

## ABSTRACT

Title of Dissertation: DIET AND STOMACH MICROBIOTA OF GULF MENHADEN, A KEY FORAGE FILTER FEEDING FISH SPECIES

Ammar Hanif, Doctor of Philosophy, 2020

Dissertation directed by: Professor, Rosemary Jagus, and University of Maryland Center for Environmental Science

Menhaden represent a family of important filter feeding forage fish that serves as a trophic link between plankton and piscivorous predators in the marine environment. Dietary analysis is difficult because diet items are small and >80 % of the stomach content is amorphous material. DNA metabarcoding combines mass-amplification of short DNA sequences (barcodes) with high-throughput sequencing. This application allows the simultaneous identification of many taxa within the same environmental sample, as well as the analysis of many samples simultaneously, providing a comprehensive assessment of diet items and gut microbiota. Here we present a methodological approach using DNA metabarcoding suitable for a small filter feeding fish to identify the stomach contents of juvenile Gulf menhaden (*Brevoortia patronus*), collected within Apalachicola Bay, Florida. I describe the optimization of DNA extraction, comparison of two primers and sequencing protocols, estimation of menhaden DNA contamination, quality filtering of sequences, post-sequence processing and taxonomic identification of recovered sequences. I characterized the prokaryotic community using 16S universal ribosomal RNA (rRNA) gene sequencing primers in the V3-V4 hypervariable regions. Using two different sequencing protocols employing different “universal” 16S rRNA gene sequencing primers. Although no difference in overall operational taxonomic units (OTUs) was found, the two sequencing protocols gave differences in the relative abundancies of several bacterial classes. The dominant OTUs resulting from 16S rRNA gene sequencing

at the phylum level were assigned to Proteobacteria, Acidobacteria, Actinobacteria and Chloroflexi and included oil eating bacteria consistent with the Gulf of Mexico location. Stomach microbiota and diet were compared in juvenile Gulf menhaden, *Brevoortia patronus*, caught at two locations, Two Mile Channel and St. Vincent Sound, in Apalachicola Bay, FL in May and July of 2013. The stomach microbiota of samples from both locations showed a predominance of Proteobacteria, Chloroflexi, Bacteroidetes, Acidobacteria and Actinobacteria, although significant differences in composition at the class level were seen. The stomach microbiota from fish from Two-Mile Channel showed a higher level of taxonomic richness and there was a strong association between the microbiota and sampling location, correlating with differences in salinity. Approximately 1050 diet items were identified, although significant differences in the species represented were found in samples from the two locations. Members of the Stramenopile/ Alveolate/Rhizaria (SAR) clade accounted for 66 % representation in samples from Two Mile Channel, dominated by the diatoms *Cyclotella* and *Skeletonema*, as well as the ciliate *Oligotrichia*. In contrast, Metazoa (zooplankton) dominated in samples from St. Vincent Sound, accounting for over 80 % of the reads. These are mainly *Acartia* copepods. Since ciliates are considered to be microzooplankton, this means there is just over 60 % representation of phytoplankton in samples from Two Mile Channel and over 90 % representation of zooplankton in samples from St. Vincent Sound. Overall, I demonstrate the diversity of juvenile menhaden stomach contents that supports a characterization of menhaden as environmental samplers.

**DIET AND STOMACH MICROBIOTA OF GULF MENHADEN, A KEY  
FORAGE FILTER FEEDING FISH SPECIES**

by

Ammar W. Hanif

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
**Doctor of Philosophy**  
2020

Advisory Committee:

Professor Rosemary Jagus, Chair  
Professor Allen Place  
Professor David Secor  
Professor Jeff Shields  
Dr. Kevin Friedland

© Copyright by  
Ammar W. Hanif  
2020

## Preface

Chapter 2 of this thesis will be submitted for publication in *Limnology and Oceanography Methods* and is formatted under the guidelines specified for that journal. Chapter 3 of this thesis will be submitted for publication in *Frontiers in Microbiology* and is formatted under the guidelines specified for that journal.

## Foreword

This project began as a joint project between the NOAA-EPP funded Cooperative Science Centers established to help train under-represented minorities in marine sciences. The partners of the Cooperative Science Centers, the Living Marine Resources Cooperative Science Center, LMRCSC, lead institution, the University of Maryland Eastern Shore, UMES, along with the Environmental Cooperative Science Center, ECSC, lead institution Florida A&M University's, FAMU, embarked on a multi-institutional study of forage fish. The skills and resources of the partner institutions and their students were brought together to work on unanswered questions and gaps in our knowledge of these key species. Atlantic and Gulf menhaden are among the largest US fisheries used primarily for feeding to farmed animals and aquaculture fish. There are concerns that this will cause unintended and unconsidered consequences, affecting not only the populations of predator fish that depend on the prey fish, but more widely on ecosystem integrity and function. My part in this has been to bring new technology to investigating the microbiota and diet of the Gulf menhaden, *Brevoortia patronus*, using DNA metabarcoding.

## Dedication

This work is dedicated to my children as an example of what they can accomplish with patience, perseverance, prayer, and support and to all those that came before me.

## Acknowledgements

I would like to thank those who assisted me in this endeavor. My special regards go out to all my committee members, who have provided excellent support and the skills to move forward. My deepest gratitude is reserved for Dr. Rosemary Jagus, my mentor and advisor. Many thanks and much appreciation to fellow graduate students. This work was supported by the National Oceanic and Atmospheric Administration, Educational Partnership Program award to the Living Marine Resources Cooperative Science Center (LMRCSC) NA11SEC4810002 and NA16SEC4810007. I was supported in part by MD SeaGrant Graduate Student Fellowship, NA10OAR4170072 SA7528129-D and in part as an LMRCSC Graduate Fellow. Thanks to Dr. Stacy Smith, Delaware State University, for coordinating the menhaden collection. Thanks to James White, Resphera Bioscience, for assistance with data analysis, Sabeena Nazar, for input and assistance with sequencing, Ernest Williams, Institute for Marine and Environmental Technology, for assistance on many experiments, Dr. Ryan Woodland, Chesapeake Bay Laboratories, for discussions on stable isotope analysis and Dr. Russell Hill, Executive Director of the Institute for Marine and Environmental Technology who supported and encouraged me during my final push to completion. Thanks also to LMRCSC summer undergraduate interns Alexis Peterson, Marcus Hughes, Nicolas Dawson, Rashawnda Wright, and Malisa Smith and to Centennial High School intern, Eric Sibanda for assistance with this work.



# Table of Contents

Preface.....	ii
Foreword.....	iii
Dedication.....	iv
Acknowledgements .....	v
Table of Contents .....	vi
List of Tables .....	viii
List of Figures .....	ix
List of Abbreviations .....	xi
Chapter 1: Introduction.....	1
1.1 Forage fish .....	1
1.2 Menhaden.....	2
1.3 Gulf menhaden.....	4
1.4 Apalachicola Bay, FL .....	8
1.6 Menhaden diet.....	10
1.7 DNA-based approaches to the identification of diet.....	10
1.8 Use of DNA metabarcoding for the analysis of the stomach content of juvenile Gulf menhaden.....	15
1.9 Research questions.....	16
Chapter 2: Methodology for the identification of stomach contents in the filter feeding fish ( <i>Brevoortia patronus</i> ) using DNA metabarcoding.....	17
2.1 Abstract.....	17
2.2 Introduction.....	18
2.3 Materials and Procedures .....	23
2.3.1 <i>Sample collection</i> .....	24
2.3.2 <i>Optimization of DNA extraction</i> .....	25
2.3.3 <i>Sample preparation and DNA extraction</i> .....	26
2.3.4 <i>High throughput sequencing</i> .....	26
2.3.5 <i>Post-sequencing pipeline</i> .....	28
2.3.6 <i>Analysis of diversity</i> .....	29
2.3.7 <i>Menhaden 18S rRNA gene sequence and estimation of menhaden DNA contamination</i> .....	29
2.4 Assessment of method .....	30
2.4.1 <i>Optimization of DNA extraction and quality assessment</i> .....	30
2.4.2 <i>Assessment of menhaden DNA contamination in stomach DNA samples</i> .....	31
2.4.3 <i>Comparison of results using different sequencing methods</i> .....	35
2.4.4 <i>Variation in taxonomic composition in each sample at the class level using the two different primers/sequencing protocols</i> .....	43
2.4.5 <i>Validation of the presence of taxa found by high throughput sequencing with end-point PCR using group specific primers</i> .....	47
2.5 Discussion.....	49
2.5.1 <i>Strengths of DNA metabarcoding</i> .....	49
2.5.2 <i>Limitations of DNA metabarcoding</i> .....	50
2.5.3 <i>Other considerations</i> .....	51
2.6 Comments and recommendations .....	52
2.7 Acknowledgments.....	53

Chapter 3: The stomach microbiota of juvenile Gulf menhaden ( <i>Brevoortia patronus</i> ) differs with location in Apalachicola Bay, FL.....	55
3.1 Abstract.....	55
3.2 Introduction.....	56
3.3 Methods.....	58
3.3.1 <i>Sample collection</i> .....	58
3.3.2 <i>Sample preparation, DNA extraction, and estimation of DNA quality</i> .....	60
3.3.3 <i>MiSeq library preparation and high throughput sequencing</i> .....	60
3.3.4 <i>Post-sequencing pipeline</i> .....	61
3.3.5 <i>Microbiota data analysis</i> .....	62
3.4 Results.....	62
3.4.1 <i>Characterization of stomach bacteria communities from menhaden caught at Two-Mile Channel and St. Vincent Sound</i> .....	62
3.4.2 <i>Shared and unique OTUs</i> .....	67
3.4.3 <i>Alpha and beta diversity analysis</i> .....	74
3.5 Discussion.....	77
3.6 Acknowledgements.....	82
Chapter 4: Diet of juvenile Gulf Menhaden ( <i>Brevoortia patronus</i> ) using DNA metabarcoding.....	83
4.1 Abstract.....	83
4.2 Introduction.....	84
4.3 Methods.....	87
4.3.1 <i>Sample collection</i> .....	87
4.3.2 <i>Sample preparation and DNA extraction</i> .....	89
4.3.3 <i>High throughput sequencing</i> .....	89
4.3.4 <i>Post sequencing pipeline</i> .....	90
4.3.5 <i>Data analysis</i> .....	90
4.3.6 <i>Stable isotope analysis</i> .....	91
4.4 Results.....	91
4.4.1 <i>Characterization of stomach 18S rRNA gene sequences from menhaden caught at Two-Mile Channel and St. Vincent Sound</i> .....	91
4.4.2 <i>Alpha and beta diversity analyses show that stomach samples from TMC are more diverse than those from SVS</i> .....	99
4.4.3 <i>Stable isotope analysis</i> .....	102
4.5 Discussion.....	104
4.6 Acknowledgements.....	108
Chapter 5: Future steps and final thoughts .....	112
5.1 Limitations .....	112
5.2 How should we convert sequence reads to dietary data?.....	113
5.3 Studying ecosystems with DNA metabarcoding .....	113
5.4 Use of filter feeders as environmental biomonitors in environmental metabarcoding studies.....	114
5.5 Accounting for functional ecological importance/significance of the stomach microbiota .....	116
5.6 The evolution of DNA metabarcoding analysis methods.....	117
Bibliography .....	119

## List of Tables

**Table 2.1:** Primers used for sequencing

**Table 2.2:** DNA recovery and quality using **different DNA extraction kits:**

**Table 2.3:** Effect of different gene sequencing methods on the number of raw sequence and post-processing sequence reads

**Table 2.4:** Mean alpha diversity metrics from 16S rDNA gene sequences

**Table 2.5:** Significant differences in mean 16S rRNA reads at the class level using Fadrosh versus Illumina sequencing protocols

**Table 2.6:** Number of 18S rRNA reads and OTUs per sample

**Table 2.7:** Taxon-specific primers for validation of sequence assignments

**Supplemental Table 2.1:** Characteristics of water at collection sites in Apalachicola Bay

**Table 3.1:** Water quality measurements from each site

**Table 3.2:** Comparison of the number of the number of raw reads and post-processing reads from Two Mile Channel and St Vincent Sound samples

**Table 3.3:** Representation of the most abundant proteobacteria from Two Mile Channel and St Vincent Sound samples

**Table 3.4:** Taxonomic composition of the shared OTUs from all samples

**Table 3.5:** Unique OTUs from Two Mile Channel and St Vincent Sound samples

**Table 4.1:** Water quality measurements from each sample site in Apalachicola Bay

**Table 4.2:** Comparison of the number of the number of raw reads and post-processing reads from Two Mile Channel and St Vincent Sound samples

**Supplemental Table 4.1:** Representation of the most abundant higher taxa from Two Mile Channel and St Vincent Sound samples

**Supplemental Table 4.2:** Taxonomic composition of the SAR clade at the genus level of stomach contents from Two Mile Channel and St Vincent Sound

**Supplemental Table 4.3:** Taxonomic composition of Metazoa at the genus level of stomach contents from Two Mile Channel and St Vincent Sound

