than the target aquatic species (Nukazawa et al., 2020). This could have a significant impact on the accuracy of the results and the reliability of the data.

However, eDNA methods meet the legal standards for being accepted as evidence in most court cases, indicating that the reliability of eDNA is not the issue (Sepulveda et al., 2020). Instead, the connection between the results and management needs to be improved, as few tools are available to integrate uncertainty into decision-making. Possible solutions include decision-support trees based on molecular best practices that

incorporate temporal and spatial trends in eDNA positives in relation to salmon risk tolerance (Sepulveda et al., 2020).

REFERENCES

Amarasiri, M., Furukawa, T., Nakajima, F., & Sei, K. (2021). Pathogens and disease vectors/hosts monitoring in aquatic environments: {Potential} of using {eDNA}/{eRNA} based approach. *Science of The Total Environment*, *796*, 148810. https://doi.org/10.1016/j.scitotenv.2021.148810

Bass, D., Christison, K. W., Stentiford, G. D., Cook, L. S. J., & Hartikainen, H. (2023). Environmental DNA/RNA for pathogen and parasite detection, surveillance, and ecology. *Trends in Parasitology*. https://doi.org/10.1016/J.PT.2022.12.010

Comeau, A. M., Li, W. K. W., Tremblay, J. É., Carmack, E. C., & Lovejoy, C. (2011). Arctic Ocean Microbial Community Structure before and after the 2007 Record Sea Ice Minimum. *PLOS ONE*, *6*(11), e27492. https://doi.org/10.1371/JOURNAL.PONE.0027492

FAO. 2018 *The state of world fisheries and aquaculture*. Rome, Italy: Food Agricultural Organization of the United Nations. https://doi.org/I9540en/1/ 07.18

Flye, M. E., Sponarski, C. C., Zydlewski, J. D., & McGreavy, B. (2021). Understanding collaborative governance from a communication network perspective: A case study of the Atlantic Salmon recovery framework. *Environmental Science & Policy*, *115*, 79–90. https://doi.org/10.1016/J.ENVSCI.2020.10.001

Huang, S., Yoshitake, K., Watabe, S., & Asakawa, S. (2022). Environmental DNA study on aquatic ecosystem monitoring and management: Recent advances and prospects. *Journal of Environmental Management*, *323*, 116310. https://doi.org/10.1016/J.JENVMAN.2022.116310

Krkošek M. 2017 Population biology of infectious diseases shared by wild and farmed fish. *Can. J. Fish. Aquat. Sci*. 74, 620–628. https://doi.org/10.1139/ cjfas-2016-0379

Lafferty KD, Harvell CD, Conrad JM, Friedman CS, Kent ML, Kuris AM, Powell EN, Rondeau D, Saksida SM. 2015 Infectious diseases affect marine fisheries and aquaculture economics. *Annu. Rev. Mar. Sci*. 7, 471–496. https://10.1146/annurev-marine-010814015646

Maine Department of Marine Resources. (2021). Chapter 24: Importation of Live Marine Organisms. In Maine Department of Marine Resources (Ed.), *Department of Marine Resources Regulations (13 188)* (pp. 1–28). essay, Maine Department of Marine Resources.

Nowak BF. 2007 Parasitic diseases in marine cage culture- an example of experimental evolution of parasites? *Int. J. Parasitol*. 37, 581–588. https://doi.org/10. 1016/j.ijpara.2007.01.003

Nukazawa, K., Akahoshi, K., & Suzuki, Y. (2020). Are bacteria potential sources of fish environmental {DNA}? *PLOS ONE*, *15*(3), e0230174. https://doi.org/10.1371/journal.pone.0230174

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10.1111/1462-2920.13023

Saunders, R., Hachey, M.A. and Fay, C.W. (2006), Maine's Diadromous Fish Community. Fisheries, 31: 537-547. https://doi.org/10.1577/1548-8446(2006)31[537:MDFC]2.0.CO;2

Sepulveda, A. J., Nelson, N. M., Jerde, C. L., & Luikart, G. (2020). Are {Environmental} {DNA} {Methods} {Ready} for {Aquatic} {Invasive} {Species} {Management}? *Trends in Ecology \& Evolution*, *35*(8), 668–678. https://doi.org/10.1016/j.tree.2020.03.011

Seymour, M. (2019). Rapid progression and future of environmental {DNA} research. *Communications Biology*, *2*(1), 80. https://doi.org/10.1038/s42003-019-0330-9

Shea, D., Bateman, A., Li, S., Tabata, A., Schulze, A., Mordecai, G., Ogston, L., Volpe, J. P., Neil Frazer, L., Connors, B., Miller, K. M., Short, S., & Krkošek, M. (2020). Environmental {DNA} from multiple pathogens is elevated near active {Atlantic} salmon farms. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1937), 20202010. https://doi.org/10.1098/rspb.2020.2010

Stoll, J. S., Leslie, H. M., Britsch, M. L., & Cleaver, C. M. (2019). Evaluating aquaculture as a diversification strategy for Maine's commercial fishing sector in the face of change. *Marine Policy*, *107*, 103583. https://doi.org/10.1016/J.MARPOL.2019.103583

Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J. A., Jansson, J. K., Caporaso, J. G., Fuhrman, J. A., Apprill, A., & Knight, R. (2015). Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. mSystems, 1(1), e00009-15. https://doi.org/10.1128/mSystems.00009-15

Wang, P., Yan, Z., Yang, S., Wang, S., Zheng, X., Fan, J., & Zhang, T. (2019). Environmental {DNA}: {An} {Emerging} {Tool} in {Ecological} {Assessment}. *Bulletin of Environmental Contamination and Toxicology*, *103*(5), 651–656. https://doi.org/10.1007/s00128-019-02720-z

Wood, S. A., Biessy, L., Latchford, J. L., Zaiko, A., von Ammon, U., Audrezet, F., Cristescu, M. E., & Pochon, X. (2020). Release and degradation of environmental {DNA} and {RNA} in a marine system. *Science of The Total Environment*, *704*, 135314. https://doi.org/10.1016/j.scitotenv.2019.135314

AUTHOR BIOGRAPHY

Noah J. Burby was born in Bangor, Maine on April 10, 2001. He was raised in Winterport, Maine and graduated from Hampden Academy in 2019. Majoring in microbiology, Noah has a minor in women's, gender, and sexuality studies. He is a member of Phi Beta Kappa, Beta Beta Beta, Iota Iota Iota, Alpha Lambda Delta, and Golden Key. He has received a Joyce Vorakiarthy Emery and Larry L. Emery Scholarship, George E and Louise Lord '24, '25 Scholarship, and a Dr. Carolyn E. Reed '72 M.D. Pre-Medical Honors College Thesis Fellowship.

Upon graduation, Noah plans to work on a Doctor of Osteopathic Medicine degree.