Using Environmental DNA (eDNA) to Assess the Occurrence of Prey in Deep-Diving Cetaceans off Cape Hatteras, NC

By

Madysen Gilbert

Dr. Andrew Read, Advisor

May 2022

Master's project submitted in partial fulfillment of the requirements for the Master of Environmental Management degree in the Nicholas School of the Environment of Duke University

Executive Summary

The productive waters off Cape Hatteras support high levels of diversity due to the convergence of the cold Labrador Current and the warm Gulf Stream waters. This area is home to many marine predators, including short-finned pilot whales (*Globicephala macrorhynchus*). The biology of many of these predators, including pilot whales, remains poorly understood, however. Recent advancements in technology have provided new opportunities for collecting information including the use of environmental DNA (eDNA) to fill some of these gaps. We collected eDNA from the foraging habitat of pilot whales off Cape Hatteras to determine whether it is possible to identify potential prey organisms using this method and, if so, whether the method allows identification of a comprehensive list of prey species. We first conducted a feasibility test by collecting eDNA from known species in coastal waters to assess whether our method was able to accurately identify these species. We analyzed samples using polymerase chain reaction (PCR) tested with different primer sets and nanopore sequencing to determine DNA sequences. The method allowed identification of organisms to the genus level in the feasibility test. Unfortunately, the samples from Cape Hatteras did not contain meaningful results for the prey of pilot whales. This technique is relatively inexpensive and noninvasive, making this a good choice for future studies of marine mammal ecology. However, I recommend increasing both the sampling frequency and the total number of samples collected during each sampling session to increase the likelihood of finding meaningful patterns. Additionally, a sequence library of potential prey species would help streamline the process and improve species identification. The use of different primer sets also has the potential to improve outcomes of future studies using eDNA. Information gained through such studies will improve our understanding of the trophic ecology of pilot whales. More intensive sampling could identify seasonal and geographical variation in prey availability. The continued enhancement of this method will improve future conservation efforts for pilot whales and other marine mammal predators.

Table of Contents

Introduction

Off the coast of North Carolina, Cape Hatteras provides a suitable habitat for many marine mammals including a number of cetaceans, pinnipeds, and occasionally sirenians. The high amount of biodiversity is a result of the convergence of the Labrador Current (Slope Sea Gyre) and the Gulf Stream in the Cape Hatteras area (Mullin and Fulling, 2003). The Labrador Current brings cold water from the north in the Labrador Sea and the Gulf Stream brings in warm water from the south and the convergence of these two water currents is where the uppermost range for southern species and the lowermost range for northern species collide. During the summer, cetacean species that are expected to be in the area include the Atlantic spotted dolphin (*Stenella frontalis*), bottlenose dolphin (*Tursiops truncatus*), Cuvier's beaked whale (*Ziphius cavirostris*), Risso's dolphin (*Grampus griseus*), long-finned pilot whale (*Globicephala melas*), short-finned pilot whale (*Globicephala macrorhynchus*), and sperm whale (*Physeter macrocephalus*) (Mullin and Fulling, 2003). Of these marine mammals, pilot whales are of particular interest. These deep-diving odontocetes have differing ranges with long-finned pilot whales preferring north, boreal waters and short-finned pilot whales inhabiting tropical to subtropical waters, though the range overlap of these species remains unknown (Isojunno et al., 2017; Thorne et al., 2017; Waring et al., 2015). Both long-finned pilot whales and short-finned pilot whales can be found in the northwest Atlantic and reside in Cape Hatteras in the late summer, though the species are difficult to distinguish when observed (Thorne et al., 2017). While the opportunity for studying marine mammals has been greatly increasing due to advancements in tagging technology, their elusive behavior creates challenges that leave gaps in our knowledge (Sepulveda et al., 2020). Though these challenges can limit our understanding of marine mammal behavior, new methods using environmental DNA (eDNA) can help fill these gaps.

Pilot whales make deep dives that can be over 500 meters deep to forage for calorific squid (Baird et al., 2002; Isojunno et al., 2017; Jensen et al., 2011). These dives have a high energetic cost, but the opportunity to locate high calorie squid at depth allows this strategy to be beneficial and rewarding (Aoki et al., 2017; Goldbogen et al., 2019; Isojunno et al., 2011; Knight, 2017). During the dive, pilot whales will use echolocation clicks to locate potential prey, and once located, will execute a prey capture attempt that is signified by the emission of a series of rapid clicks, referred to as buzzing (Goldbogen et al., 2019; Isojunno et al., 2011; Miller et al., 2004).

Pilot whales mainly feed on squid (including long-finned squid) and other high calorie prey (including Atlantic mackerel) found in the mesopelagic zone of the water column, though there is not a comprehensive prey species list (Aoki et al., 2017; Desportes and Mouritsen, 1993; Gannon et al., 1997). While tagging and recording foraging dives is allowing researchers to gain more information about foraging techniques and preferences from these whales, these foraging adaptations are still not fully understood and the previous reliance of opportunistic stomach content studies done during necropsies creates issues in the known data (Goldbogen and Madsen, 2018; Mintzer et al., 2008; Pyenson, 2011; Tyack et al., 2006). Though useful, this technique is not dependable as the main method of studying pilot whale diets due to extreme limitations and biases, showing only what the whales fed on prior to stranding (Mintzer et al., 2008). Firstly, the individual whales being examined cannot represent the whole population since the contents merely show as the final foraging attempts before the death of the animal and is not representative of either a healthy whale and does not inform on diet changes due to seasonal habitat changes and prey availability (Mintzer et al., 2008; Stephanis et al., 2008). Secondly, stomach content analyses can be bias due to prey digestibility since some prey species digest in less time than others (Gannon et al., 1997; Stephanis et al., 2008). Thirdly, though fish may take longer to digest than squid, squid beaks remain intact, but difficulty in identification of squid through their beaks makes this avenue difficult to pursue (Gannon et al., 1997; Stephanis et al., 2008). To better study pilot whale prey preferences, finding less invasive strategies for identifying target species is necessary.

eDNA is the DNA that is left behind by an organism through the sloughing of cells or excretion of waste material (Closek et al., 2019). These trace amounts of DNA can be analyzed and used to identify species without the need to directly interact with the organism. This approach is less invasive than other marine mammal study practices and does not require researchers to see an animal to collect useful data. Water samples containing eDNA can be collected and analyzed using metabarcoding techniques to determine the species that are present in the sample.

Our goal was to determine if it is possible to identify potential prey animals in samples taken from locations where pilot whale presence is known and identify if the identified method is feasible for future studies. The study that inspired this project focused solely on cephalopod prey, but we wanted to expand this concept and determine if there is a feasible way of identifying

a comprehensive list of target prey species of these deep-diving cetaceans. For this study, we conducted an eDNA analysis of water samples collected at varying depths from a study location off Cape Hatteras, NC and performed a polymerase chain reaction (PCR) with primers that are universal and primers that are specific to the species we expect to see in the area. Using a combination of family specific primer sets and universal primer sets may allow for species to be recognized that may otherwise be missed if only one type of primer was used. Since some species were already expected, but we wanted to identify all that were present, using the combination of primers increased the likelihood of identifying more species in each sample.

Materials and Methods

1. Sample Collection

During July 2021, a research team from the Duke Marine Laboratory travelled to Cape Hatteras on the RV Shearwater to collect information on pilot whales. During this research trip, eDNA samples were collected six times from different depths in the water column, as shown in Table 1. A conductivity, temperature, and depth instrument (CTD) was lowered into the water from the back of the boat and contained a rosette holding Niskin bottles that would collect the water. The Niskin bottles have lids and bottoms that are controlled by the scientists onboard the vessel and are closed one by one at varying depths to capture stratified samples. Six samples were collected on July 16th, 2021, in a location known to have pilot whales with the coordinates: 35.5324, -74.6528 as shown in Figure 1. Water was collected at depth and filtered through a Millipore Sterivex 0.22 μm filter unit using a 50 mL SOFT-JECT Luer Lock. The samples were then preserved on the boat and stored at -80 °C once back at the lab.

Table 1. Cape Hatteras sample information.