**Supplemental Table 4.4:** Length of fish caught at TMC and SVS

## List of Figures

- Figure 1.1:** Conceptual life history model of Gulf menhaden
- Figure 1.2:** Perspective map of Apalachicola Bay, Florida
- Figure 2.1:** Summary of workflow
- Figure 2.2:** Collection sites in Apalachicola Bay
- Figure 2.3:** Percentage of menhaden 18S rRNA gene sequences in stomach DNA samples
- Figure 2.4:** Multiple alignment of Gulf menhaden 18S rRNA gene sequence
- Figure 2.5:** Schematic of the 16S rRNA gene sequencing primers
- Figure 2.6:** Comparison of 16S rRNA gene read counts generated by Fadrosh and Illumina protocols pre- and post-processing
- Figure 2.7:** Comparison of alpha diversity metrics of OTUs at the class level generated by the Fadrosh or Illumina protocols
- Figure 2.8:** Comparison of apparent community beta diversity by unweighted and weighted UniFrac measures
- Figure 2.9:** Comparison of the relative abundance of OTUs at the class level derived using Fadrosh and Illumina protocols for 16S rRNA gene sequencing
- Figure 2.10:** Comparison of the relative abundance of OTUs at the phylum level determined using primers for 18S rRNA gene sequences
- Figure 2.11:** Validation of taxa found by endpoint PCR
- Figure 3.1:** Collection sites in Apalachicola Bay
- Figure 3.2:** Taxonomic composition at the phylum level of menhaden stomach microbiota from Two Mile Channel and St Vincent Sound samples
- Figure 3.3:** Taxonomic composition at the class level of Proteobacteria and Chloroflexi from Two Mile Channel and St Vincent Sound samples
- Figure 3.4:** Variation in taxonomic composition at the class level of stomach microbiota in each sample from Two Mile Channel and St Vincent Sound
- Figure 3.5:** Taxonomic composition of the shared OTUs from all samples
- Figure 3.6:** Comparison of the differences in the microbial community from each sample by alpha diversity metrics
- Figure 3.7:** Comparison of the differences in the microbial community from each sample by beta diversity analysis
- Figure 4.1:** Collection sites in Apalachicola Bay
- Figure 4.2:** Taxonomic composition at the higher taxa level of menhaden stomach contents from Two Mile Channel and St Vincent Sound samples

**Figure 4.3:** Variation in taxonomic composition of stomach contents of the major taxa in each sample from Two Mile Channel and St Vincent Sound

**Figure 4.4:** Variation in taxonomic composition of stomach content OTUs within the SAR clade in each sample from Two Mile Channel and St Vincent Sound

**Figure 4.5:** Variation in taxonomic composition of stomach content OTUs within the Metazoa in each sample from Two Mile Channel and St Vincent Sound

**Figure 4.6:** Comparison of the differences in stomach contents from each sample by alpha diversity metrics

**Figure 4.7:** Comparison of the differences in stomach contents from each sample by beta diversity analysis

**Figure 4.8:** Significant differential abundance in OTUs from Two Mile Channel and St Vincent Sound

**Figure 4.9:** Stable isotope analysis of muscle

## List of Abbreviations

16S rRNA:	16S (Svedberg) ribosomal ribonucleic acid
18S rRNA:	18S (Svedberg) ribosomal ribonucleic acid
ANERR:	Apalachicola National Estuarine Research Reserve
ARISA:	automated ribosomal intergenic spacer analysis
BASLab:	Bioanalytical Service Laboratory
BLAST:	basic local alignment search tool
bp:	base pairs
CASIF:	Central Appalachian Laboratory Stable Isotope Facility
COI:	cytochrome c oxidase subunit I
CTAB:	cetyl trimethyl ammonium bromide
DGGE:	denaturing gradient gel electrophoresis
DSU:	Delaware State University
DNA:	deoxyribonucleic acid
ECSC:	Environmental Cooperative Science Center
eDNA:	environmental deoxyribonucleic acid
FAMU:	Florida Agriculture and Mechanical University
FDR:	false discovery rate
FISH:	fluorescence in situ hybridization
GSI:	gonadal somatic index
GUI:	graphical user interface
HTS:	high-throughput sequencing
ITS:	internal transcribed spacer
KNIME:	Konstanz information miner
LMRCSC:	Living Marine Resources Cooperative Science Center
OTU:	operational taxonomic unit
PAH:	polycyclic aromatic hydrocarbon
PCoA:	principal coordinates analysis
PCR:	polymerase chain reaction
PSU:	practical salinity units
QIIME:	quantitative insights into microbial ecology
RNA:	ribonucleic acid
RT-PCR:	reverse transcriptase-polymerase chain reaction
SAR:	Stramenopiles, Alveolates and Rhizaria
SEDAR:	Southeast data, assessment, and review
SDS:	sodium dodecyl sulfate
SL	standard length
SVS:	Saint Vincent Sound
TL:	
TMC:	Two-mile Channel
T-RFLP:	terminal restriction fragment length polymorphism

# Chapter 1: Introduction

## 1.1 *Forage fish*

Forage fish, also called prey fish, are small pelagic fish that are preyed on by larger predators for food. Predators include larger fish, seabirds and marine mammals. Typical ocean forage fish feed near the base of the food chain on plankton, often by filter feeding. Forage fish species play both an ecological role in food webs and an economic role in commercial fisheries (Pikitch *et al*, 2012; Pikitch *et al*, 2014). Forage fish species are exceptionally important to the structure and functioning of marine ecosystems, serving as the main conduit of energy flow from lower to upper trophic levels (Pikitch *et al*, 2012). They can exert middle out control on both plankton and predators.

Marine ecosystems that exhibit this community configuration, featuring many species at the lower and upper trophic levels, but constricted to one, or at most several, dominant planktivorous forage-fish species at the crucial mid-level, have been referred to as “wasp-waist” ecosystems (Bakun *et al* 2006; Bakun *et al*, 2009; Alder *et al*, 2008; Cury *et al*, 2000). Wasp-waist species support a high diversity of larger predators that are highly susceptible to fluctuations in prey biomass (Cury *et al*, 2000). Variations in the abundance of forage fish species will propagate to both higher trophic levels (which may depend on them as a major food item) and to lower trophic levels (on which they may exert very heavy grazing pressures) (Bakun, 2006). Numerous ecosystem models have shown the overall importance of forage fish to the ecosystem (Geers *et al*, 2016; O’Farrell *et al*, 2017; Plaganyi & Essington, 2014). Forage fish are the main diet for many fish, seabirds, and marine mammals (Alder *et al*, 2008; Essington *et al*, 2015). It is estimated that forage fish can make up to 20 % of the diet of marine mammals and 12.5 % of the diet of predatory seabirds (Alder *et al*, 2008; Essington *et al*, 2015). Furthermore, they can be vital to local economies by supplying fish for large volume fisheries that support industrial or reduction fisheries. Such fisheries are found throughout the world’s oceans, except Antarctica. Some of the largest

fisheries occur on the west coast of South America (southeast Pacific), northern Europe, and the United States (east and Gulf coasts and Alaska) where the principal catches are of Peruvian anchovy, capelin, and Atlantic and Gulf menhaden respectively (Alder 2008). Historically herring, sardines and menhaden have been the main forage fish species targeted for the reduction fisheries. In the late 1950s, Peruvian anchovy was added to this list and fishing for this species increased. Pikitch found that while the global catch of forage fish worldwide was valued at \$5.6 billion, the fisheries they supported were valued at more than twice that at \$11.3 billion (Pikitch *et al*, 2014).

## 1.2 Menhaden

Menhaden (*Brevoortia spp*) are a small common pelagic schooling genus of the family Clupeidae occupying the coasts and estuaries of the United States Atlantic and Gulf regions. There are four recognized species of menhaden in North American marine waters, three of which are found in the Gulf of Mexico. Recent taxonomic work, using DNA sequence comparisons, have organized these into large-scaled (Gulf and Atlantic menhaden, *B. patronus* and *B. tyrannus*, respectively) and small-scaled (Finescale and Yellowfin menhaden, *B. gunteri* and *B. smithi*, respectively) (Anderson, 2007). There is higher relatedness within the small-scaled and large-scaled species than amongst other members of the genus. Of these, only the large-scaled, *B. patronus* and *B. tyrannus*, support an established reduction fishery where fish are reduced to fish oil and fishmeal. Atlantic menhaden range along the Atlantic coast from Nova Scotia to southeastern Florida. Gulf menhaden dominate the menhaden fishery in the Gulf of Mexico, with the other two Gulf species menhaden species representing less than 1 % of the annual catch (Ahrenholz, 1981). Gulf menhaden range from Veracruz, Mexico to southwestern Florida. Though Finescale and Gulf menhaden stay within the Gulf of Mexico, the Yellowfin menhaden overlap the ranges of the other menhaden and can be found from the Mississippi River Delta to Virginia. Finescale menhaden range just east of the Mississippi River Delta to Campeche, Mexico.



Menhaden form large schools, which can be found migrating throughout estuaries and near-shore regions. Menhaden depend on these environments for spawning and nursery grounds. Although Atlantic menhaden have been considered to be the most migratory of the four species (Ahrenholz, 1991), recent reassessments of tagging data show their seasonal movements to be more regional than coastal (Liljestrand *et al*, 2019). The adult population is generally distributed from Florida to Maine. During May-June, an estimated 86 % of Atlantic menhaden from North and South Carolina move northwards. They remain largely within the same coastal region from June to October. In winter, approximately 20 % of tagged fish north of the Chesapeake Bay move southward to the Chesapeake Bay and North and South Carolina. However, most appear to over-winter in the northern part of their range (Liljestrand *et al*, 2019). This is consistent with high Atlantic menhaden larval abundance in near-shore waters during the winter in regions north of the Maryland-Delaware line and the Chesapeake Bay region (Simpson *et al*, 2016). The other menhaden species do not exhibit extensive migration patterns.

More is known of the spawning behavior of Atlantic menhaden compared to the other species. Analysis of Northeast Fisheries Science Center ichthyoplankton surveys (1977-1987 and 2000-2013) shows that Atlantic menhaden spawning occurs primarily near shore over a large spatial range, from southern New England to North Carolina (Simpson *et al*, 2017) with hotspots in the Mid Atlantic Bight between the Chesapeake Bay and the Delaware Bay and near Long Island in New York. Spawning activity takes place throughout the year and population range, but peaks during November and December.

Neither Yellowfin nor Finescale menhaden show any evidence of systemic seasonal migration but rather remain nearshore or in the estuaries throughout the year (Gunter, 1945; Dahlberg, 1970; Ahrenholz, 1991). Based on the collection of ripening adults and times when eggs are found in the water column, the spawning activity of both Yellowfin and Finescale menhaden is presumed to be from November to March where they filter-feed extensively on phytoplankton

and zooplankton. As larvae, menhaden are attack feeders that subsequently undergo an ontogenic change in feeding behavior to filter feeding. Due to their collective filtering capacity, it has been postulated that menhaden can exert a significant grazing pressure on common phytoplankton blooms that occur near and in estuaries (Deegan, 1993; Lynch *et al*, 2010).

### 1.3 Gulf menhaden

Gulf menhaden (*Brevoortia patronus*) is the main menhaden species in the Gulf of Mexico, ranging from the northern Gulf from Brazos Santiago, Texas, to Tampa Bay, Florida (Christmas & Gunter, 1960). This species distributes along the U.S Gulf coast nearshore waters during late spring and summer then utilize deeper waters offshore beginning in October and for the winter (Ahrenholz, 1991). Though Gulf menhaden do not stratify with age, as seen in Atlantic menhaden, there are data to indicate a tendency for Gulf menhaden from the extreme eastern and western ranges to move toward the center of their range with age (Ahrenholz, 1981). The species is an important forage fish species along the Gulf coast providing an important food source for fish, seabirds and marine mammals (Vaughan *et al*, 2007). Many of the commercially and recreationally harvested fish species along the Gulf coast rely on the abundant schools of menhaden, including king mackerel (*Scomberomorus cavalla*), Spanish mackerel (*Scomberomorus maculates*), dorado (*Coryphaena hippurus*), crevalle jack (*Caranx hippos*), tarpon (*Megalops atlanticus*), red drum and bonito (*Sarda sarda*) (Dailey *et al*, 2008; Franklin, 2007; Sagarese *et al*, 2016). Among other species, the diet of the blacktip shark (*Carcharhinus limbatus*) and the brown pelican (*Pelecanus occidentalis*), Louisiana's state bird, can consist of over 95 % menhaden (Franklin, 2011).

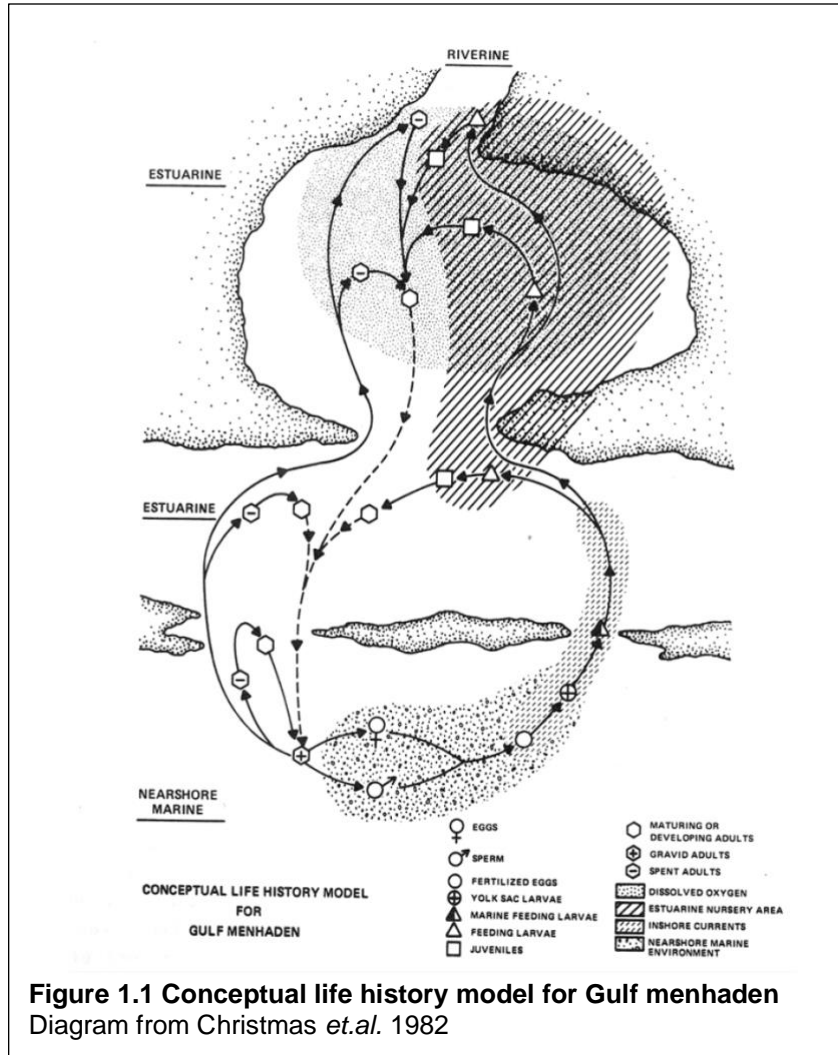
Gulf menhaden support a large directed reduction fishery and, along with shrimp, support the largest fisheries by landings and by revenue in the Gulf of Mexico (Vaughan *et al*, 2007; O'Farrell *et al*, 2017). Gulf menhaden are reduced to fish oil and meal used in livestock feed, aquaculture feed and omega-3 fatty acid-rich menhaden oil, is used in pharma- and nutraceuticals, cosmetics and

other consumer products (Olsen *et al*, 2014; SEDAR, 2013; Menhaden Advisory Commission, 2015). Assessment of the fishery in 2018 concluded that Gulf menhaden are not overfished or undergoing overfishing (SEDAR63, 2018). However, the assessment panel concluded that data and techniques are insufficient at present to incorporate factors that could describe the ecosystem value of Gulf Menhaden adequately. Without adequate assessment of the ecological role of the species, the determination of fishery reference points remains inadequate. In particular, if a wasp-waist species such as Gulf menhaden decreases in abundance, the architecture of energy flows can become highly vulnerable and unreliable (Jordan *et al*, 2005; Robinson *et al*, 2015; Geers *et al*, 2016). This raises the possibility that the potential exists for the large reduction fishery to impose a substantial ecological impact (Sagarese *et al*, 2016). One deficit in understanding the ecological role of Gulf menhaden is the dearth of specific dietary data for menhaden. Most of the menhaden dietary studies have been on the allopatric species, the Atlantic menhaden, *Brevoortia tyrannus*, which does not overlap with *B. patronus* spatially; hence emphasizing the need for a more complete understanding of the diet of Gulf menhaden.

The valid scientific name for Gulf menhaden is *Brevoortia patronus* (Goode). The life history of Gulf menhaden has been described by several authors (Hoode & Fore, 1973; Ahrenholz 1991). In general, Gulf menhaden life history is typical of the cycle followed by most estuarine-dependent species in the Gulf of Mexico (**Figure 1.1**). Gulf menhaden have been found spawning from near-shore to sixty miles offshore along the entire U.S. Gulf coast from October through early March (Suttkus, 1956; Turner, 1969; Combs, 1969; Fore, 1970; Christmas & Waller, 1975; Lassuy, 1983; Shaw *et al*, 1985). Peak spawning periods fluctuate from year-to-year, probably in response to varying environmental conditions (Suttkus 1956). Using samples archived at the NMFS Beaufort Laboratory from 1944-2014, as well as fresh samples collected from Mississippi and Louisiana waters from 2014-2016, gonadal somatic index (GSI) values were found to increase in early October for both males and females, and reach peak values for females by the second half of October (Brown-Peterson *et al*, 2017). Female GSI remained

elevated but gradually decreased from late October through March, whereas male GSI remained elevated from early October through March without a gradual decline. Mean male and female GSI values suggest a spawning season of 5.5 months, extending from early October through the end of March (Brown-Peterson *et al*, 2017). Work looking at the histology of female fish concluded that Gulf menhaden are asynchronous batch spawners, potentially spawning every 4-7 days. The Brown-Peterson study provided an estimated total annual fecundity of 10-20 times higher than that estimated in earlier studies (Lewis & Roithmayr, 1981) and the value used in the last benchmark assessment of menhaden stock (SEDAR 13).

Spawning occurs offshore, and the larvae move into estuarine nursery areas where they spend the early part of their lives (**Figure 1.1**) (Christmas *et al*, 1982; Reid, 1955). Egg hatch and early growth of planktonic larvae occur when currents from offshore spawning grounds transport them to low-salinity estuary nursery grounds (Minello & Webb, 1997). The use of estuaries as nursery habitat is a common theme in the life history of marine fishes because the protected environment and abundant food provide an ideal location (Able, 2005; Potter *et al*, 2013). Gulf menhaden are unusual in that the juveniles depend on estuaries and can be considered to be marine estuarine opportunists (Potter *et al*, 2013). Planktonic larvae make their way into estuaries. At hatching, larvae are poorly developed with undeveloped mouths and fin rays as well as nonfunctional, unpigmented eyes (Reintjes 1962; Houde & Fore 1973). The metamorphosis of Gulf menhaden larvae to juveniles occurs between 28-35 mm SL (Deegan 1986) and at a reported age range of 88-103 days (Deegan & Thompson 1987). Juvenile growth and development occur primarily in estuaries (Robinson *et al*, 2015). The duration of this stage, and the ultimate size reached, varies based on estuarine conditions and the absolute age of individual fish (relative to when they were spawned during the season) (Lassuy 1983; Ahrenholz 1991). At the time of hatching, larval Gulf menhaden are from 2.8-3.1 mm standard length (SL). First feeding occurs at 2.9-5.7 days at 4.3 mm (SL) Powell (1993). As larvae transform



into juveniles, body depth and weight increase substantially with only a minimal increase in length (Ahrenholz 1991). Significant changes in internal morphology occur; the maxillary and dentary teeth become nonfunctional and disappear. Gill rakers increase in length, number, and complexity, and pharyngeal

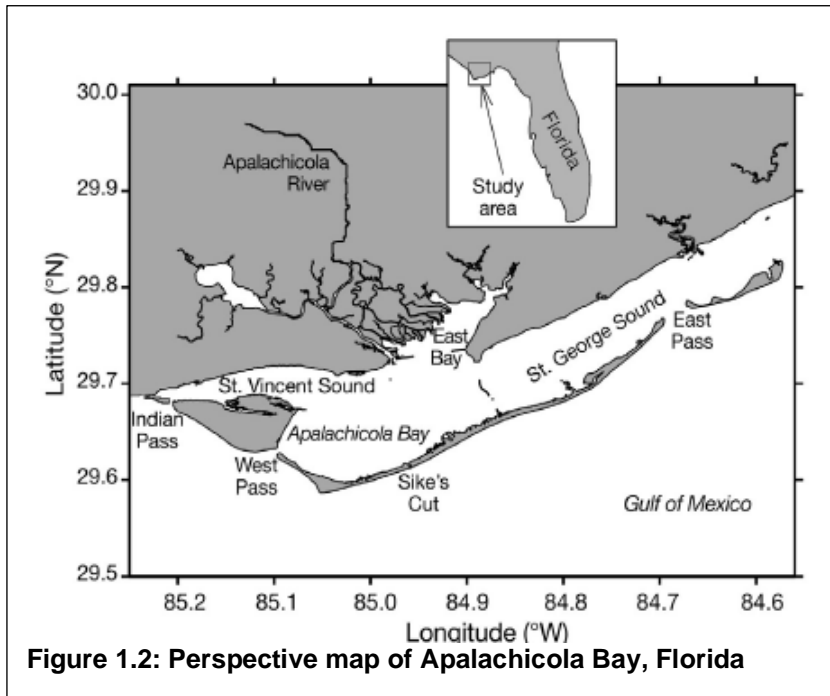
pockets appear. The alimentary tract folds forward, and a muscular stomach (gizzard) and many pyloric caeca develop while the intestine forms several coils (June & Carlson 1971).

Young-of-the-year Gulf menhaden are ubiquitous members of the northern Gulf of Mexico estuarine nekton communities and occupy fresh to brackish waters. After transformation from the larval form, juveniles remain in low salinity, near-shore areas where they travel about in dense schools, often near the surface (Lassuy, 1983). Menhaden are omnivorous filter feeders, feeding by straining phytoplankton and zooplankton from water. As juveniles, Gulf menhaden live in tidal creeks, marsh and open bay areas where they filter the water column via their gill rakers and where they remain until late summer or fall

(Deegan, 1990). The migration pattern of juvenile Gulf menhaden involves the sequential use of marsh creek and open bay areas, coinciding with the productivity peaks in those areas (Deegan, 1990). By occupying tidal creeks early in the year, they can take advantage of high primary productivity stimulated by the influx of nutrients with high spring river flow, the flushing of detritus off the marsh surface from the river mouth, as well as temperatures warmer than the open bay area (Deegan, 1993). The combination of warm water and high productivity in tidal creeks in the spring provides an environment that promotes rapid growth. When food availability in the tidal creeks begins to decline, the fish move to the open bay area where phytoplankton and zooplankton are increasing. By fall, the schools disappear from near-shore waters and are thought to move offshore, wintering on the inner and middle continental shelf (Roithmayr & Waller, 1963). The extent of the offshore range is unknown (SEDAR 63, 2018).

#### 1.4 Apalachicola Bay, FL

My model system for studying Gulf menhaden stomach contents is the Apalachicola Bay estuary. It is a highly productive lagoon and barrier island complex on the upper Gulf coast of Florida (**Figure 1.2**). The high productivity is a result of the Apalachicola River delivering freshwater and nutrients to the bay (Livingston 1984, Mortazavi *et al* 2000a, 2000b, 2001). Nutrient input supports high levels of phytoplankton productivity (Mortazavi *et al*, 2000b) which in turn supports the bay's secondary productivity (Chanton & Lewis 2002). It covers about 212 square miles and serves as the interface between the river system and the Gulf of Mexico. Four barrier islands bound the bay: St. Vincent Island, St. George Island, Little St. George Island, and Dog Island. The bay area, including Apalachicola Bay, East Bay, St. George Sound, St. Vincent Sound, Indian Lagoon, and Alligator Harbor, is about 65 km long and 5 to 10 km wide. Apalachicola Bay is a river-dominated system with the major source of freshwater input coming from the alluvial Apalachicola River. The Apalachicola River is the largest Florida river in terms of flow (NWFWM report 2017).



Maximum river flows occur during late winter to early spring months and are highly correlated with Georgia rainfalls (Meeter *et al*, 1979). The bulk of seawater flow is from the east entering St. George Sound. The western end of Apalachicola

Bay is linked to the Gulf of Mexico by Indian Pass, the narrow channel between St. Vincent Island and the mainland, with a maximum water depth of about 4 m. St. Vincent Sound itself is shallow, with an average depth of little more than 1 m, containing numerous oyster bars. Within the bay's shallow waters, with an average depth of 3 m, are numerous oyster reefs and sandy shoals. The surrounding wooded lowlands consist of saltwater and freshwater marshes, and freshwater swamps. Two-Mile Channel follows the coastline from the west side of the Apalachicola River estuary for approximately 2 miles.

Apalachicola Bay has been designated by NOAA as a National Estuarine Research Reserve (ANERR). There is ongoing work with Florida A&M University's (FAMU) Environmental Sciences Institute as part of the Environmental Cooperative Science Center (ECSC) to develop a conceptual model of Apalachicola Bay to help in management decisions and fill in data gaps about the system. This made it an ideal site in which to pursue my studies on the diet of juvenile Gulf menhaden

### 1.6 Menhaden diet

Assessment of menhaden diet is technically challenging because the visible food items are small (5-100  $\mu\text{m}$ ) and menhaden have a gizzard-like stomach that grinds ingested items to an amorphous paste (Friedland *et al*, 1984). As a result, most of the menhaden stomach content is unrecognizable and has been described as amorphous material (Lewis & Peters, 1994) making microscopic techniques extremely difficult for identification. Microscopic examination also has disadvantages that include being labor intensive and needing highly skilled individuals for the morphological identification of semi-digested or fully digested plant and animal fragments (Holechek *et al*, 1982; Ingerson-Mahar, 2002; Moreby, 1988). Friedland *et al* (1984) showed in the allopatric species, Atlantic menhaden (*Brevoortia tyrannus*), that the minimum-sized filtered particle for juvenile menhaden is 7 to 9  $\mu\text{m}$ ; however, maximum filtration efficiency is for particles approximately 100  $\mu\text{m}$ . It has been postulated that the detrital material plays a role in the retention of these smaller particles (Friedland *et al*, 1984). Detritus is primarily structural material of plant origin that also commonly includes bacteria, fungi, microalgae, protozoa and small animals (Deegan *et al*, 1990). However, detrital material will be reflective of the water filtered since constituents of detritus vary by location (VanValkenburg, 1978). In addition, the amount of detrital material in the stomach may differ depending on location of feeding (Lewis 1994). Furthermore, although the role of ingested detritus has largely been ignored (Lewis & Peters, 1994), it has been shown, using physiological and stable isotope evidence that detrital material can be used as a food source in juvenile Gulf menhaden (Deegan *et al*, 1990; Olsen *et al*, 2018).

### 1.7 DNA-based approaches to the identification of diet

In this study, I sought to establish the molecular technique of DNA metabarcoding to provide an unambiguous forensic tool to identify stomach contents of menhaden. DNA metabarcoding has become the method of choice in characterizing living communities in any environment. This approach provides a comprehensive culture-independent approach that has been used to obtain a full



inventory of gut microbiota (Jami *et al*, 2015; Tarnecki *et al*, 2017; Egerton *et al*, 2018) and prey items (Jakubavičiūtė *et al*, 2017; Riccioni *et al*, 2018; Waraniak *et al*, 2019) from multiple fish species. Advances in DNA based metabarcoding have made comprehensive assessment of diet and gut microbiota feasible by combining mass-amplification of short DNA sequences (barcodes) with high-throughput sequencing. Through molecular barcoding methods, organisms in the stomach contents of filter feeders, where most prey items lack diagnostic taxonomic features, can now be assessed (Pompanon *et al*, 2012). Similarly, DNA barcoding and high-throughput sequencing can also be used to evaluate the microbiota of stomach contents (King *et al*, 2012) and for characterizing the biodiversity of microbial communities.