Sample	Depth Collected	Water Column Zone	Date Collected
Bottle #1	521 m	Mesopelagic	July 16th, 2021
Bottle #2	380 m	Mesopelagic	July 16th, 2021
Bottle #3	380 m	Mesopelagic	July 16th, 2021
Bottle #4	261 m	Mesopelagic	July 16th, 2021
Bottle #5	260 m	Mesopelagic	July 16th, 2021
Bottle #6	100 m	Epipelagic	July 16th, 2021

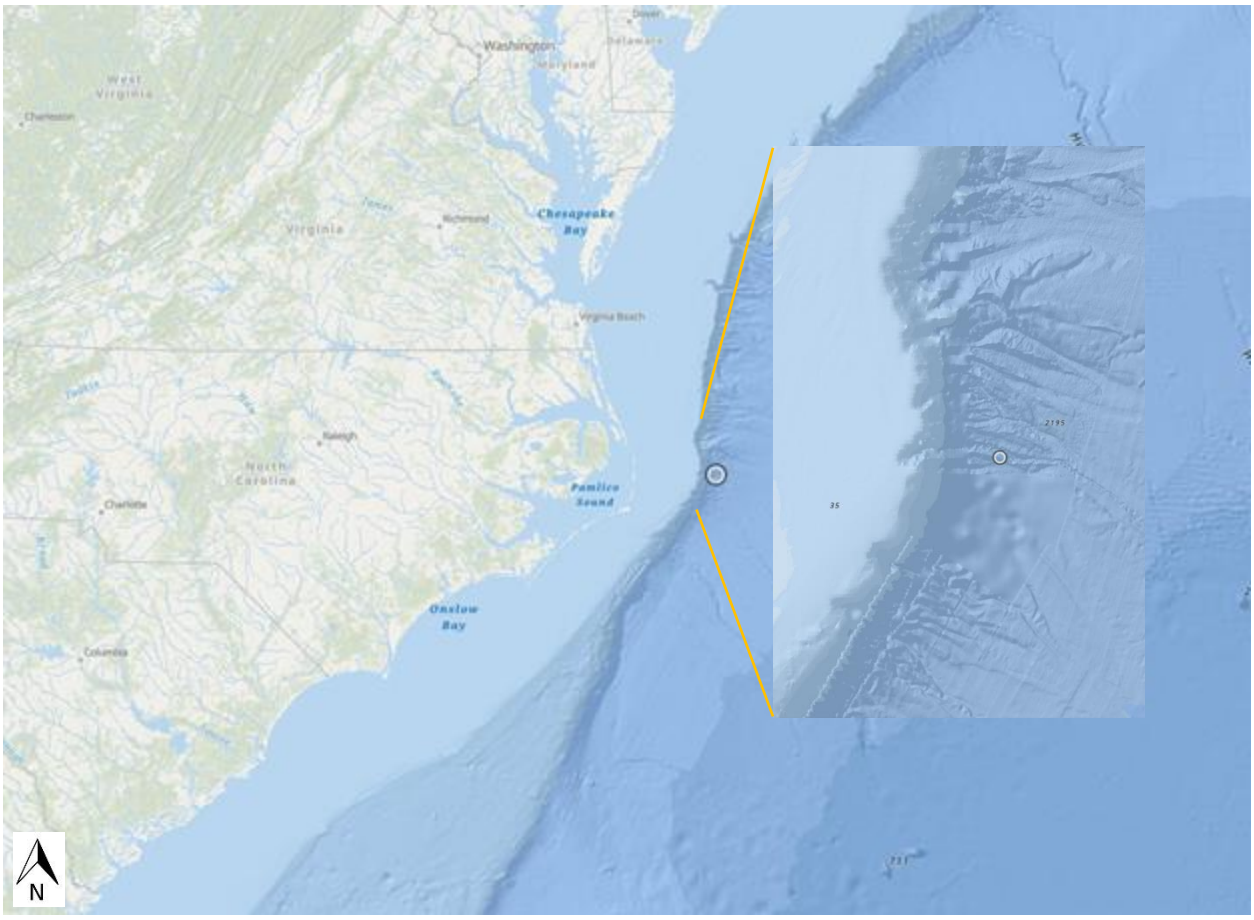


Figure 1. Map showing the study site location off Cape Hatteras, NC, 35.5324 °N and -74.6528 °W.

2. eDNA Analysis Trial Run

2.1 eDNA Collection and Extraction

To create a method for testing the unknown samples from Cape Hatteras, a practice run was constructed. The goal of the practice run was to collect eDNA from known species and process the sample to determine if the method was able to accurately identify the species. The practice run involved collecting water samples from separate tanks containing a different species of marine invertebrate including a blue crab (*Callinectes sapidus)*, oysters (*Crassostrea virginica)*, polychaete worms (*Chaetopterus variopedatus)*, urchins (*Strongylocentrotus purpuratus)*, and shrimps (Palaemonidae sp). Each tank held only one invertebrate species and used running seawater from the Duke Marine Lab seawater system. Each water sample was collected after the subjects were left in separate containers of still seawater for one hour. Samples were then filtered using 0.2 µm mixed cellulose ester Membrane Filters (Whatman) and stored in the

freezer at -20 C. The filters were then divided into fourths using a sterile razor blade and petri dish. One quarter was collected using sterile tweezers and used for the DNA extraction. Each sample was put into an autoclaved microcentrifuge tube and combined with 275 µl of the Digestion Solution Master Mix from the Wizard SV Genomic DNA Purification System and incubated and vortexed overnight (approximately 18 hours) at 55 C to denature the DNA. The samples were then combined with 250 µl of Wizard SV Lysis Buffer and centrifuged to separate the lysate from the precipitate. The lysate was extracted from the microcentrifuge tube and transferred to a wizard SV minicolumn assembly. The samples were then centrifuged and washed with the Column Wash Solution to bind the DNA to the minicolumn assembly. Afterwards, 100 µl of nuclease-free water were added to the minicolumn assembly and then left to incubate at room temperature for two minutes. The tube and assembly constructions were then centrifuged at 8,000 RPM for one minute to filter the nuclease-free water through the assembly and collect the DNA before being stored in the freezer at -20 C until next step.

2.2 Round One PCR

The extracted DNA was then subjected to PCR. After extraction and denaturing, the next step in processing the DNA was to add primers to the template DNA and add Taq polymerase to extend these primers. The universal primers that were chosen for these samples include 18S and LCO HCO. The 18S primer targets the 18S small subunit of the rRNA gene and the LCO HCO primer targets larger regions of the COI gene. For each sample of DNA, 2 µl were taken and placed into a tube in the VWR PCR 8-Well Tube Strips. A solution containing 2 µl 10x PCR Buffer, 1.6 µl $MgCl_2$ (25 mM), 1 µl primer (10 mM), 0.2 µl Taq polymerase, and 11.6 µl nuclease-free water was added per sample (20 µl total in each tube). In addition to the five known samples, one tube was prepared with a negative control of nuclease-free water instead of the 2 µl of DNA. The negative control tests for contamination and to act as quality assurance for the primers (Sepulveda et al., 2020). Bubble cap strips were placed on the tube strips and the samples were placed into the GeneAmp PCR System 9700 (Applied Biosystems) and cycled 35 times under the Standard COI PCR program to anneal the DNA (see supplemental materials for details). Once removed from the thermocycler, the samples were taken to be loaded into the VWR Power Source 300V at 100V for 30 minutes. For each sample, 8 µl were mixed with loading dye and pipetted into the wells of a 2% agarose gel containing SYBR Safe DNA gel

stain (Invitrogen). The remaining sample was stored in the freezer until later use. A PCR marker (Promega) was mixed with loading dye and added into the last well after each sample set based on primer. The Promega 1kb ladder PCR marker provided known base pair lengths to compare the results of the samples for that run. Once complete, the gel was taken to the Ingenius Syngene Bio Imaging machine and an image was taken via GeneSnap under UV light and analyzed.

2.3 Second Round PCR and Nanopore Sequencing

For the next stage, the previous steps were repeated. For each primer used, 20 µl reactions of each sample using the previous mix 2 µl of DNA, 2 µl 10x PCR Buffer, 1.6 µl MgCl2 (25 mM), 1 µl primer (10 mM), 0.2 µl Taq polymerase, and 11.6 µl nuclease-free water were prepared and run through the thermocycler. Once complete, 5 µl were loaded onto an agarose gel and run through the VWR Power Source 300V at 100V for 30 minutes. The gel was then viewed under UV lighting to visualize the amplifications and analyze the quality of the amplicons. The remaining 15 µl from each sample underwent MagBead purification. For each sample, 15 µl of MagBeads were added. MagBeads were collected using a magnet and the supernatant was pipetted out and discarded. Without disturbance of the pellet, the MagBeads were washed a total of two times using 50 µl of fresh 70% EtOH. The EtOH was removed, and the pellet was then allowed to air dry for 2-5 minutes until dry. The pellet was then eluted in 60 µl of nanopure water.