The basic principle of DNA-based methods is to analyze the DNA extracted directly from a sample derived from a site of interest. This could be water, sediment, or gut contents. The earliest DNA-based methods extracted DNA from a microbial community and probed for targeted genes of interest using the technique fluorescent *in situ* hybridization (FISH). This method uses fluorescently labeled, specific oligonucleotides probes as marker genes, which hybridize to the target DNA (Amann *et al*, 1995). An alternative method was to sequence amplicons of specific gene regions that were subsequently cloned into *Escherichia coli* (Ward *et al*, 1990). These methods are only sufficient for low-throughput applications, they do not deliver exhaustive insights into microbial or prey diversity and are expensive and time consuming. Other DNA-based methods such as automated ribosomal internal transcribed spacer analysis (ARISA), terminal restriction fragment length polymorphism (T-RFLP), denaturing gradient gel electrophoresis (DGGE) and microarrays were also pursued. However, the development of high-throughput sequencing technology quickly replaced those methods as the go-to method for DNA-based organism identification in either microbial community or diet analysis by being able to generate millions to billions of short-fragment DNA reads within hours.

High-throughput sequencing for identification uses a phylogenetically conserved marker region as a DNA barcode to identify organisms. A suitable barcode sequence needs to have conserved sequences at the 5'- and 3'-ends with hypervariable region(s) in between. This allows amplification of the bar code region by polymerase chain reaction. Over the last decade, barcode-based approaches have been applied to study microbial communities and diet (Hiergeist *et al*, 2015; Pompanon *et al*, 2012). These studies have provided insights into diet, host health, complex trophic interactions and ecosystem function. Barcode based approaches use the DNA sequence of the amplified product to identify organisms. There are several types of barcodes available, each with their own set of limitations and specific applications. Perhaps the most widely used are the 16S (prokaryote) or 18S (eukaryote) ribosomal RNA gene (rRNA) sequences or the mitochondrial cytochrome oxidase I (COI) sequence (eukaryote). These targets have allowed a much more detailed analysis of gut and fecal contents compared to earlier microscopic examinations. Barcoding methods allow the identification of organisms in the stomach content of filter feeders where prey items lack diagnostic taxonomic features (Pompanon *et al*, 2012). Early barcoding approaches to diet analysis and microbial communities used a DNA profiling technique through amplification of the sample using general or group-specific primers followed by denaturing gradient gel electrophoresis (DGGE) analysis (Deagle *et al*, 2005; Harper *et al*, 2006; Hiergeist *et al*, 2015). However, this method does not lend itself to high throughput analysis. Another early method to identify taxa included bacterial cloning and sequencing of amplicons generated by general or group-specific primers. Jo *et al* (2016) used DNA barcoding and cloning of amplicons to identify diet items consumed by brown trout, an invasive generalist feeder in Tasmanian lakes. Using primers for COI (specific for metazoans), they identified a 1.4-fold higher number of dietary items overall compared to visual quantification. However, the diversities of coleopterans and dipterans identified via DNA barcoding were 2.7-fold and 7-fold higher, respectively. This application has been used in the diet analysis of passerine birds, whales, marine amphipods and bivalves, penguins, dolphins,

bats, and humans (Blankenship & Yayanos, 2005; Deagle *et al*, 2007; Dunshea *et al*, 2008; Jarman *et al*, 2004; Rollo *et al*, 2002; Zeale *et al*, 2011).

DNA barcoding is a method of species identification using a short section of DNA from a specific gene or genes. DNA metabarcoding, using high-throughput sequencing technologies, allows the simultaneous amplification and identification of bar code sequences from many taxa within the same environmental sample and also allows the analysis of many samples simultaneously. DNA metabarcoding coupled with high-throughput sequencing technologies has allowed information on diet and microbiome studies to be obtained more rapidly and has uncovered much higher diversity (Pompanon *et al* 2012). Within recent years the lower cost of high throughput sequencing, expansion of sequence databases, and more user-friendly data analysis and bioinformatics tools have increased the use of DNA metabarcoding and high-throughput sequencing for microbiome and diet studies (Deagle *et al* 2013; Pompanon *et al*, 2012). The first papers using this method primarily dealt with analyzing fecal material from larger mammals such as Australian fur seals, penguin, and bats (Bohmann *et al*, 2011; Deagle *et al*, 2009; Deagle *et al*, 2010; Murray *et al*, 2011). In the marine environment DNA metabarcoding and high-throughput sequencing have become more common in studying the diet and microbiota of ecologically significant animals as well as the microbial community of water and sediments. Harms-Tuohy *et al* (2016) used this method to study the stomach content of the invasive lionfish. They identified species that were missed in the digested liquiform material from the stomach, using traditional microscopic methods. Given their results they argued that DNA analyses of fish gut contents could be used in monitoring or evaluating biodiversity. Filter feeders such as forage fish and mollusks provide an opportunity to supplement evaluations of biodiversity or environmental monitoring because they are effective environmental samplers. King *et al* (2016) filled a gap in oyster gut and stomach microbiome diversity by analyzing the stomach and gut contents of the eastern oyster. By using DNA metabarcoding and high-throughput sequencing, they found two distinct rich microbial communities in oyster stomachs and guts collected from two different

sites in Louisiana, USA. Furthermore, they found an approximate 8-fold increase in the number of species in gut and stomach core microbiota than had been identified previously. Previous studies had been predominantly culture-dependent, focusing on characterizing human or oyster pathogens. Even culture-independent studies using fingerprinting methods found only a few important taxa to compare the microbial patterns in the gut of different oyster populations.

Although DNA metabarcoding and high-throughput sequencing have substantially enhanced diet and microbiome studies, heedless application of this method can lead to the introduction of bias. **Chapter 2** discusses some of the technical issues that can introduce bias into DNA metabarcoding analysis and describes optimization of the method for looking at Gulf menhaden stomach contents. The choice of sequencing platforms is also important. There are several available high-throughput sequencing platforms each with their own set of limitations contributing to the sources of bias. To minimize errors and avoid bias, the appropriate sequencing application must be selected based on research goals. Perhaps the primary variable in sequencing technology is read length and error rates. Earlier microbiome and diet studies used the pyrosequencing technology of the Roche/454 platform (Roche), which had the advantage of producing longer read lengths over other available platforms. However, this came at the cost of having higher error rates that can lead to an overestimation of diversity. The Illumina sequencing platform has lower error rates, but the shorter read lengths, posing a challenge for designing barcodes that have sufficient discriminatory power for species level identification. Though the Illumina platform is currently a popular choice due to cost and lower sequencing errors, near full-length fragments longer than 1,300 bps are required for a comprehensive and reliable estimation of taxa richness (Yarza *et al*, 2014). Unfortunately, only 23 % of the 16S rRNA sequences are longer than 900 bps. Clooney *et al* (2016) did a comparison study across sequencing platforms, amplicon choice and sequence analysis software and found that the choice of taxonomic binning software proved to be more important in discriminatory power over sequencing platform and choice of amplicon. High-throughput sequencing technologies produce

millions of reads giving gigabytes worth of sequencing data. It is only by using bioinformatics tools that these data become usable to answer research questions. These tools are charged with the task of discarding sequences with errors, sorting the remaining sequences according to their tags, clustering them based on sequence similarity and assigning them to a taxon. Once this is complete, sequences can then be analyzed for ecological significance, e.g. by measuring diversity or statistically determining group abundance. Such processing tools also have the potential to introduce bias error giving a false sense of diversity. For example, one artifact of sequencing technology is the generation of chimeric sequences. Comparing the sequence to a known database or determining distance from other sequences usually finds chimeras (Soininen *et al*, 2009). Another critical step in the bioinformatics analysis is sorting sequences into clusters of operational taxonomic units (OTU) in which 16S or 18S rRNA sequences are considered to be from the same taxon (Edgar, 2013). Operational taxonomic unit is an operational definition to group sequences by similarity, equivalent to classical Linnaean or evolutionary taxonomy, i.e. OTUs are proxies for “species”. This is useful since not every organism has an rRNA sequence in available databases. Sequences can be clustered according to their similarity to each other, based on a similarity threshold (usually 97 %). Clusters are sorted (binned) based on a percent similarity depending on the sequence threshold set (usually 97 %). Poor threshold assignment can lead to errors in the estimation of diversity. Taxonomic assignment accuracy is largely dependent on the database being used. It is important to recognize that public databases may contain sequencing errors and incorrectly assigned taxa (Harris *et al*, 2003). In addition, databases may not contain sufficient taxonomic breadth.

### 1.8 Use of DNA metabarcoding for the analysis of the stomach content of juvenile Gulf menhaden

Despite the importance of Gulf menhaden as a dominant prey fish and its economic importance in fisheries, there is limited specific dietary information for this species and a more complete understanding of the diet is needed.

Assessments of the fishery have concluded that Gulf menhaden are not overfished or undergoing overfishing (Vaughan *et al*, 2007). However, this assessment did not fully consider the ecological role of the species and its ecological importance has not been adequately quantified. This suggests that the potential exists for this large fishery to yield an ecological impact (Olsen *et al*, 2014; Sagarese *et al*, 2016).

To understand food web dynamics, the whole dietary breadth needs to be described. I have developed a method for determining the stomach content of juvenile menhaden using DNA metabarcoding to look at diet items in the stomach, as well as the stomach microbiota. I have compared the results in fish caught at two different locations within Apalachicola Bay, one a low salinity location close to the Apalachicola estuary at Two Mile Channel (May 2013) and the other high salinity location in St. Vincent Sound (July 2013).

### 1.9 Research questions

1. Can a method based on DNA metabarcoding using rRNA gene sequences be used to describe the microbiota (using 16S rRNA sequences) and prey items (18S rRNA sequences) from the stomach contents of Gulf menhaden?
2. Does the stomach microbiota, as assessed by 16S rRNA metabarcoding differ in menhaden caught at different locations? i.e. could Gulf menhaden, as filter feeders, function as environmental samplers, and therefore be used as a possible biomonitor species to assess microbial diversity in inland bays and estuaries?
3. Does the application of DNA metabarcoding provide a wider description of menhaden prey items compared with previous methods?
4. Do the stomach diet items change with the location at which the fish are caught? This could differentiate between selective filtration of food items reflecting size or developmental stage of menhaden from an opportunistic feeding strategy.

## Chapter 2: Methodology for the identification of stomach contents in the filter feeding fish (*Brevoortia patronus*) using DNA metabarcoding

### 2.1 Abstract

Menhaden are filter feeding forage fish that serve as a trophic link between plankton and piscivorous predators in the marine environment. Dietary analysis is difficult in juvenile menhaden because >80 % of the stomach content is amorphous material. DNA metabarcoding allows a comprehensive assessment of stomach contents by combining mass-amplification of short DNA sequences (bar codes) with high-throughput sequencing. Here we describe a method for the assessment of diet items and gut microbiota of juvenile Gulf menhaden (*Brevoortia patronus*), collected within Apalachicola Bay, FL. The method describes the optimization of DNA extraction, effects of different sequencing protocols, estimation of menhaden DNA contamination, quality filtering of sequences, post-sequence processing and taxonomic identification of sequences. We characterized the stomach prokaryotic community using universal 16S ribosomal RNA (rRNA) gene sequencing primers in the V3-V4 hypervariable region. We explored the effects of two different sequencing protocols employing different “universal” 16S rRNA gene sequencing primers. The two protocols gave differences in the relative abundances of several bacterial classes. The dominant OTUs resulting from 16S rRNA gene sequencing were assigned to Proteobacteria, Acidobacteria, Actinobacteria and Chloroflexi and included oil eating bacteria consistent with the Gulf of Mexico location. Eukaryotic diet items were determined using universal sequencing primers targeting the V4-V5 hypervariable region of the 18S rRNA gene sequence. We identified OTUs belonging predominantly to copepods and diatoms. Overall, this study demonstrated a greater taxonomic richness of stomach contents than previously described, consistent with a depiction of menhaden as environmental samplers.

## 2.2 Introduction

Menhaden (*Brevoortia spp*) are small common pelagic schooling members of the family Clupeidae occupying the coasts and estuaries of the United States Atlantic and Gulf regions. Gulf menhaden (*Brevoortia patronus*) range from Veracruz, Mexico to southwestern Florida. They form large schools, which can be found moving throughout estuaries and near-shore regions in late spring and summer (Ahrenholz, 1991). Gulf menhaden is an important forage fish species along the Gulf coast, providing forage for several commercially important fishes in the Gulf of Mexico, including mackerel, bluefish, sharks, white and spotted seatrout, longnose gars and red drum (Sagarese *et al* 2016; Etzold & Christmas, 1979; Reintjes, 1970; Simmons & Breuer, 1964). They also constitute up to 97 % of food consumed by birds such as the brown pelican (Arthur 1931) and common loon (Pendleton, 1989; Ahrenholz, 1991). Furthermore, Gulf menhaden support the largest fishery by landings in the Gulf of Mexico. Gulf menhaden are reduced to fish oil and meal for use in livestock feed, aquaculture, pharma- and nutraceuticals, cosmetics and other consumer products (Nicholson, 1978).

Gulf menhaden are obligate filter feeders and juveniles have been shown to feed on phytoplankton, zooplankton and particulate organic matter. However, neither the full breadth of species consumed, nor the representation of any species in the diet can be determined by visual methods, because food is ground to an amorphous paste in the gizzard-like stomach (Lewis & Peters, 1994). To understand food web dynamics, the whole dietary breadth needs to be measured. Targeted PCR identification using species-specific bar codes can only show the presence or absence of species already thought to be present. DNA metabarcoding can greatly reduce the bias of sequence-specific methods by combining amplification of a “universal” gene region as a DNA barcode with high-throughput sequencing (next generation sequencing). Unbiased community structure could be monitored by denaturing gradient gel electrophoresis (DGGE) and clone library methods based on rRNA gene sequences. However, this approach requires extensive man-hours for picking/cloning/sequencing. DNA



barcoding is a method of species identification using a short section of DNA from a specific gene or genes. DNA metabarcoding allows for simultaneous identification of many taxa within the same environmental sample and allows the analysis of many samples simultaneously. DNA metabarcoding methods have been shown to enhance diet studies substantially in a range of fish species and can be used with the total and partially degraded DNA extracted from fish stomachs (Jakubavičiūtė *et al*, 2017; Waraniak *et al*, 2019). Using molecular barcoding methods, it is possible to identify prey items that lack visible diagnostic taxonomic features (Pompanon *et al*, 2012). Similarly, DNA metabarcoding and high-throughput sequencing can also be used to evaluate the microbiota of stomach contents (King *et al*, 2012).

Although DNA barcoding and high-throughput sequencing have substantially enhanced diet and microbiome studies, inappropriate application of this method without an appreciation of its limitations can lead to the introduction of bias. A number of steps, that include DNA extraction and PCR amplification, may hamper the objective of obtaining results that are truly representative of the source of DNA studied. Two factors that affect the successful amplification of extracted DNA are the quality and quantity of DNA (Eichmiller *et al* 2016; Li *et al*, 2018; Majaneva *et al*, 2018). Uneven or inefficient DNA extraction can result in only the most abundant organisms being sequenced and the subsequent underestimation of the diversity of diet and/or microbiota. To allow all organisms present to be identified requires uniform and efficient DNA extraction. However, it should be noted that there is no universal best method for isolating DNA from stomach contents and optimization is needed for each species and source investigated (Pollock *et al*, 2018).

An extensive literature can be found comparing DNA extraction methods often looking at different target taxa and environments and showing differences in DNA yield and PCR amplification success amongst methods (reviewed Schiebelhut *et al*, 2017). Commercial spin-column based methods make use of the binding of DNA to a silica membrane in the presence of a high concentration

of chaotropic salts (Boom *et al*, 1990) and non-target substances are rinsed off while the target DNA is bound to the silica membrane and can be eluted. Published DNA extraction protocols vary widely depending on sample type and intended use. Some choices include enzymatic digestion with lysozyme and/or Proteinase K, use of surfactants like sodium dodecyl sulfate (SDS) or cetrimonium bromide (CTAB), strong chaotropic agents like guanidine thiocyanate or urea, or physical methods such as bead-beating or freeze-thaw cycles. A combination of strategies is essential to maintain the balance between maximum cell disruption, low DNA degradation and efficient DNA extraction from all cell types. The gut microbial community can contain organisms from more than thirty different phyla that range in susceptibility to cell breakage and release of DNA and therefore efficiency of DNA extraction (Lagier *et al*, 2012; Rajilić-Stojanović *et al*, 2007; Hoffmann *et al*, 2013). If cells from only some species are lysed, the community analysis data will be skewed. The inclusion of a bead-beating step, in which a sample is agitated rapidly with beads or balls in a device that shakes the homogenization vessel, has been linked to higher DNA yields (Schiebelhut *et al*, 2017; Ushio, 2019). The inclusion of a bead beating step has been shown to allow more efficient extraction of DNA from Gram-positive and spore-forming bacteria uncovering a higher bacterial diversity (Han *et al*, 2019; Jiang *et al*, 2019; Ketchum *et al*, 2018). DNA can also be lost by adsorption to the surface of various particles in the sample from sludge or soil and so contribute to bias (Vanysacker *et al* 2010). Extracting DNA from the stomach contents of a filter feeding marine organism such as menhaden, in which phytoplankton and plant detritus can be present, can lead to adsorption to the surface of various particles in the sample (Vanysacker *et al*, 2010).

Co-extraction of PCR inhibitors with DNA can interfere with downstream amplification (Claassen *et al*, 2013). Complex polysaccharides, bile salts, lipids, and urates in stomach contents are all known PCR inhibitors that require additional steps to be removed (Schrader *et al*, 2012). Many manufacturers have developed DNA extraction kits designed to remove PCR inhibitors. Other products are available to cope with the presence of PCR inhibitors when included

in PCR amplification reactions and sequencing. PCR inhibition may lead to only the most abundant organisms being sequenced and the subsequent underestimation of the diversity of diet and/or microbiota. Conversely, extra purification steps can reduce DNA yield and lead to the underestimation of diversity.

Another essential aspect demanding careful consideration is the choice of barcode. The 16S rRNA gene sequence has long been the gold standard for identifying prokaryotes, making use of variable regions that occur between highly conserved sites within the 16S rRNA gene of bacteria (Degnan *et al*, 2012; Quast *et al*, 2013). In eukaryotes, the 18S rRNA gene or the mitochondrially encoded cytochrome c oxidase I (COI, also called COX1) gene are most commonly used. (Folmer *et al*, 1994; Martin *et al*, 2006; Jarman *et al*, 2006; Pompanon *et al*, 2012; King *et al*, 2012; Wang *et al*, 2014; Hugerth *et al*, 2014). However, Klindworth *et al* (2012) have shown that the barcodes should be optimized in respect to their overall coverage and phyletic spectrum expected. The mitochondrially encoded COI was developed as a potential bar code for eukaryotes. It works well for vertebrates and many metazoa, but COI gene sequences lacks the discriminatory power to identify protist and fungal species (Bellemain *et al*, 2010). Newer barcodes targeting 18S rRNA genes have been developed for identification of species across eukaryotes (Hugerth *et al*, 2014) and are used here for identifying eukaryotic diet items. 18S rRNA gene sequences provide better coverage of eukaryotic taxa than that provided by available sequences of mitochondrial CO1 (Deagle *et al*, 2014) and the SILVA database is superior for annotating 18S rDNA sequences to finer taxonomic levels than the NCBI nt database (Pruesse *et al* 2007; Lindeque *et al*, 2013). Pertinent to our study here, significantly higher total OTU richness was recovered from the marine zooplankton community off the Florida Keys using 18S rRNA gene sequencing data compared with COI (Djurhuus *et al*, 2018).

A suitable barcode sequence needs to have conserved sequences at the 5'- and 3'-ends with hypervariable region(s) in between. Primers to amplify these

regions are designed using multiple sequence alignment to identify sites that are conserved within a group of taxa, but unique between groups, ideally with equal amplification efficiency from all taxa sequenced (Jarman *et al*, 2004; Jarman *et al*, 2006; King *et al*, 2008). Again, having *a priori* knowledge of the system is useful as the primers chosen can selectively introduce bias by underestimating one taxon and over-estimating another. When looking at communities, the design of “universal” PCR primers can have an effect on phylogenetic resolution. No primer set is truly universal and some commonly used 16S rRNA gene sequence primers have proved ineffective at amplifying biologically relevant bacteria, (Gołębiewski & Tretyn, 2019). Here we compare results using two different “universal” 16S rRNA gene sequence primers and a modified sequencing protocol that provide differential amplification of DNA.

Another source of bias can be in the choice of sequencing platforms. There are several available high-throughput sequencing platforms each with their own set of limitations contributing to the sources of bias. To minimize errors and avoid bias, the appropriate sequencing application must be selected based on research goals. The primary variables in sequencing technology are read length and error rates. The Illumina MiSeq system has become the most commonly used sequencing platform for 16S and 18S rRNA gene metabarcoding and is used in this study. In general, the MiSeq platform produces the longest and most accurate sequences and has a much higher throughput than the other platforms. This enables more samples to be sequenced at higher depth or lower cost (Forin-Wiart *et al*, 2018; Quail *et al*, 2012).

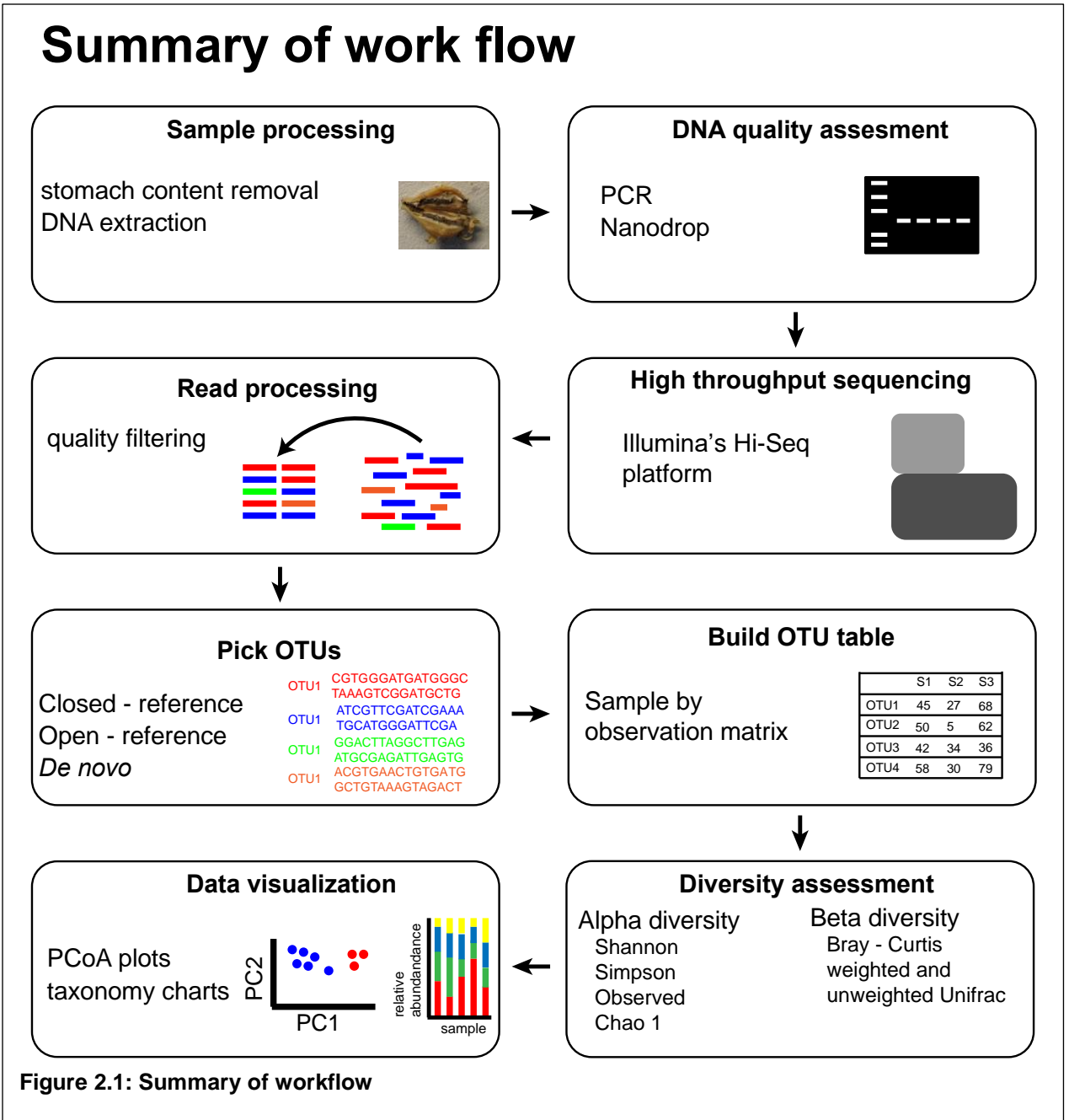
High-throughput sequencing technologies produce millions of sequences giving gigabytes worth of sequencing data. It is only by using bioinformatics tools to analyze big data sets that these data become usable to answer research questions. Such tools are needed for quality checks to discard sequences with errors. Sequences can be sorted based on sequence similarity to sequences in the 16S and 18S rRNA databases and assigned to a taxon (Grabowski & Rappsilber, 2019). Sequence analysis software and the choice of taxonomic

binning software can also make a significant difference to results (Clooney *et al*, 2016).

The purpose of this study was to develop a method to characterize stomach contents of a key filter feeding forage fish species, Gulf menhaden. We did this by reducing sources of bias such as that introduced by the DNA extraction method, refining the amplification and sequencing methods, choosing primers wisely, and analyzing the results using common bioinformatic tools and diversity metrics. The method allowed characterization of the stomach prokaryotic community, as well as eukaryotic diet items. We intend to use this method in the future to examine the effects of location, season and developmental stage on the diet and microbiota of menhaden. It could also be used to monitor Gulf menhaden stomach contents over time to tease out effects of climate change. We also anticipate applying the method to other filter feeding species. Our overall purpose was to provide a path to gain a better understanding of microbial diversity in menhaden stomachs, as well a better understanding of their diet.

### 2.3 Materials and Procedures

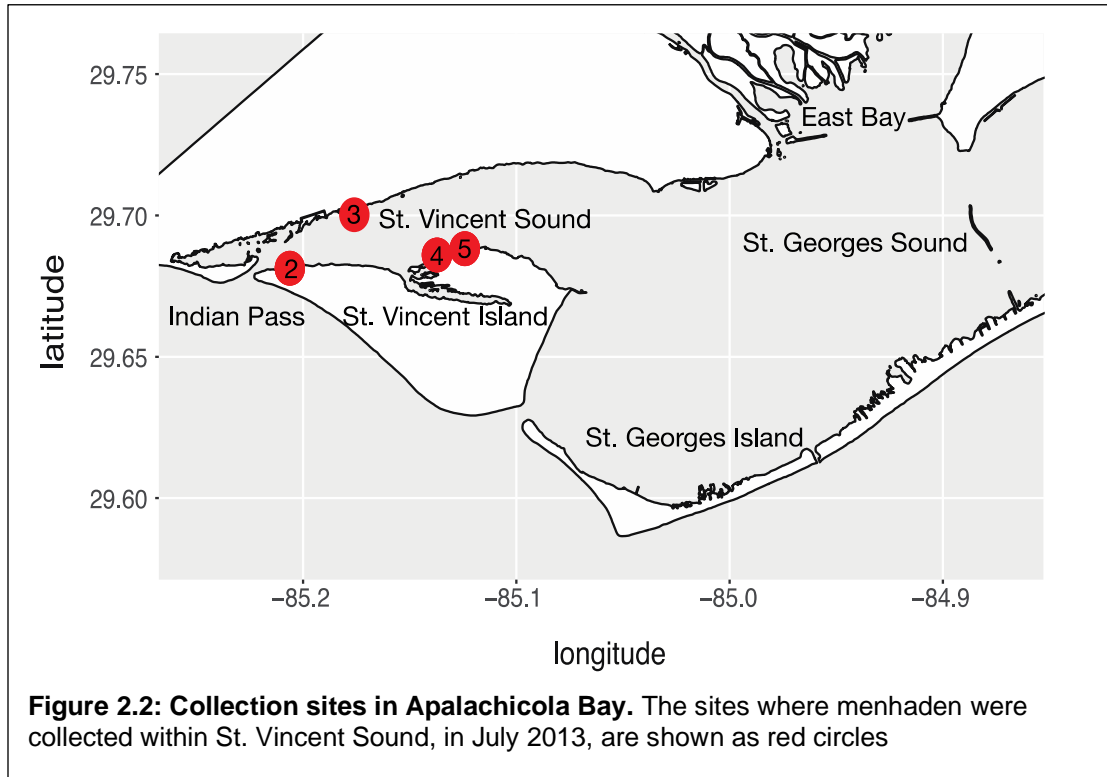
A summary of the workflow involved in the identification of the stomach contents of juvenile Gulf menhaden (*Brevoortia patronus*) using DNA metabarcoding and high throughput sequencing are shown in **Figure 2.1**.