To prepare for the second round PCR, 2 µl of the eluted DNA sample and 1 µl of oligonucleotide (ONT) target primers were added to strip tubes and combined with 2 µl 10x PCR Buffer, 1.6 µl MgCl2 (25 mM), 0.2 µl Taq polymerase, and 11.6 µl nuclease-free water. The samples were then placed in the thermocycler on the 58oC anneal program for 35 cycles (see supplemental materials for details). The ONT barcodes attach to the samples while in the thermocycler and allow for the samples to be pooled and sequenced together.

Once complete, the PCR products were combined into one Eppendorf tube and approximately 400 µl of MagBeads were added, pipetted to mix together, and allowed to incubate at room temperature for five minutes. The tube was then placed on a magnetic plate to allow the beads to

collect and rinsed twice with 85% EtOH and allowed to dry. The beads were then eluted in 70 µl of nanopure water and 65 µl were transferred to a new Eppendorf tube.

The next step includes repairing and dA-tailing the sheared ends of the DNA samples to prepare for the adapter ligation step. A mix of end-prep reagents from the NEBNext Ultra II End repair/dA-tailing kit including 0.5 µl DCS, 24 µl DNA, 2 µl end-prep buffer, and 1.5 µl end-prep enzyme mix were combined in a PCR tube and incubated in the thermocycler at 20 °C for 5 minutes and 65 °C for 5 minutes. The end-prep reaction was then combined with 30 µl of MagBeads and incubated for 5 minutes at room temperature. The MagBeads were then collected using the magnet tray and the supernatant was pipetted out and discarded. The pellet was rinsed twice with 85% EtOH and allowed to dry before being eluted with 35 µl of nanopure water and incubated at room temperature for two minutes. The supernatant was then pipetted out and transferred to a new Eppendorf tube.

In a separate Eppendorf tube, the adapter ligation mix was prepared using 30 µl of DNA, 12.5 µl of Long Fragment Buffer (LNB), 5 µl of Quick t4 Ligase, and 2.5 µl of Adapter Mix (AMX). This reaction was then incubated at room temperature for 10 minutes and then combined with 50 µl of MagBeads and incubated again for 5 minutes. The tube was then placed on a magnet tray to allow a pellet to form, and the supernatant was pipetted off and discarded. The pellet was then washed with 125 µl of LFB, returned to the magnet tray, and the supernatant was pipetted out and discarded twice. The pellet was allowed to dry before being resuspended in 10 µl of Elution Buffer (EB) and incubated at room temperature for 10 minutes. The tube was then placed on the magnet tray and the supernatant was discarded. The beads were then eluted in 7 µl of nanopure water, creating the prepared library.

The final prepared barcoded sequencing library was then prepped for loading onto the Flongle Flow Cell in the MinIon device to sequence the amplicons (Srivathsan et al., 2021). The Flongle Sequencing Expansion Kit was used to create a mix to prime the flow cell using a combination of 117 µl Flush Buffer (FB) and 3 µl of Flush Tether (FLT). To prepare the library, 15 µl Sequencing Buffer (SQB II), 10 µl Loading Beads (LB II), and 5 µl DNA library were combined. The flongle was then primed using the FB and FLT mix and the library was loaded in via the SpotON sample port and allowed to run.

The nanopore sequencing data results, identified using the barcodes, were unzipped, concatenated, and filtered by size. The data were mapped into groups of sequences that were similar for each individual barcode and then further polished before undergoing the Basic Local Alignment Search Tool (BLAST) to find regions of similarity between nucleotide sequences. The BLAST results were then used to analyze the sequences. Originally, the analysis process involved using the software Metagenome Analyzer (MEGAN), but due to the large amount of small sequences that each eDNA sample yielded, this avenue was not sufficient in identifying the most significant hits and was abandoned. Instead, the BLAST results were then run through the python package IsoCon to derive consensus sequences. IsoCon is a tool that is used for reconstructing highly similar sequences found in datasets from long noisy reads (Sahlin et al., 2018; Vierstraete and Braeckman, 2022). These consensus sequences were filtered in excel to count the number of duplicates and how many times those duplicates occurred for each read to determine which sequences were significant hits. Sequences with more than 100 duplicates were then searched in the National Center for Biotechnology Information's (NCBI) genetic sequence database, GenBank, to determine species identification possibilities.

2.4 Cape Hatteras Samples

When transitioning to the samples collected from Cape Hatteras, this first round PCR method was applied for each primer that was tested, but only primers that yielded significant visual results were analyzed with the second round PCR and nanopore sequencing. The Cape Hatteras samples all underwent DNA extraction at the same time using a similar method to the practice run but requiring modifications to incubate the Sterivex filter that was used. Samples were incubated in the Labnet Mini Incubator for 18 hours at 55 °C and then centrifuged into a 50 mL Falcon tube for collection. Once collected, the samples underwent the same process for DNA extraction and denaturing as the practice run samples. The first primers that were chosen were the universal 18S primer and LCO HCO primer that were used in the practice run. These samples were amplified in the thermocycler using the standard COI PCR program and run on a gel with the practice run samples serving as positive controls to ensure that the primers worked. Next, the MiFish primer which amplifies the hypervariable region of the 12 rRNA gene and the 16S primer which amplifies the entire 16S region were used with a flounder eDNA sample serving as the positive control and using the standard COI PCR program (Miya et al., 2015).

Then, the 16S Fish primer, amplifying the mitochondrial 16S gene for marine vertebrates/fish, was tested alongside the 18S primer with the flounder eDNA sample used for the positive control and using the standard COI PCR program. The 18S primer was used in second round PCR and nanopore sequencing, using the protocol outlined for the practice run. The nanopore data were converted into BLAST files which were inputted into IsoCon, analyzed using excel, and checked in GenBank.

To target potential cephalopod species, a combination of three cephalopod specific primers were used including Ceph18S, CephMLS, and S_Cephalopoda. The Ceph18S primer targets the nuclear 18S rRNA gene and both the CephMLS primer and S_Cephalopoda primer target the mitochondrial 16S rRNA gene (Jonge et al., 2021). Initially, these primers were tested using the standard touchdown PCR program, but the program was switched to the ones outlined in the paper by Visser et al. which specified different protocols for the 18S and 16S primers.

The nanopore sequencing outputs were used to create BLAST files that identify regions of similarity in sequences which were then refined using the IsoCon python package to condense replicates and identify significant reads in each barcode. The subsequent reads were then searched in NCBI BLAST to identify the possible identities in each sample.

Results

1. eDNA Practice Run Results

For the practice run samples, the primers 18S and LCO HCO were tested with results shown in Table 2. Since these samples only contained marine invertebrates, they could not be used to test the marine vertebrate specific primers including 16S Fish or MiFish and did not represent cephalopod species so were excluded from the cephalopod primers as well. Table 3 shows that the primers both worked on the samples, but Figures 2a and 2b show that the visualization from the agarose gel revealed that the 18S primer yielded stronger bands than the LCO HCO primer and was favored for the analysis portion of the project. The practice run was fully analyzed using the 18S primer in first round PCR, second round PCR, and nanopore sequencing. The BLAST results showed a myriad of nucleotide sequences found in each of the samples and was used in MEGAN (Metagenome Analyzer) to identify and visualize present species. MEGAN

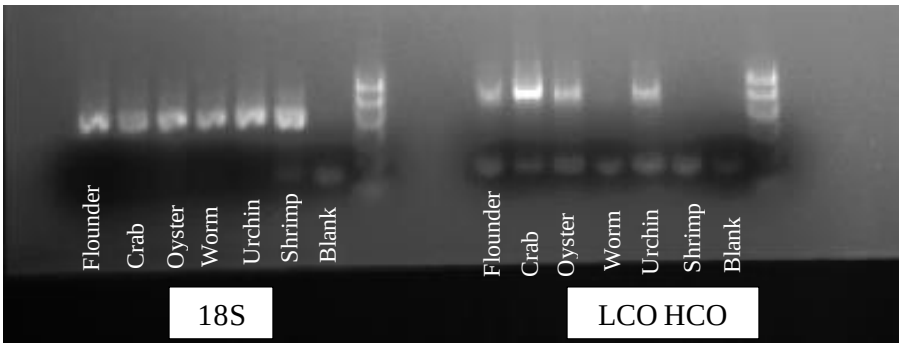
showed the results in the form of a phylogenetic tree and could be modified to represent different levels of identification.