### 2.3.1 Sample collection

Ten Gulf menhaden stomach samples were received from our Delaware State University and Florida Fish and Wildlife collaborators. The fish were collected from Apalachicola Bay, FL on July 2, 2013, at four sites (SVS02, SVS03, SVS04, SVS05) in St. Vincent Sound (SVS) with latitude and longitude coordinates corresponding to 29.68, -85.2; 29.70, -85.1; 29.68, -85.1; 29.6, -85.1,

respectively (**Figure 2.2**). Collections were made using a seine net at 1 m depth. Water quality measurements of temperature, salinity, pH, dissolved oxygen, and turbidity were recorded using a YSI 556 multiparameter water quality meter (**Supplemental Table 2.1**).



### 2.3.2 Optimization of DNA extraction

Given that menhaden are filter feeders and their stomachs contain amorphous environmental material, we assumed that the stomach contents could possess many potent PCR inhibitors. Furthermore, the menhaden diet is known to consist of organisms that resist cell lysis necessitating the need to evaluate DNA extraction and quality. There are a range of commercial kits available to produce high quality DNA free of PCR inhibitors in high yields. We extracted the DNA of menhaden stomach contents using several of these and compared DNA quality, DNA yield, and its ability to be amplified by PCR. Four commercial DNA extraction kits, DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany), PowerSoil DNA Isolation kit (Mo Bio, Santa Anna, CA, US), *Quick-DNA* Miniprep kit (Zymo, Irvine, CA, US), and High Pure PCR Template DNA Preparation kit (ROCHE

Diagnostics, Indianapolis, IN, US) were compared to ascertain the best method for extraction of menhaden stomach contents. Each kit was used as per the manufacturer's instructions, although all included a bead beating step. DNA quality was assessed spectrophotometrically using the Nanodrop 1000 (ThermoFisher Scientific, Waltham, MA, US) and by PCR amplification using the universal prokaryotic 16S rRNA gene sequence primers 27F-1492R (Weisburg *et al*, 1991). PCR amplicons were visualized by ethidium bromide staining after 1 % agarose gel electrophoresis.

### *2.3.3 Sample preparation and DNA extraction*

Whole stomachs were removed from fish and placed in 95 % ethanol prior to transit to IMET. Using a surgical blade, stomachs were cut in half and gently shaken in 1 ml of 100 % ethanol until contents were released.

DNA extraction was done using the High Pure PCR Template DNA Preparation kit (ROCHE Diagnostics, Indianapolis, IN, US) with some modifications to the manufacturer's protocol. The stomach contents in ethanol were left at 60 °C until dry. Two hundred µl of lysis buffer (ROCHE Diagnostics, Indianapolis, IN, US) was added to the sample along with 10 µl Proteinase K (20 mg/ml) and incubated at 70 °C for 1 h. This extract was transferred to a 2 ml screw cap tube containing garnet beads (Mo Bio Santa Ana, CA, US). Disruption of the sample was achieved by bead beating using the FastPrep FP120 bead beating apparatus (Savant Instruments, Inc., Holbrook, NY) for 20 sec at 4500 rpm. Once this step was completed, manufacturer's protocols were followed for the subsequent cellular debris removal, washes and DNA elution steps.

### *2.3.4 High throughput sequencing*

Sequencing of both the 16S rRNA gene and 18S rRNA gene amplicons of all 10 samples was performed on Illumina's MiSeq platform located in the BioAnalytical Services Laboratory (BAS Lab) at the University of Maryland Center for Environmental Science-Institute of Marine and Environmental Technology. To assess bias in sequencing methods, 16S rRNA genes were



sequenced using two different protocols and primers that amplify the V3-V4 variable regions in 16S rRNA gene sequence. The first method used the recommended Illumina protocol for the MiSeq as described in the Nextera DNA Library Prep Reference Guide (Illumina Document #15027987v1) and is referred to as the Illumina dataset. The second method used a dual indexing amplification and sequencing approach as described in Fadrosch *et al* (2014) and is referred to as the Fadrosch data set. In the Fadrosch protocol, the inclusion of heterogeneous spacers at the ends of the primers increases the efficiency of primer hybridization and overcomes the under-representation of low frequency and low primer homology sequences during the first several cycles of a sequencing run. Both the recommended Illumina and Fadrosch sequencing protocols use sequencing primers targeting the V3-V4 variable regions of 16S rRNA gene sequence, as described in Klindworth *et al* (2012) and Fadrosch *et al* (2014) were sequenced in both directions. The primer sets are shown in **Table 2.1**.

For eukaryotic species, the sequencing was done as described above for 16S rRNA sequencing except that the sequencing primer set used was for the 18S rRNA gene sequence. The primers used were 574\*F (CGGTAAYTCCAGCTCYV) and 1132R (CCGTCAATTHCTTYAART) in the 18S rRNA gene V4-V5 region as developed by Hugerth *et al* (2014). The position numbering refers to the position in the rRNA sequence, as identified in the *Saccharomyces cerevisiae* strain FM-sc-08 18S rRNA gene, NCBI accession number Z75578. The 18S rRNA primers are also given in **Table 2.1**.

The list of primers is given in **Table 2.1**. Individual "index" sequences are added to DNA fragments from each sample during the library preparation so that sequences from each sample can be identified, edited and sorted before the final data analysis.

Name	Sequence	Source
319F	ACTCCTACGGGAGGCAGCAG	Fadrosh et. al 2014
806R	GGACTACHVGGGTWTCTAAT	Fadrosh et. al 2014
S-D-Bact-0341-b-S-17	CCTACGGGNGGCWGCAG	Klindworh et. al 2013
S-D-Bact-0785-a-A-21	GACTACHVGGGTATCTAATCC	Klindworh et. al 2013
574*F	CGGTAAYTCCAGCTCYV	Hugerth et. al 2014
1132R	CCGTCAATTHCTTYAART	Hugerth et. al 2014

**Table 2.1: Primers used for sequencing**  
Primer sequences used for amplification and sequencing of 16S and 18S rRNA gene sequences.

### 2.3.5 Post-sequencing pipeline

Sequence identity and data quality assessment were performed on the MiSeq instrument itself. MacQIIME was used to process and assess quality of output sequences (called reads) from the sequencing primers. Post-sequence processing was done using the recommended QIIME pipeline for Illumina reads (Caporaso *et al*, 2010). Removal of index sequences, called de-multiplexing, takes place on the MiSeq instrument. The MacQIIME script `join_paired_ends.py` was used to join forward and reverse sequences. Paired reads were filtered for low quality reads (quality score of <25) and short read length (<200 bp) and removed from the library using the `split_libraries.py` command. Chimeric sequences were identified *de novo* using the USEARCH61 (v6.1.544) algorithm with the script `identify_chimeric_seqs.py` (Edgar 2010). This was followed by the removal of PhiX sequences by a BLAST analysis with the scripts `parallel_blast.py` and `filter_fasta.py`. After the removal of chimeric and PhiX sequences, the remaining reads were used for operational taxonomic unit (OTU) picking using the script `pick_otus.py`. For prokaryotic species in particular, many more species exist than have been cultured and identified taxonomically. An OTU is an operational term, most commonly defined as a cluster of reads with 97 % similarity, based on the expectation that OTUs can be used as a proxy for species (Sneath & Sogal, 1973; Mysara *et al*, 2017). A UCLUST *de novo* clustering method within the USEARCH61 (v6.1.544) algorithm was used to pick OTUs for both 16S sequencing protocols (Edgar 2010). Taxonomic assignment was done using the script `assign_taxonomy.py` with Greengenes `gg_13_8` as a

reference database at 90 % similarity (Lawrence Berkeley National Laboratory, <http://greengenes.lbl.gov>). These data were used to construct an OTU table for subsequent relative sequence abundance and diversity analysis.

The 18S rRNA gene sequences were also processed in MacQIIME using the appropriate scripts as described above. OTUs generated from 18S rRNA gene sequencing were picked and taxonomy was assigned using the UCLUST method against the Silva 111 Eukaryote-only database (Edgar 2010; Quast *et al*, 2013).

The resulting OTU tables and metadata table are available at <https://github.com/Hanif82/gulfmenhadenotutable.git>.

### 2.3.6 *Analysis of diversity*

Observed OTUs, Good's coverage, Fisher's alpha, Chao1 richness, Shannon index, and Inverse Simpson index metrics were used to assess alpha diversity using the programs R and MacQIIME. Metastats was used to test for differentially abundant taxa with p-value adjustment using the False Discovery Rate (FDR) (White *et al*, 2009). To analyze how closely the samples were related to each other and to compare observed differences in the microbial community from the two sequencing methods, beta diversity analyses were determined based on the unweighted and weighted UniFrac phylogenetic distances metric and visualized using the ordination method principal coordinate analysis (PCoA) into two-dimensional plots.

### 2.3.7 *Menhaden 18S rRNA gene sequence and estimation of menhaden DNA contamination*

Using the SILVA 111 eukaryote-only database for taxonomic assignment, we observed a significant number of reads that were assigned to Reeves shad, *Tenulosa reevesii*, a clupeid fish closely related to Gulf menhaden. Given that there is no record of Reeves shad in the Gulf of Mexico, we assumed these reads to be of the host, Gulf menhaden. Based on this, we calculated the percentage of reads assigned to Reeves shad as a proxy for estimating the percentage of reads from menhaden DNA in the stomach DNA samples. Using

the 18S rRNA DNA sequence deposited in GenBank for Reeves shad (accession number EU12003.1) five primer sets (a combination of universal eukaryotic primers from the literature and primers designed in this study) were used to amplify and sequence the menhaden 18S rRNA gene producing a partial sequence of 1489 base pairs (accession number MN335220, **Supplemental Figure 2.1**).

#### *2.4 Assessment of method*

##### *2.4.1 Optimization of DNA extraction and quality assessment*

In order to make the extraction method transferable to other investigations, we opted for the use of commercially available kits with the expectation of having to do some adjustment to the manufacturer's protocol.

Spectrophotometry is a commonly used method to assess DNA quality and quantity. Here we used Nanodrop 1000 (ThermoFisher Scientific, Waltham, MA, US) to measure absorbance from 230-320 nm to calculate the concentration of DNA and determine A260/A280 and A260/A230 ratios. Good quality DNA should have a A260/A280 and A260/A230 ratio of 1.7-2.0 and 2.0-2.2 respectively. Strong absorbance at 230 nm and/or 280 nm will lower expected ratios indicating contaminants such as organic or chaotropic salts. Several commercial kits use silica to bind nucleic acids which can co-elute and has a measured absorbance of approximately 230 nm lowering the A260/A230 ratio. However, this does not

interfere with downstream molecular applications such as PCR. Of the four kits we tested, the Qiagen DNeasy Blood & Tissue kit and the Roche High Pure PCR Template Preparation kit gave DNA with acceptable A260/A280 ratios and the Roche kit gave the best A260/A230 ratios. However, only the High Pure PCR

DNA extraction kit	ng/ul	260/280	260/230	PCR
Qiagen	83.96	1.72	0.73	-
Qiagen	62.56	1.65	0.82	-
Qiagen	78.33	1.62	0.79	-
Qiagen	60.71	1.71	0.76	-
MoBio	13.92	1.59	0.83	-
Mo Bio	34.06	1.3	1.04	-
Roche	56.84	1.69	1.05	+
Roche	82.62	1.64	0.95	+
Roche	87.99	1.82	1.27	+
Roche	81.51	1.98	1.67	+
Zymo	3.7	1.9	0.11	-
Zymo	0.52	-0.07	0	-
Zymo	6.55	1.45	0.16	-