Table 2. Primer characteristics.

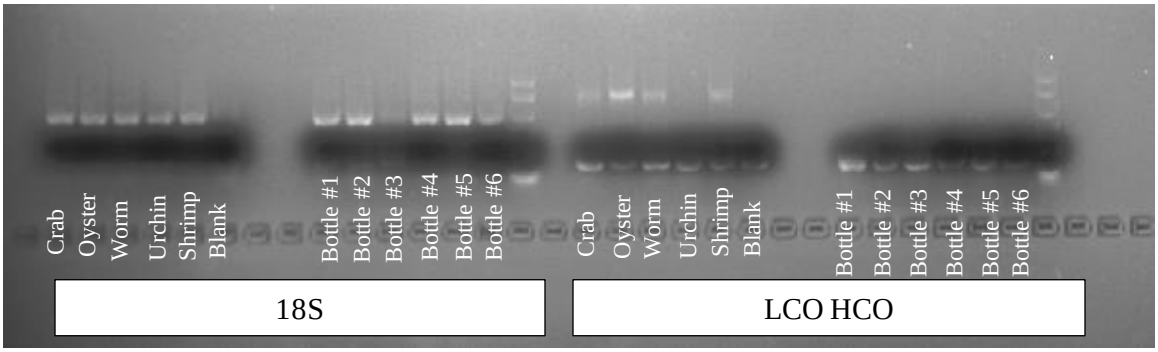
Primer	Primer Type	Forward	Reverse	Base Pairs	Target Region
18S	Universal	5'- TTT CTG TTG GTG CCA TAT GCT TGT CT -3'	5'- ACT TGC CTG TCG CTG CTG CCT TCC TT -3'	336-423 bp	V1-V3 hypervariable regions of 18S rRNA gene
LCO HCO	Universal	5'- TTT CTG TTG GTG CAT AAA GAT ATT GG -3'	5'- ACT TGC CTG TCG CGA CCA AAA AAT CA -3'	658 bp	COI region
16S	Universal	5'- GCA GTC GAA CAT GGC MGC CGC GGT AA -3'	5'- TGG ATC ACT TGT GNV GGG TWT CTA AT -3'	1500 bp	Entire 16S rRNA gene
16SFish	Family Specific	5'- TTT CTG TTG GTG CAG ACC CTD TGG AG -3'	5'- ACT TGC CTG TCG CCG NTG TTA TCC CT -3'	200 bp	Mitochondrial 16S rRNA gene; marine vertebrates/fish
MiFish	Family Specific	5'- TTT CTG TTG GTG CAA CTC GTG CCA GC -3'	5'- ACT TGC CTG TCG CTA ATC CCA GTT TG -3'	163-180 bp	Hypervariable region of 12 rRNA gene
Ceph18S	Family Specific	5'- CGC GGC GCT ACA TAT TAG AC -3'	5'- GCA CTT AAC CGA CCG TCG AC -3'	140-190 bp	Nuclear 18S rRNA gene
CephMLS	Family Specific	5'- TGC GGT ATT WTA ACT GTA CT -3'	5'- TTA TTC CTT RAT CAC CC -3'	212-244 bp	Mitochondrial 16S rRNA gene
S_Cephalopoda	Family Specific	5'- GCT RGA ATG AAT GGT TTG AC -3'	5'- TCA WTA GGG TCT TCT CGT CC -3'	70-73 bp	Mitochondrial 16S rRNA gene

Table 3. Visual results from the agarose gel for each primer tested on each sample.

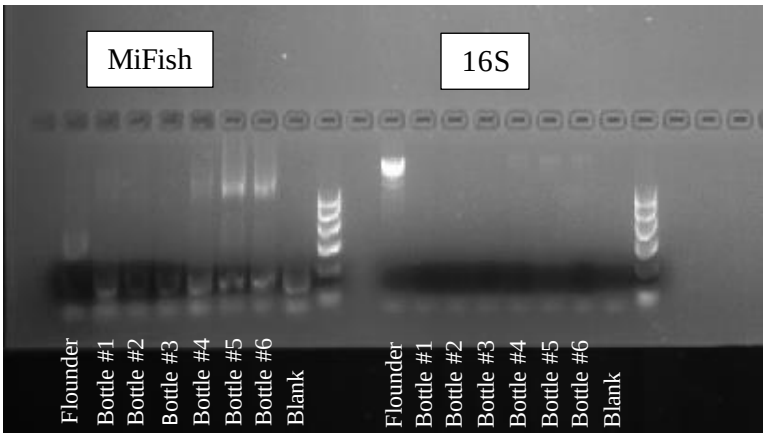
Primer	PCR Program	Crab	Oyster	Chaetopterus	Urchin	Shrimp	Bottle 1	Bottle 2	Bottle 3	Bottle 4	Bottle 5	Bottle 6	Flounder	NC Squid	CA Squid	Blank
18S	Standard COI	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-	-	No
LCO HCO	Standard COI	Yes	Yes	Yes	No	Yes	No	No	No	No	No	No	-	-	-	No
16S	Standard COI	-	-	-	-	-	No	No	No	Yes	Yes	Yes	Yes	-	-	No
16SFish	Standard COI	-	-	-	-	-	No	No	No	No	No	No	No	-	-	No
MiFish	Standard COI	-	-	-	-	-	No	No	No	Yes	Yes	Yes	No	-	-	No
Ceph18S	Touchdown	-	-	-	-	-	No	No	No	No	No	No	No	-	-	No
CephMLS	Touchdown	-	-	-	-	-	No	No	No	No	No	No	No	-	-	No
S_Cephalopoda	Touchdown	-	-	-	-	-	No	No	No	No	No	No	No	-	-	No
Ceph18S	Ceph 18S	-	-	-	-	-	No	No	No	No	No	No	-	Yes	Yes	No
CephMLS	Ceph 16S	-	-	-	-	-	No	No	No	No	No	No	-	Yes	Yes	No
S_Cephalopoda	Ceph 16S	-	-	-	-	-	No	No	No	No	No	No	-	Yes	Yes	No



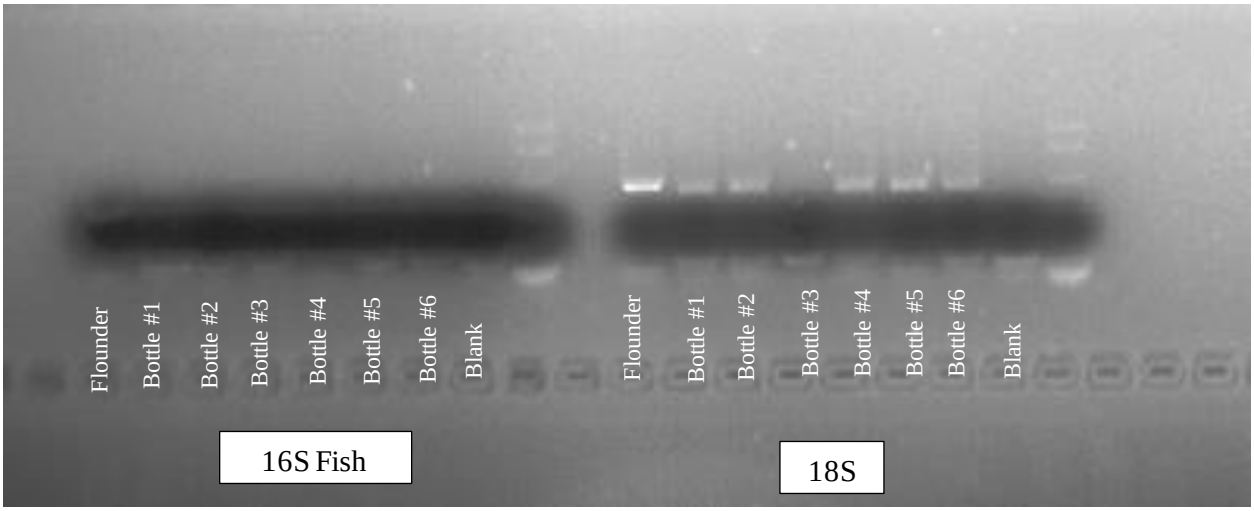
A. Gel showing the results of the 18S primer (left) and LCO HCO primer (right) on the flounder positive control sample and five known invertebrate samples.