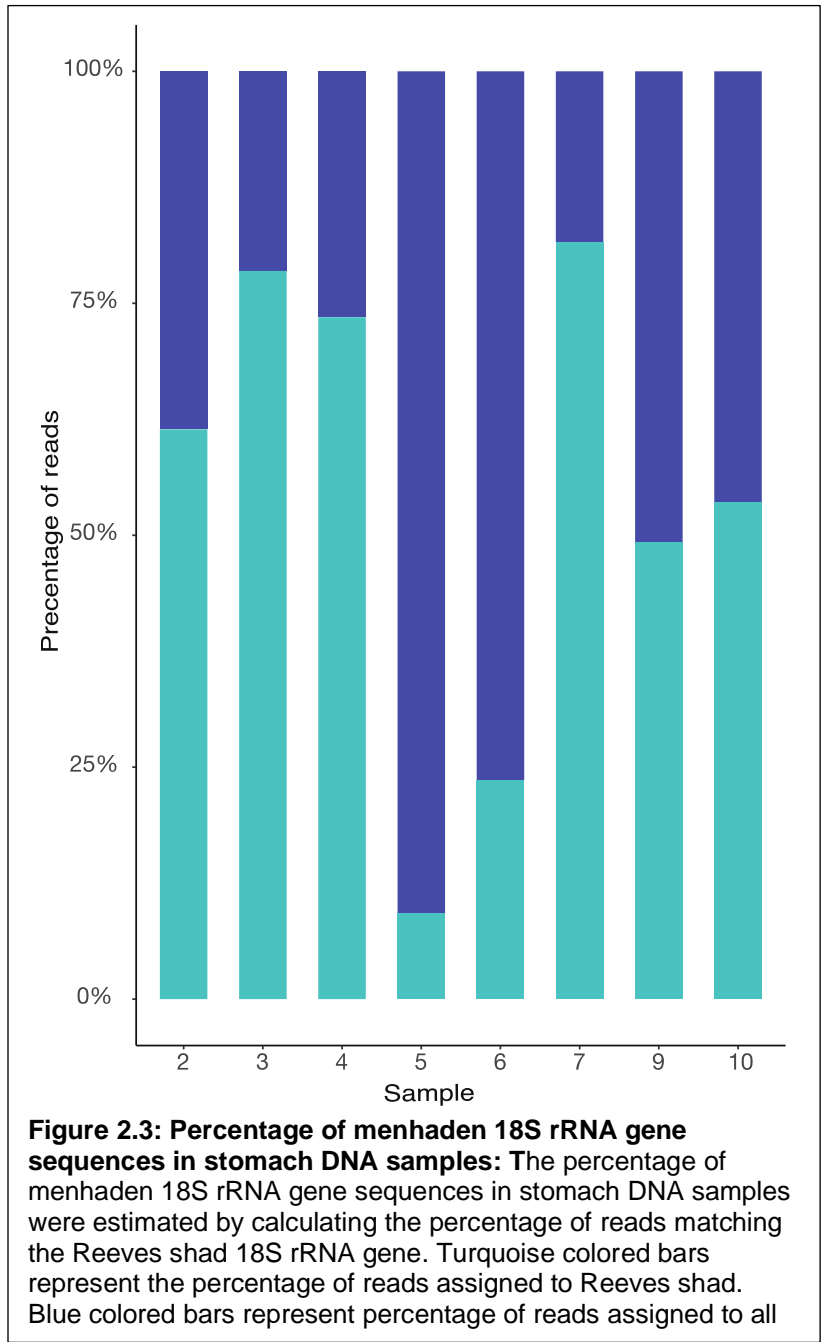
**Table 2.2: DNA recovery and quality using different DNA extraction kits.** DNA was extracted from menhaden stomach contents using four different commercial DNA extraction kits. The quality and quantity of DNA recovered was assessed by spectrophotometry from A230 to A320 using NanoDrop 1000, as well as by its ability to be amplified by universal primers 27F and 1492R for 16S rRNA genes (Weisburg *et al*, 1991).

Template Preparation kit (Roche, Diagnostics, Indianapolis, IN, US) gave extracted DNA that was amplifiable using the 16S rRNA gene sequence primers 27F and 1492R (Weisburg *et al*, 1991) (**Table 2.2**). The size of the amplicon, 1465 bp was also an indication that little degradation had occurred during purification.

#### 2.4.2 Assessment of menhaden DNA contamination in stomach DNA samples

In the analysis of a microbial community in stomach contents by high throughput sequencing, too much host DNA contamination can bias results. Though care was taken during DNA extraction to minimize menhaden DNA contamination, it was not eliminated. Initially there was no straightforward way to test for this prior to the high-throughput sequencing step because the sequence for menhaden 18S rRNA gene had not been found in gene databases. However,

after the taxonomic assignments were made using the SILVA 111 database, we



noticed that a high percentage of reads closely matched a related clupeid fish, *Tenulosa reevesii*, Reeves shad, (Richardson, 1846). Reeves shad belongs to the same family as menhaden but is native to the Northwest Pacific Ocean. Given the close phylogenetic relationship and no known reports of Reeves shad in the Gulf of Mexico, we assumed these reads represented menhaden DNA. Using the sequence of Reeves shad 18S rRNA gene sequence as a proxy for

menhaden 18S rRNA sequence, we estimated the amount of menhaden DNA contamination in individual fish. The percentage of the 18S rRNA reads considered to be from menhaden ranged from 9 to 81 % (**Figure 2.3**). In view of these results, the reads considered to be menhaden were excluded from the

eukaryotic community analysis in all samples. Despite this, the samples gave good depth of coverage, as indicated by the Goods coverage index (see below).

Prior to this study, sequence for the menhaden 18S rRNA gene was absent in gene databases. This lack of information in the database is a prime example of the shortcomings of metabarcoding methods that rely on well curated gene databases for high taxonomic resolution and assignment. Primer sets based on the sequence of *T. reevesii* allowed amplification of menhaden 18S rRNA DNA using DNA purified from menhaden fin clips and generated a partial sequence of 1489 bp (accession number MN335200) (**Supplemental Figure 2.1**).

A BLAST analysis of the menhaden 18S rRNA partial gene sequence recovered against the GenBank database showed >99 % identity to several other closely related Clupeidea. An alignment of the menhaden sequence against the top hits, *Tenulosa reevesii* (accession number EU12003.1), *Potamulosa richmondia*, (accession number KJ774739.1) and *Nematalosa erebi* (accession number HQ615575.1) is shown in **Figure 2.4**. All four species are from different parts of the world, but all are from the family Clupeidae. Identification of the menhaden 18S rRNA sequence confirmed our assumption that the apparent *Tenulosa reevesii* sequences represented Gulf menhaden 18S rRNA gene sequences and justified our exclusion of these sequences from the community analysis.

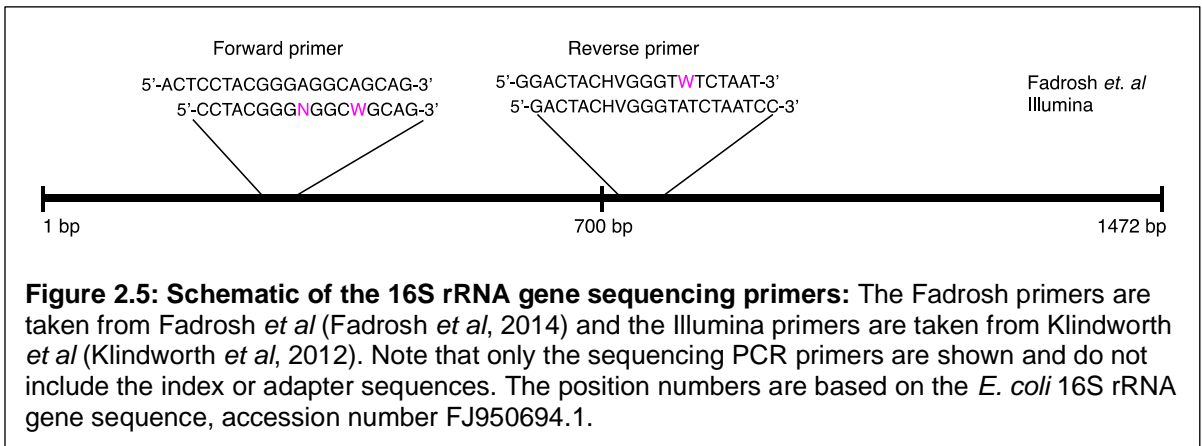
*Menhaden P. richmondia T. reevesii N. erebi	CCAGT AACAT AT GCT T GT CT CAAAAGAT AAGCGAT GCAGGT CT AAGT GCACACGGCCGGT ACAGT GAAACT GCGAAT GGCT CATT AAAT CAGT TAT GGT T CCT T T GAT GC AAGCGAT GCAGGT CT AAGT GCACACGGCCGGT ACAGT GAAACT GCGAAT GGCT CATT AAAT CAGT TAT GGT T CCT T T GAT GC AAGCGAT GCAGGT CT AAGT GCACACGGT GGGT ACAGT GAAACT GCGAAT GGCT CATT AAAT CAGT TAT GGT T CCT T T GAT GC	110 82 0 0
*Menhaden P. richmondia T. reevesii N. erebi	CT CCCACCCGT ACT T GGAT AACT GT GGCAAT T CCGAGAGT AAT ACAT GCAAAACGAGCGCT GACT GGCCCCCT T CACGGGGGGCT GCGGAT GCGT GCAT T TAT CAGAT CCA CT CCCACCCGT ACT T GGAT AACT GT GGCAAT T CCGAGAGT AAT ACAT GCAAAACGAGCGCT GACT GGCCCCCT T CACGGGGGGCT GCGGAT GCGT GCAT T TAT CAGAT CCA CT CCCACCCGT ACT T GGAT AACT GT GGCAAT T CCGAGAGT AAT ACAT GCAAAACGAGCGCT GACT GGCCCCCT T CACGGGGGGCT GCGGAT GCGT GCAT T TAT CAGAT CCA	220 192 0 192
*Menhaden P. richmondia T. reevesii N. erebi	AAAACCCAT CCGGGGGCT CGT GCCCCCGT CCGG T T GGT GACT CT AGT A ACCT CGGGCGGAT CCGGGCCCT CCGGGGGGGGACGT GT CT T CGAAT GT CT GCCCT AAAACCCAT CCGGGGGCT CGT GCCCCCGT CCGG T T GGT GACT CT AGT A ACCT CGGGCGGAT CCGGGCCCT CCGGGGGGGGACGT GT CT T CGAAT GT CT GCCCT AAAACCCAT CCGGGGGCT CGT GCCCCCGT CCGG T T GGT GACT CT AGT A ACCT CGGGCGGAT CCGGGCCCT CCGGGGGGGGACGT GT CT T CGAAT GT CT GCCCT	330 302 0 302
*Menhaden P. richmondia T. reevesii N. erebi	AT CAAT T T CGAT GGT ACT T T CCGCGCT ACCAT GGT GACCA CCGGT AACGGGAAT CAGGGT T CGAT T CCGGAGGGAACGCT GAGAAA CCGGT A CCAACAT CCAAGGAA AT CAAT T T CGAT GGT ACT T T CCGCGCT ACCAT GGT GACCA CCGGT AACGGGAAT CAGGGT T CGAT T CCGGAGGGAACGCT GAGAAA CCGGT A CCAACAT CCAAGGAA AT CAAT T T CGAT GGT ACT T T CCGCGCT ACCAT GGT GACCA CCGGT AACGGGAAT CAGGGT T CGAT T CCGGAGGGAACGCT GAGAAA CCGGT A CCAACAT CCAAGGAA	440 412 20 412
*Menhaden P. richmondia T. reevesii N. erebi	GGCAGCAGCGCGCAAA T ACCCAT T ACCGACACCGT GAGGT AGT GACGAAAAA AACAA T ACAGGT CT CT T CAGAGCCCT GT AAT T GGAAT GAGCGT AT CCT AAACCC GGCAGCAGCGCGCAAA T ACCCAT T ACCGACACCGT GAGGT AGT GACGAAAAA AACAA T ACAGGT CT CT T CAGAGCCCT GT AAT T GGAAT GAGCGT AT CCT AAACCC GGCAGCAGCGCGCAAA T ACCCAT T ACCGACACCGT GAGGT AGT GACGAAAAA AACAA T ACAGGT CT CT T CAGAGCCCT GT AAT T GGAAT GAGCGT AT CCT AAACCC	550 522 130 522
*Menhaden P. richmondia T. reevesii N. erebi	AT GGGCGAGGACCCAT T GGAGGGCAAGT CT GGT GCGAGCAGCGGGT AAT T CCGAGT CCAAT AGGCT AT T AAAGT T GCT CAGT T AAAAAAGT CCGT AGT T GGAAT T T C AT GGGCGAGGACCCAT T GGAGGGCAAGT CT GGT GCGAGCAGCGGGT AAT T CCGAGT CCAAT AGGCT AT T AAAGT T GCT CAGT T AAAAAAGT CCGT AGT T GGAAT T T C AT GGGCGAGGACCCAT T GGAGGGCAAGT CT GGT GCGAGCAGCGGGT AAT T CCGAGT CCAAT AGGCT AT T AAAGT T GCT CAGT T AAAAAAGT CCGT AGT T GGAAT T T C	660 632 240 632
*Menhaden P. richmondia T. reevesii N. erebi	GGAGT GGGT GGGGT CCGCCGAGCGGT GCGACCGT CT GT CCGT ACCCT GCGT CCGGGGCT CCGCGAT GCGCT AGCT GGGT GT CCGT CCGGGCCCGGAGCG GGAGT GGGT GGGGT CCGCCGAGCGGT GCGACCGT CT GT CCGT ACCCT GCGT CCGGGGCT CCGCGAT GCGCT AGCT GGGT GT CCGT CCGGGCCCGGAGCG GGAGT GGGT GGGGT CCGCCGAGCGGT GCGACCGT CT GT CCGT ACCCT GCGT CCGGGGCT CCGCGAT GCGCT AGCT GGGT GT CCGT CCGGGCCCGGAGCG	769 741 349 742
*Menhaden P. richmondia T. reevesii N. erebi	T T T ACT T T GAAAAA T T AGAGT GT T CAAAGCAGCGCCGACGACCGCT GAAT ACCGAGCT AGGAAT AAT GGAAT AGGACT CCGGT T CT AT T T T GGGT T T T CCGAACCA T T T ACT T T GAAAAA T T AGAGT GT T CAAAGCAGCGCCGACGACCGCT GAAT ACCGAGCT AGGAAT AAT GGAAT AGGACT CCGGT T CT AT T T T GGGT T T T CCGAACCA T T T ACT T T GAAAAA T T AGAGT GT T CAAAGCAGCGCCGACGACCGCT GAAT ACCGAGCT AGGAAT AAT GGAAT AGGACT CCGGT T CT AT T T T GGGT T T T CCGAACCA	879 851 459 852
*Menhaden P. richmondia T. reevesii N. erebi	GGCCAT GAT T AAGAGGACCGCGGGGCAT T CGT AT T GCGCCCT AGAGT GAAAT T CT T GGACCGGCGCAAGCGGAGGAAAGCAAGCAT T T GCCAAGAAT GT T T GGCCAT GAT T AAGAGGACCGCGGGGCAT T CGT AT T GCGCCCT AGAGT GAAAT T CT T GGACCGGCGCAAGCGGAGGAAAGCAAGCAT T T GCCAAGAAT GT T T GGCCAT GAT T AAGAGGACCGCGGGGCAT T CGT AT T GCGCCCT AGAGT GAAAT T CT T GGACCGGCGCAAGCGGAGGAAAGCAAGCAT T T GCCAAGAAT GT T T	989 961 569 962
*Menhaden P. richmondia T. reevesii N. erebi	T CAT T AAT CAAGAACGAAAGT CCGAGGT T CGAAGCAGT CAGAT ACCGT GGT AGT T CCGACCGT AAACGAT GCCGACCCCGAT CCGGGCGGT T AT T CCGAT GACCCG T CAT T AAT CAAGAACGAAAGT CCGAGGT T CGAAGCAGT CAGAT ACCGT GGT AGT T CCGACCGT AAACGAT GCCGACCCCGAT CCGGGCGGT T AT T CCGAT GACCCG T CAT T AAT CAAGAACGAAAGT CCGAGGT T CGAAGCAGT CAGAT ACCGT GGT AGT T CCGACCGT AAACGAT GCCGACCCCGAT CCGGGCGGT T AT T CCGAT GACCCG	1099 1071 679 1072
*Menhaden P. richmondia T. reevesii N. erebi	CGGGCAGCT GCGGGAAACCAAGT CT T T GGT T CCGGGGGAGT AT GGT T GCAAAGCT GAAACT T AAAGGAA T T GACGGAAGGGCACACCAGGAT GGAAGCT GCGG CGGGCAGCT GCGGGAAACCAAGT CT T T GGT T CCGGGGGAGT AT GGT T GCAAAGCT GAAACT T AAAGGAA T T GACGGAAGGGCACACCAGGAT GGAAGCT GCGG CGGGCAGCT GCGGGAAACCAAGT CT T T GGT T CCGGGGGAGT AT GGT T GCAAAGCT GAAACT T AAAGGAA T T GACGGAAGGGCACACCAGGAT GGAAGCT GCGG	1209 1181 789 1182
*Menhaden P. richmondia T. reevesii N. erebi	CT T AAT T T GACT CAACACGGGAACCT CACCGGGCGGGAACGGAAAGGAT T GAGAGT T GAT AGCT CT T T CT CAGT T CT GT GGGT GGT GGT GCAT GCGCGT CT T AGT CT T AAT T T GACT CAACACGGGAACCT CACCGGGCGGGAACGGAAAGGAT T GAGAGT T GAT AGCT CT T T CT CAGT T CT GT GGGT GGT GGT GCAT GCGCGT CT T AGT CT T AAT T T GACT CAACACGGGAACCT CACCGGGCGGGAACGGAAAGGAT T GAGAGT T GAT AGCT CT T T CT CAGT T CT GT GGGT GGT GGT GCAT GCGCGT CT T AGT	1319 1291 899 1292
*Menhaden P. richmondia T. reevesii N. erebi	T GGT GGAAGGAT T T GT CT GGT T AAT T CCGAT AACGAACGAGCT CCT CCAT GCT AAAT AGT T ACGGGCCCT GCGT CCGCGT T T CAACT CT T AGAGGGAAGAAT GG T GGT GGAAGGAT T T GT CT GGT T AAT T CCGAT AACGAACGAGCT CCT CCAT GCT AAAT AGT T ACGGGCCCT GCGT CCGCGT T T CAACT CT T AGAGGGAAGAAT GG T GGT GGAAGGAT T T GT CT GGT T AAT T CCGAT AACGAACGAGCT CCT CCAT GCT AAAT AGT T ACGGGCCCT GCGT CCGCGT T T CAACT CT T AGAGGGAAGAAT GG	1429 1401 1009 1402
*Menhaden P. richmondia T. reevesii N. erebi	CGT T T AGCCACCGGAGT GGAACAAT AACAGGT CT GT GAT GCCC CGT T T AGCCACCGGAGT GGAACAAT AACAGGT CT GT GAT GCCC CT T AGAT GT CCGGGGCT GCAACCGGGCCACAAAT GGGT GAT CAGCGT GT GT CT ACT CT CCGCCGACA CGT T T AGCCACCGGAGT GGAACAAT AACAGGT CT GT GAT GCCC CT T AGAT GT CCGGGGCT GCAACCGGGCCACAAAT GGGT GAT CAGCGT GT GT CT ACT CT CCGCCGACA	1473 1511 1119 1512
*Menhaden P. richmondia T. reevesii N. erebi	GGCGGGT AACCGGT GAAACCACT CGT GAT T GGGACT GGGAT T GAAACT AT T T CCGAT CAACGAGGAA T CCAAGT AAGCGGGT CAT AAGCT CCGCT T GAT T AA GGCGGGT AACCGGT GAAACCACT CGT GAT T GGGACT GGGAT T GAAACT AT T T CCGAT CAACGAGGAA T CCAAGT AAGCGGGT CAT AAGCT CCGCT T GAT T AA GGCGGGT AACCGGT GAAACCACT CGT GAT T GGGACT GGGAT T GAAACT AT T T CCGAT CAACGAGGAA T CCAAGT AAGCGGGT CAT AAGCT CCGCT T GAT T AA	1473 1621 1622 1622
*Menhaden P. richmondia T. reevesii N. erebi	GT CCGT CCGCT T T CT ACACACCGCCGCT CCGT AGT AGGAT T CCGAT GGT T T AGT CAGCT CCGT CCGAT CCGCGCGCGGGGCT CCGT CCGGGGCT CCGCGACCGCGGAG GT CCGT CCGCT T T CT ACACACCGCCGCT CCGT AGT AGGAT T CCGAT GGT T T AGT CAGCT CCGT CCGAT CCGCGCGCGGGGCT CCGT CCGGGGCT CCGCGACCGCGGAG GT CCGT CCGCT T T CT ACACACCGCCGCT CCGT AGT AGGAT T CCGAT GGT T T AGT CAGCT CCGT CCGAT CCGCGCGCGGGGCT CCGT CCGGGGCT CCGCGACCGCGGAG	1473 1731 1202 1732
*Menhaden P. richmondia T. reevesii N. erebi	AAGACGAT CAAACT T GACT AT CT AGAGGA AAGACGAT CAAACT T GACT AT CT AGAGGA AAGACGAT CAAACT T GACT AT CT AGAGGA	1473 1761 1202 1775