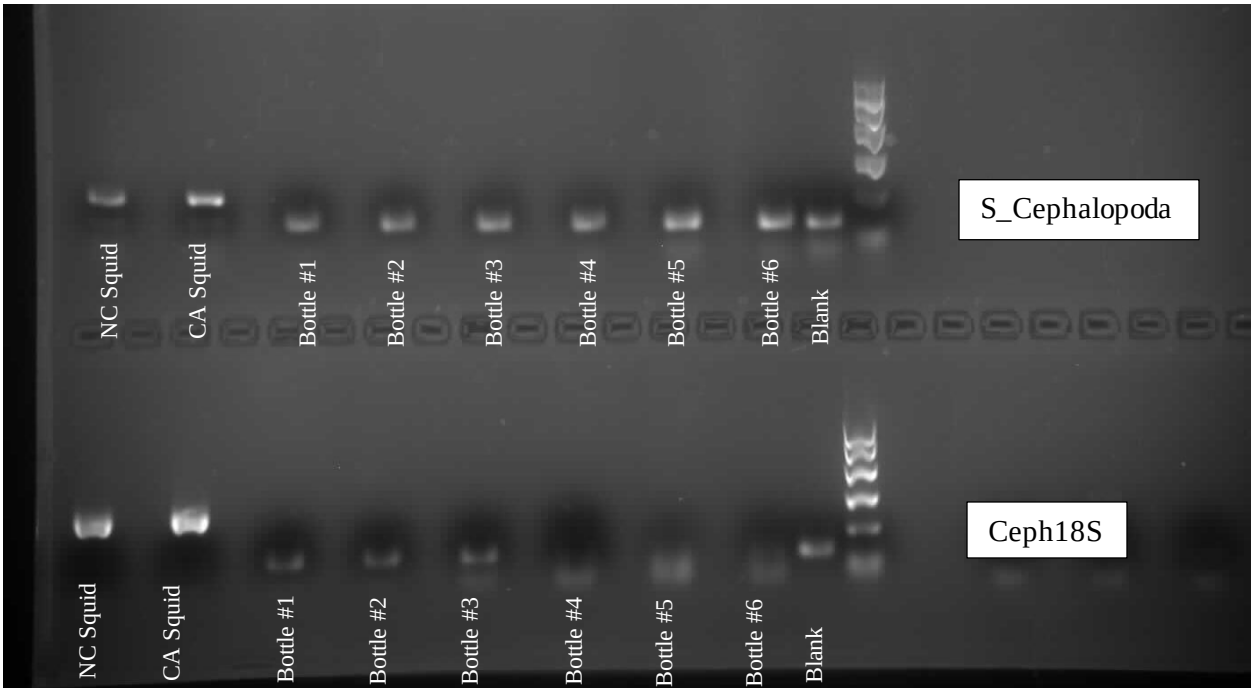
B. Gel showing the results of the 18S primer (left side) and the LCO HCO primer (right side) for both the five known invertebrate samples and the unknown Cape Hatteras samples.



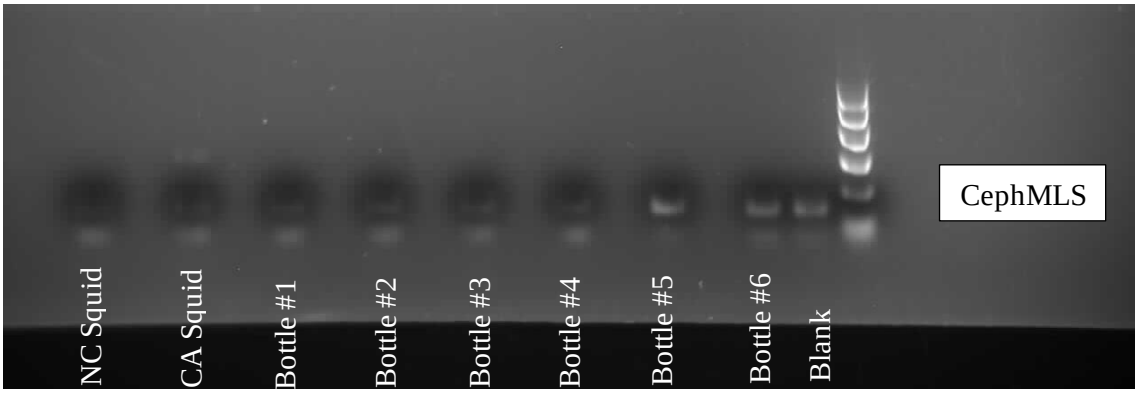
C. Gel showing the results for the MiFish primer (left) and the 16S primer (right) for the flounder positive control sample and the unknown Cape Hatteras samples.



D. Gel showing the results for the 16S Fish primer (left) and the 18S primer (right) for the flounder positive control sample and the unknown Cape Hatteras samples.



E. Gel showing the results for the S_Cephalopoda primer (top) and the Ceph18S primer (bottom) for the two squid positive control samples and the unknown Cape Hatteras samples.



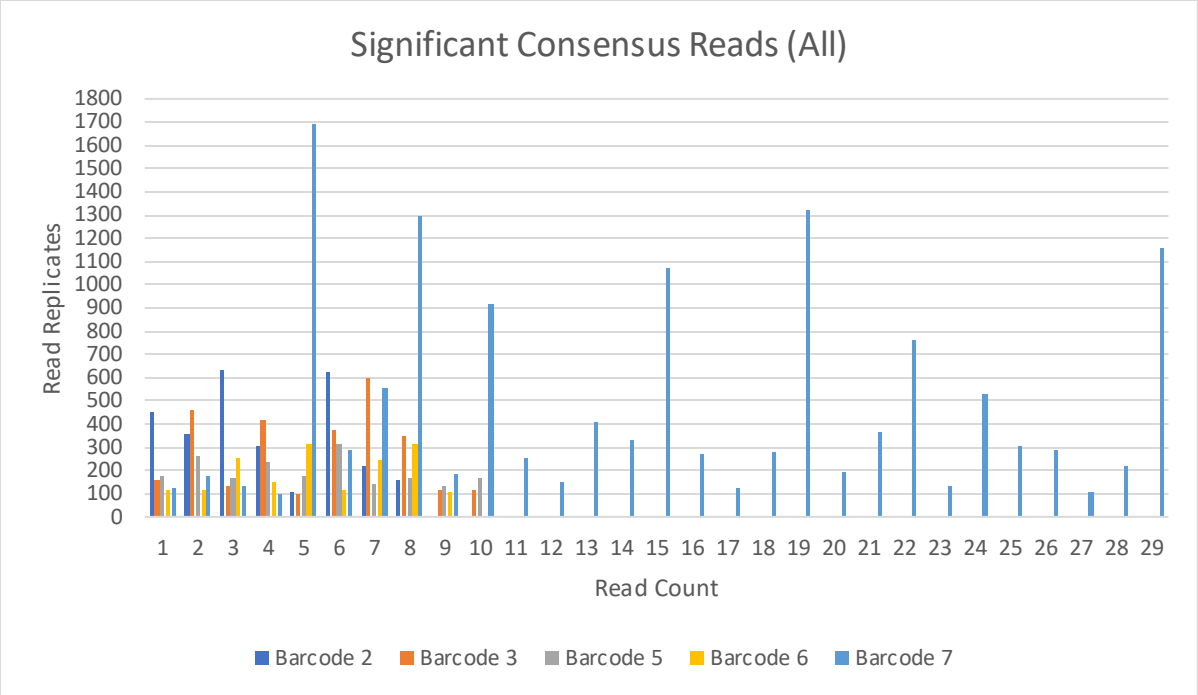
F. Gel showing the results for the CephMLS primer for the two squid positive control samples and the unknown Cape Hatteras samples.

Figure 2. Agarose gels visualizing round one PCR results showing the quality of the amplicons for each primer tested.

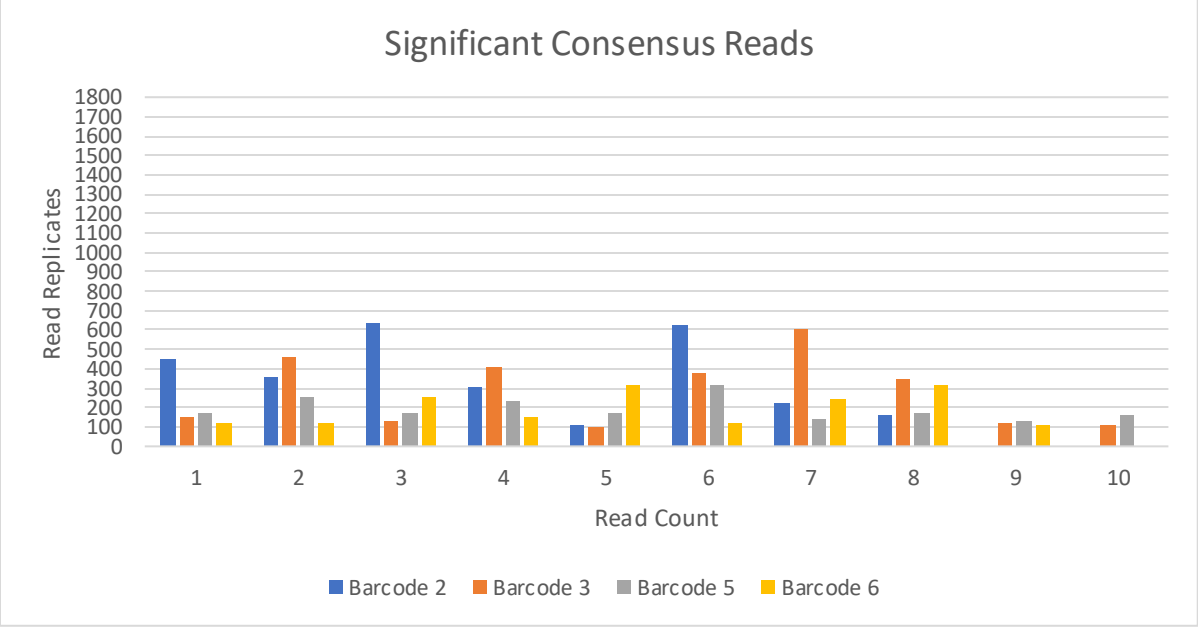
2. Cape Hatteras Samples Results

For the Cape Hatteras samples, the primers 18S, LCO HCO, MiFish, 16S, 16S Fish, S_Cephalopoda, Ceph18S, and CephMLS were used, as shown in Figures 2b, 2c, 2d, 2e, and 2f. Table 2 shows the characteristics of each primer that was used on these samples, indicating that 18S, LCO HCO, and 16S were universal primers, and 16S Fish, MiFish, and the cephalopod primers were family specific primers. Figures 2b and 2c show that of these primers, only 18S, 16S and MiFish were successfully visualized on the gels. However, MiFish did not work for Bottles 1, 2, 3, and 6, so this primer was dropped. The gel for 16S had very streaky bands and incorporated too many base pairs, so it was also dropped. The LCO HCO primer did not show any bands for the Cape Hatteras samples, but successfully showed the practice run marine invertebrate samples, so it was dropped for this sample set. Ultimately, only the 18S primer was chosen for the full run. The full protocol included conducting first round PCR, second round PCR, and nanopore sequencing and examining the results in an analysis using BLAST, IsoCon python package, and GenBank.

The nanopore sequencing results included FASTA files that contained the sequence data. These sequence data were uploaded to the IsoCon consensus algorithm python package to identify the consensus sequences that appear the most for each barcode. The resulting file was sorted and filtered in excel to determine the identity of the sequences that appeared the most with the highest percent identity match and base pair length. Once sorted, the sequences that had a replicate value of 100 (Figure 3) or more were recorded and searched in BLAST. Table 4 shows that barcodes 2, 3, 5, and 6 had similar results for the significant reads after the excel filtering while barcode 4 did not have any significant reads and barcode 7 had 29 results at almost three times as much as the other barcodes. Out of these results, there were 9 unique results across the barcodes with the most abundant result being uncultured eukaryote clone. Figure 4 shows the final results for the organisms that were identified in each barcode and the percentage that each makes up.



A. Graph showing the significant consensus reads for each barcode that yielded results. Barcode 7 had the most results (29) compared to the other barcodes.



B. Graph showing the significant consensus reads for barcodes 2, 3, 5, and 6.

Figure 3. Graphs showing the number of significant consensus sequences that resulted from the excel analysis. Barcode 7 had the most, barcode 4 had no significant sequences, and the remaining barcodes had similar results.

Table 4. Number of significant DNA sequence reads for each barcode and the number of unique organisms from those significant reads.

Barcode	Barcode 2	Barcode 3	Barcode 4	Barcode 5	Barcode 6	Barcode 7
Significant IsoCon Reads	8	10	0	10	9	29
Unique BLAST Identities	4	2	0	1	3	5

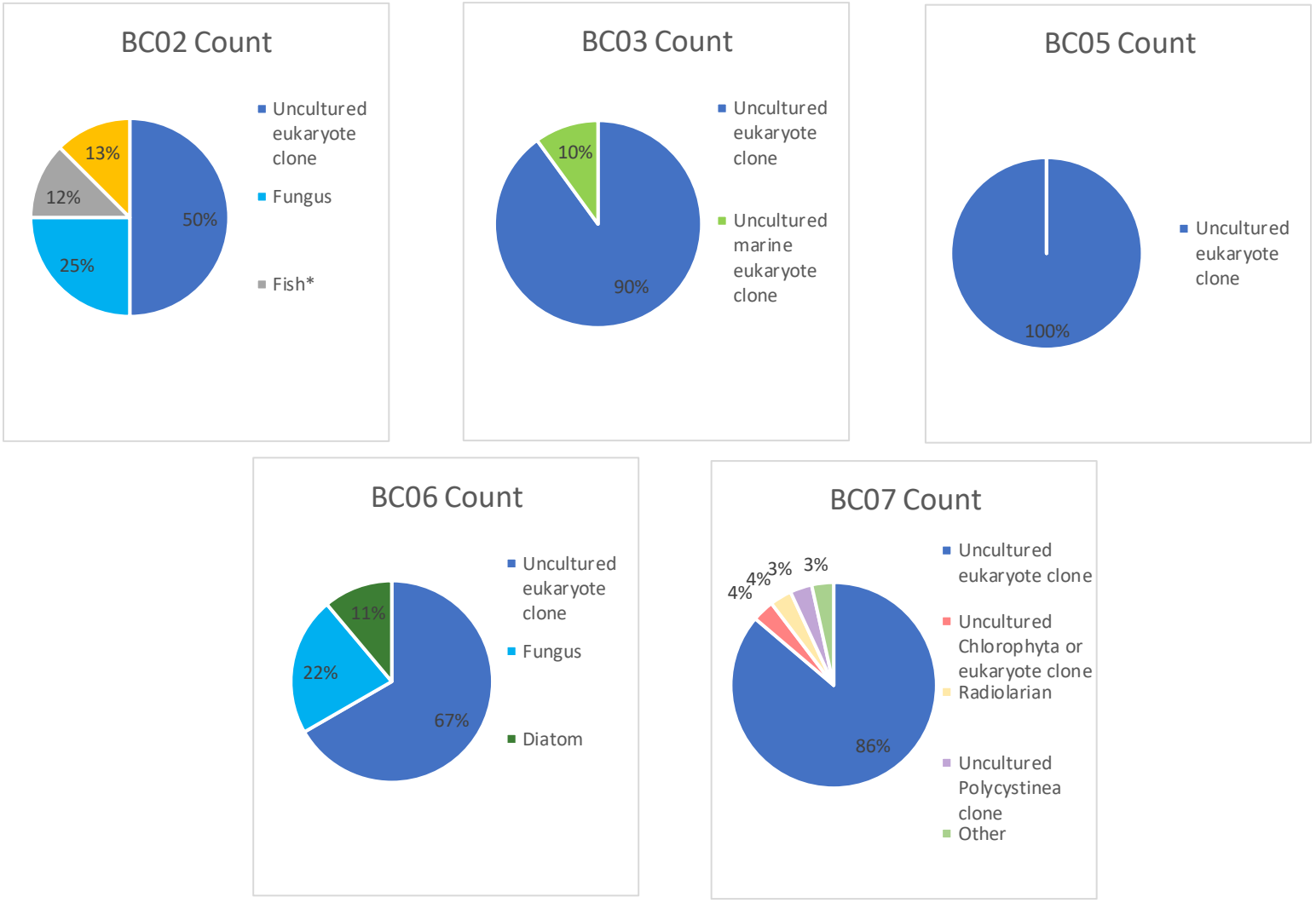


Figure 4. Graphs showing the percentage of identified organisms from each barcode. The majority for each barcode was uncultured eukaryote clone. Barcode 7 had the most diversity in results. *Refers to contamination.

Discussion

The results show that the universal 18S primer yielded the most results during each stage of the PCR process. This primer only targets a small region of 336 to 423 base pairs on the 18S rRNA gene, a smaller chain of base pairs than other universal primers, causing the results to have a lower specificity when identifying species than other primers. The 18S primer showed the brightest bands for the most samples when visualized on the gel. The indicated that this primer was able to identify the most reads out of all of the primers used. Since no other primer was able to show strong bands during the first round PCR, only the 18S primer underwent the entire PCR process. While it was expected that squid species would be found in the eDNA samples, the use of the various cephalopod primers on the samples and two positive squid controls showed that cephalopods were completely absent from these samples. Additionally, the lack of fish in the samples was unanticipated, and, though this is unexpected, the sampling was only conducted in one location on one day, so with repeated sampling, the results may become more comprehensive and better represent the species that are found in the area. The most abundant identified organism for all barcodes was uncultured eukaryote clone, and while these play a critical role in the ocean, they do not hold significant value for this study (Worden et al., 2011).

Identifying the primers that worked for the samples was simple using the gel visualization, however, there was difficulty navigating the best method for analyzing the sequencing results. At first, after creating the BLAST files, the results were uploaded directly into MEGAN, but the outputs were not trustworthy. While MEGAN did produce phylogenetic trees at different identification levels, the software would use sequences that had low amounts of identified base pairs recognized, a low match percentage to known sequences, or both. When comparing the number of reads that we deemed significant to the number of reads that MEGAN used, we found that MEGAN tended to include more, and without being able to see or modify the inputs that were used, we decided to find an alternative method for identifying the sequences. While this method failed, we discovered that we could use the IsoCon python package in conjunction with excel to filter and sort the most significant reads for each barcode. This gave a level of certainty in the significance of the reads that were analyzed and increased the confidence in the results.

The results after applying this new method to the Cape Hatteras samples indicated that it was possible to identify nucleotide sequences found in the samples, but they could only be

confidently identified to a genus level and not a species level since the sequences did not have a 100% match to those in the library. This could be due to the relatively short sequence that the universal 18S primer targets and the large amount of overlap that species within the same family may have for that region of the 18 rRNA gene. Though there were no cephalopod species present, the 18S primer was able to identify the organisms present using the significant consensus sequences as shown in table 5. While the sequences were able to be identified, the samples did not contain the target fish or squid, but this barrier can be resolved with continued sampling and creating a more comprehensive genome of the Cape Hatteras species. Potentially, instead of relying solely on GenBank and BLAST, a new collection of sequences belonging to species found in Cape Hatteras could be created to better compare the contents of samples in the future. The creation of this specific bank would help researchers involved in different projects in the Cape Hatteras area to have a better understanding and way of identifying the organisms they encounter. The establishment of a long-term sampling and identification effort would greatly increase the specificity and accuracy of identifications in future sampling.

Table 5. Organism identification of the significant consensus sequences in each barcode using GenBank. *Refers to contamination.

Identification	BC02	BC03	BC04	BC05	BC06	BC07
Uncultured eukaryote clone	4	9	0	10	6	25
Fungus	2	0	0	0	2	0
Diatom	0	0	0	0	1	0
Fish*	1	0	0	0	0	0
Radiolarian	0	0	0	0	0	1
Uncultured Chlorophyta or eukaryote clone	0	0	0	0	0	1
Uncultured fungus clone	1	0	0	0	0	0
Uncultured marine eukaryote clone	0	1	0	0	0	0
Uncultured Polycystinea clone	0	0	0	0	0	1
Other	0	0	0	0	0	1
Total	8	10	0	10	9	29

This novel technique is feasible with some adjustments to the methods and sampling frequency. Sampling eDNA is a simple technique involving water collection from a desired location that can

be adapted for freshwater and saltwater in both shallow water and at depth. In addition to being simple, it is inexpensive to collect and analyze and is not time intensive when the protocol is known. This new method is less invasive than previous sampling and data collection techniques since it does not require any organism to physically be present during the collection. This decreases perturbation to the animals and allows researchers to sample in a known habitat for their target species or in any location to determine the biodiversity that is present. The value of using eDNA to assess the occurrence of prey in deep-diving cetaceans is high due to this efficient method taking away many of the constraints that have exacerbated the elusiveness of these animals thus far. It is plausible that this method will allow for the identification of comprehensive prey species lists with long-term repeated sampling and more targeted approaches of where samples are collected and at what depth. There is high potential for eDNA to serve as a useful tool in many different genetic studies due to the efficiency in time, cost, and effort along with the realistic approaches of the entire procedure.

Conclusion

This project examined the ability of eDNA to be used for identifying the occurrence of prey in pilot whales off Cape Hatteras. The samples collected were able to show evidence of eDNA in the waters where pilot whales reside, however, the small number of samples resulted in few significant results and the analysis process was unable to identify the organisms present down to a species level. While having the identification down to the genus level is encouraging, it would be more meaningful to adjust the sampling process and frequency in order to create a more comprehensive and specific list for each study using this technique.

Though the analysis process was unable to identify expected organisms like fish or squid in the samples, if eDNA of different organisms is present in different studies, the method is viable. With some adjustments to the sampling procedure including frequency, amount of samples collected, and the addition of blank or negative controls from the sampling sites, the results could be better interpreted in future studies and lead to more useful findings. Due to the cost-effective and non-invasive nature of this technique, using eDNA to detect prey occurrence in deep-diving cetaceans is plausible and feasible with continued sampling and analysis. The results of positive controls of fish and squid were able to show that the combination of universal and family primers was effective and favorable for detecting different organisms and identifying them down to the

lowest level possible. Using the combination of primers allows for different sections of the DNA to be detected and sequenced, increasing the likelihood of correctly identifying the present organisms down to genus. Finding a way to accurately identify the results down to a species-level would strengthen this technique, though there is high confidence in the current method.

Since certain universal primers are only able to detect small ranges of nucleotide sequences, expanding the types of primers used based on the specific location or target organisms can help cover the range of species and potentially better identify organisms to a species-level due to more specific or significant nucleotide sequences being identified. Using a combination of primers in conjunction with increased sampling frequency and amount of samples collected during each trip can help improve the outcome of future projects that incorporate the use of eDNA. Continuing to adapt and enhance this technique will allow for improved management techniques for pilot whales or other target organisms by better informing on current biodiversity and prey identification along with providing a better understanding of the biology and physiology of the animals that are being studied. There is high potential for eDNA in uncovering new information in the marine environment in a noninvasive and inexpensive manner, and with increased efforts, this tool can be key in helping with marine conservation in the future.

Acknowledgements

Thank you to my advisor, Dr. Andrew Read, for all of the support you gave during this project and the encouragement throughout the process. Also, thank you to Dr. Tom Schultz and Laura Givens for all of their invaluable expertise and helpful advice they gave along the way.

References

Aoki, Kagari, Katsufumi Sato, Saana Isojunno, Tomoko Narazaki, and Patrick J. O. Miller. "High Diving Metabolic Rate Indicated by High-Speed Transit to Depth in Negatively Buoyant Long-Finned Pilot Whales." *Journal of Experimental Biology* 220, no. 20 (October 15, 2017): 3802–11. https://doi.org/10.1242/jeb.158287.

Baird, Rw, Jf Borsani, Mb Hanson, and Pl Tyack. "Diving and Night-Time Behavior of Long-Finned Pilot Whales in the Ligurian Sea." *Marine Ecology Progress Series* 237 (2002): 301–5. https://doi.org/10.3354/meps237301.

Closek, Collin J., Jarrod A. Santora, Hilary A. Starks, Isaac D. Schroeder, Elizabeth A. Andruszkiewicz, Keith M. Sakuma, Steven J. Bograd, Elliott L. Hazen, John C. Field, and Alexandria B. Boehm. "Marine Vertebrate Biodiversity and Distribution Within the Central California Current Using Environmental DNA (EDNA) Metabarcoding and Ecosystem Surveys." *Frontiers in Marine Science* 6 (2019). https://www.frontiersin.org/article/10.3389/fmars.2019.00732.

F. Visser*, V. J. Merten, T. Bayer, M. G. Oudejans, D. S. W. de Jonge, O. Puebla, T. B. H. Reusch, J. Fuss, H. J. T. Hoving. "Supplementary Materials for Deep-Sea Predator Niche Segregation Revealed by Combined Cetacean Biologging and EDNA Analysis of Cephalopod Prey," n.d.

Gannon, Damon P., Andrew J. Read, James E. Craddock, Kurt M. Fristrup, and John R. Nicolas. "Feeding Ecology of Long-Finned Pilot Whales Globicephala Melas in the Western North Atlantic." *Marine Ecology Progress Series* 148, no. 1/3 (1997): 1–10.

Goldbogen, J. A., D. E. Cade, D. M. Wisniewska, J. Potvin, P. S. Segre, M. S. Savoca, E. L. Hazen, et al. "Why Whales Are Big but Not Bigger: Physiological Drivers and Ecological Limits in the Age of Ocean Giants." *Science*, December 13, 2019. https://doi.org/10.1126/science.aax9044.

Goldbogen, J. A., and P. T. Madsen. "The Evolution of Foraging Capacity and Gigantism in Cetaceans." *Journal of Experimental Biology* 221, no. 11 (June 12, 2018): jeb166033. https://doi.org/10.1242/jeb.166033.

Isojunno, S., D. Sadykova, S. DeRuiter, C. Curé, F. Visser, L. Thomas, P. J. O. Miller, and C. M. Harris. "Individual, Ecological, and Anthropogenic Influences on Activity Budgets of Long-Finned Pilot Whales." *Ecosphere* 8, no. 12 (2017): e02044. https://doi.org/10.1002/ecs2.2044.

Jensen, Frants H., Mark Johnson, Michael Ladegaard, Danuta M. Wisniewska, and Peter T. Madsen. "Narrow Acoustic Field of View Drives Frequency Scaling in Toothed Whale Biosonar."

Current Biology 28, no. 23 (December 3, 2018): 3878-3885.e3. https://doi.org/10.1016/j.cub.2018.10.037.

Jensen, Frants H., Jacobo Marrero Perez, Mark Johnson, Natacha Aguilar Soto, and Peter T. Madsen. "Calling under Pressure: Short-Finned Pilot Whales Make Social Calls during Deep Foraging Dives." *Proceedings of the Royal Society B: Biological Sciences* 278, no. 1721 (October 22, 2011): 3017–25. https://doi.org/10.1098/rspb.2010.2604.

Jonge, Daniëlle S. W. de, Véronique Merten, Till Bayer, Oscar Puebla, Thorsten B. H. Reusch, and Henk-Jan T. Hoving. "A Novel Metabarcoding Primer Pair for Environmental DNA Analysis of Cephalopoda (Mollusca) Targeting the Nuclear 18S RRNA Region." *Royal Society Open Science* 8, no. 2 (February 2021): rsos.201388, 201388. https://doi.org/10.1098/rsos.201388.

Knight, Kathryn. "Long-Finned Pilot Whales Opt for High-Cost Dives." *Journal of Experimental Biology* 220, no. 20 (October 15, 2017): 3607–3607. https://doi.org/10.1242/jeb.171231.

Miller, P. J. O., M. P. Johnson, and P. L. Tyack. "Sperm Whale Behaviour Indicates the Use of Echolocation Click Buzzes 'Creaks' in Prey Capture." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271, no. 1554 (November 7, 2004): 2239–47. https://doi.org/10.1098/rspb.2004.2863.

Mintzer, Vanessa, Damon Gannon, Nélio Barros, and Andrew Read. "Stomach Contents of Mass-stranded Short-finned Pilot Whales (Globicephala Macrorhynchus) from North Carolina." *Marine Mammal Science* 24 (April 1, 2008): 290–302. https://doi.org/10.1111/j.1748-7692.2008.00189.x.

Miya, M., Y. Sato, T. Fukunaga, T. Sado, J. Y. Poulsen, K. Sato, T. Minamoto, et al. "MiFish, a Set of Universal PCR Primers for Metabarcoding Environmental DNA from Fishes: Detection of More than 230 Subtropical Marine Species." *Royal Society Open Science* 2, no. 7 (July 22, 2015): 150088. https://doi.org/10.1098/rsos.150088.

Mullin, Keith D., and Gregory L. Fulling. "Abundance of Cetatceans in the Southern U.S. North Atlantic Ocean during Summer 1998." *Http://Aquaticcommons.Org/Id/Eprint/15150*, 2003. https://aquadocs.org/handle/1834/31003.

Pyenson, Nicholas D. "The High Fidelity of the Cetacean Stranding Record: Insights into Measuring Diversity by Integrating Taphonomy and Macroecology." *Proceedings of the Royal Society B: Biological Sciences* 278, no. 1724 (December 7, 2011): 3608–16. https://doi.org/10.1098/rspb.2011.0441.

Sahlin, Kristoffer, Marta Tomaszkiewicz, Kateryna D. Makova, and Paul Medvedev. "Deciphering Highly Similar Multigene Family Transcripts from Iso-Seq Data with IsoCon." *Nature Communications* 9 (November 2, 2018): 4601. https://doi.org/10.1038/s41467-018-06910-x.

Sepulveda, Adam J., Patrick R. Hutchins, Meghan Forstchen, Madeline N. Mckeefry, and Anna M. Swigris. "The Elephant in the Lab (and Field): Contamination in Aquatic Environmental DNA Studies." *Frontiers in Ecology and Evolution* 8 (2020). https://www.frontiersin.org/article/10.3389/fevo.2020.609973.

Srivathsan, Amrita, Leshon Lee, Kazutaka Katoh, Emily Hartop, Sujatha Narayanan Kutty, Johnathan Wong, Darren Yeo, and Rudolf Meier. "ONTbarcoder and MinION Barcodes Aid Biodiversity Discovery and Identification by Everyone, for Everyone." *BMC Biology* 19, no. 1 (September 29, 2021): 217. https://doi.org/10.1186/s12915-021-01141-x.

Thorne, Lh, Hj Foley, Rw Baird, Dl Webster, Zt Swaim, and Aj Read. "Movement and Foraging Behavior of Short-Finned Pilot Whales in the Mid-Atlantic Bight: Importance of Bathymetric Features and Implications for Management." *Marine Ecology Progress Series* 584 (December 7, 2017): 245–57. https://doi.org/10.3354/meps12371.

Tyack, Peter L., Mark Johnson, Natacha Aguilar Soto, Albert Sturlese, and Peter T. Madsen. "Extreme Diving of Beaked Whales." *Journal of Experimental Biology* 209, no. 21 (November 1, 2006): 4238–53. https://doi.org/10.1242/jeb.02505.

Vierstraete, Andy, and Bart Braeckman. "Vierstraete 2022 - Amplicon Sorter A Tool for Reference-free Amplicon Sorting Based on Sequence." *Ecology and Evolution* 12 (March 2, 2022). https://doi.org/10.1002/ece3.8603.

Visser, F., V. J. Merten, T. Bayer, M. G. Oudejans, D. S. W. de Jonge, O. Puebla, T. B. H. Reusch, J. Fuss, and H. J. T. Hoving. "Deep-Sea Predator Niche Segregation Revealed by Combined Cetacean Biologging and EDNA Analysis of Cephalopod Prey." *Science Advances*, March 2021. https://doi.org/10.1126/sciadv.abf5908.

Waring, Gordon, Elizabeth Josephson, Katherine Maze-Foley, and Patricia Rosel. *U.S. Atlantic and Gulf of Mexico Marine Mammal Stock Assessments 2009*, 2015.

Worden, Alexandra Z, Christopher Dupont, and Andrew E Allen. "Genomes of Uncultured Eukaryotes: Sorting FACS from Fiction." *Genome Biology* 12, no. 6 (2011): 117. https://doi.org/10.1186/gb-2011-12-6-117.

Supplemental Materials

Table 6. PCR programs for thermocycler.

Program	Initial Denaturation Step	35 Cycles of Denaturing			Final Elongation Step	Hold
Standard COI Program	94 ℃ for 3 minutes	94 ℃ for 15 seconds	48 ℃ for 15 seconds	72 ℃ for 30 seconds	72 ℃ for 5 min	4 ℃
Ceph18S Program	95 ℃ for 5 minutes	98 ℃ for 20 seconds	62 ℃ for 15 seconds	72 ℃ for 1 minute	72 ℃ for 10 min	4 ℃
Ceph16S Program	95 ℃ for 5 minutes	98 ℃ for 20 seconds	55 ℃ for 15 seconds	72 ℃ for 1 minute	72 ℃ for 10 min	4 ℃