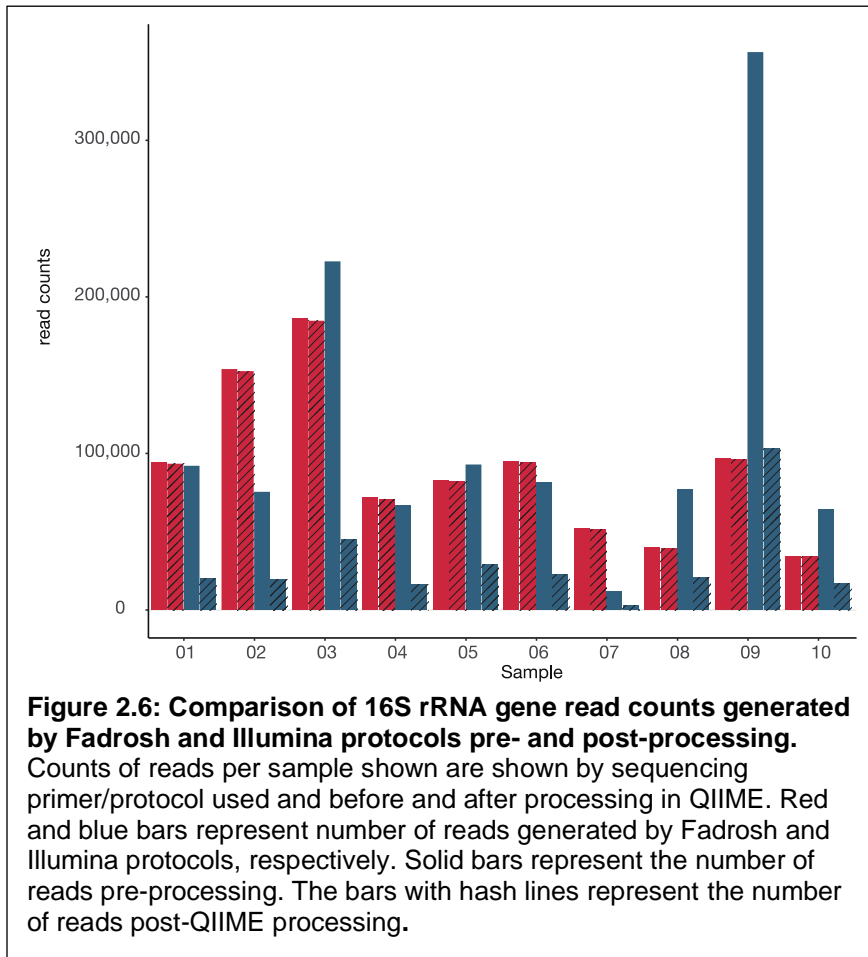
**Figure 2.4: Multiple alignment of menhaden 18S rRNA gene sequence.** The partial menhaden sequence (designated with an asterisk (\*)) (Accession number, MN335220) is shown here in a multiple alignment with 18S rRNA gene sequence from *Potamalosa richmondia*, Australian freshwater herring, accession number, KJ774739.1; *Nematalosa erebi*, Australian river gizzard shad, accession number HQ615575.1; and *Tenulosa reevesii*, accession number EU12003.1.



### 2.4.3 Comparison of results using different sequencing methods

One reason why high-throughput sequencing has become a powerful tool in microbiome studies is the identification of “universal” primers. However, for 16S and 18S rRNA gene sequences, previous studies have indicated that primer selection can influence results and bias abundance data (Meusnier *et al*, 2008; Hugerth *et al*, 2014; Pollock *et al*, 2018). In order to determine if there was introduction of community composition bias due to primer and/or sequencing protocol, all samples analyzed here were sequenced with two published sequencing protocols that use different “universal” prokaryotic primer sets. One protocol, termed “Illumina” is the Illumina-recommended sequencing protocol using primers based on the study of Klindworth *et al*, 2012). The other protocol, termed “Fadrosh”, was based on the study of Fadrosch *et al*, 2014). The Fadrosch protocol uses a dual indexing approach in which heterogeneous spacers are attached to the sequencing primers. This protocol is used to increase depth of coverage by capturing low abundance organisms. Both primer sets are based on the V3-V4 hypervariable region the *E. coli* 16S rRNA gene sequence (accession number FJ950694.1). Both primer sets give PCR products of approximately 500 bp in length. **Figure 2.5** shows the main differences between the two primer sets is that the Illumina forward primer is shorter and more degenerate in comparison to Fadrosch forward primer and the Fadrosch reverse primer is shorter and more degenerate than the Illumina reverse primer. Both forward and reverse primers have heterogeneous spacers following the 12 bp index sequence (not shown).





We retrieved a total of 1,141,141 and 906,483 raw reads using the Illumina and Fadrosh sequencing methods, respectively (Table 2.3). After post-sequencing processing to obtain quality reads, these numbers were reduced to 296,734 and

898,66 for the Illumina and Fadrosh protocols, respectively. The number of reads varied between samples, with a wide range between the highest and lowest number of reads (Table 2.3 and Figure 2.6). However, post-processing, the Fadrosh sequencing method consistently gave the higher number of reads. Using the Illumina sequencing method, sample 09 produced 356,332, the maximum number of raw reads and sample 07 produced the least number of raw reads, at 11,705. The Fadrosh sequencing method did not produce such wide differences in number of reads per sample. The most striking difference between the two methods is the comparison of the number of raw and processed reads as shown in Figure 2.6. Overall, there are far fewer raw reads lost with processing using the Fadrosh sequencing method indicating a greater percentage of high-quality reads. Looking at the average number of reads the percent difference pre-

Sample ID	Raw reads		Post processing	
	Fadrosh et. al	Illumina	Fadrosh et. al	Illumina
01	94,491	92,125	93,434	20,104
02	153,942	75,483	152,598	19,518
03	186,302	222,707	184,780	45,105
04	71,620	66,824	70,883	16,038
05	82,712	92,906	82,070	29,232
06	94,578	81,730	94,055	22,861
07	51,793	11,705	51,398	2,690
08	39,854	76,835	39,428	20,840
09	96,854	356,332	95,936	103,153
10	34,337	64,494	34,080	17,193
Total reads	906,483	1,141,141	898,662	296,734
Average reads	90,648	114,114	89,866	29,673

**Table 2.3: Effect of different gene sequencing protocols on the number of raw reads and post-processing reads.** DNA was extracted from the stomach contents of 10 juvenile menhaden. A region of approximately 469 bp encompassing the V3-V4 hypervariable regions of 16S rRNA genes was targeted for sequencing using the sequencing protocols described by Fadrosh *et al* (Fadrosh *et al*, 2014) and those recommended by Illumina (Klindworth *et al*, 2012) using the Fadrosh and Illumina sequencing protocols. The raw and post-processing reads from each were expressed as reads per sample, mean number of reads per sample and total reads.

and post-processing was 74 % and 0.86 % for the Illumina and Fadrosh sequencing method, respectively. Post-processing, sample 07 from the Illumina reads had the lowest number of reads with

2,690 reads. This was excluded in our downstream analysis because it did not meet the minimum rarefaction requirements. The difference in the number of raw reads generated by the two protocols probably reflects the stringency of hybridization of the two primer sets to rRNA sequences. The Illumina primers were smaller and had higher degeneracy and would have hybridized less stringently in early rounds of amplification giving a higher number of products that were not of sufficient quality. The Fadrosh primer set had less degeneracy and also had the spacer sequences that allowed better amplification of rare or more diverse 16S rRNA sequences after the initial rounds of amplification and gave reads of better quality.

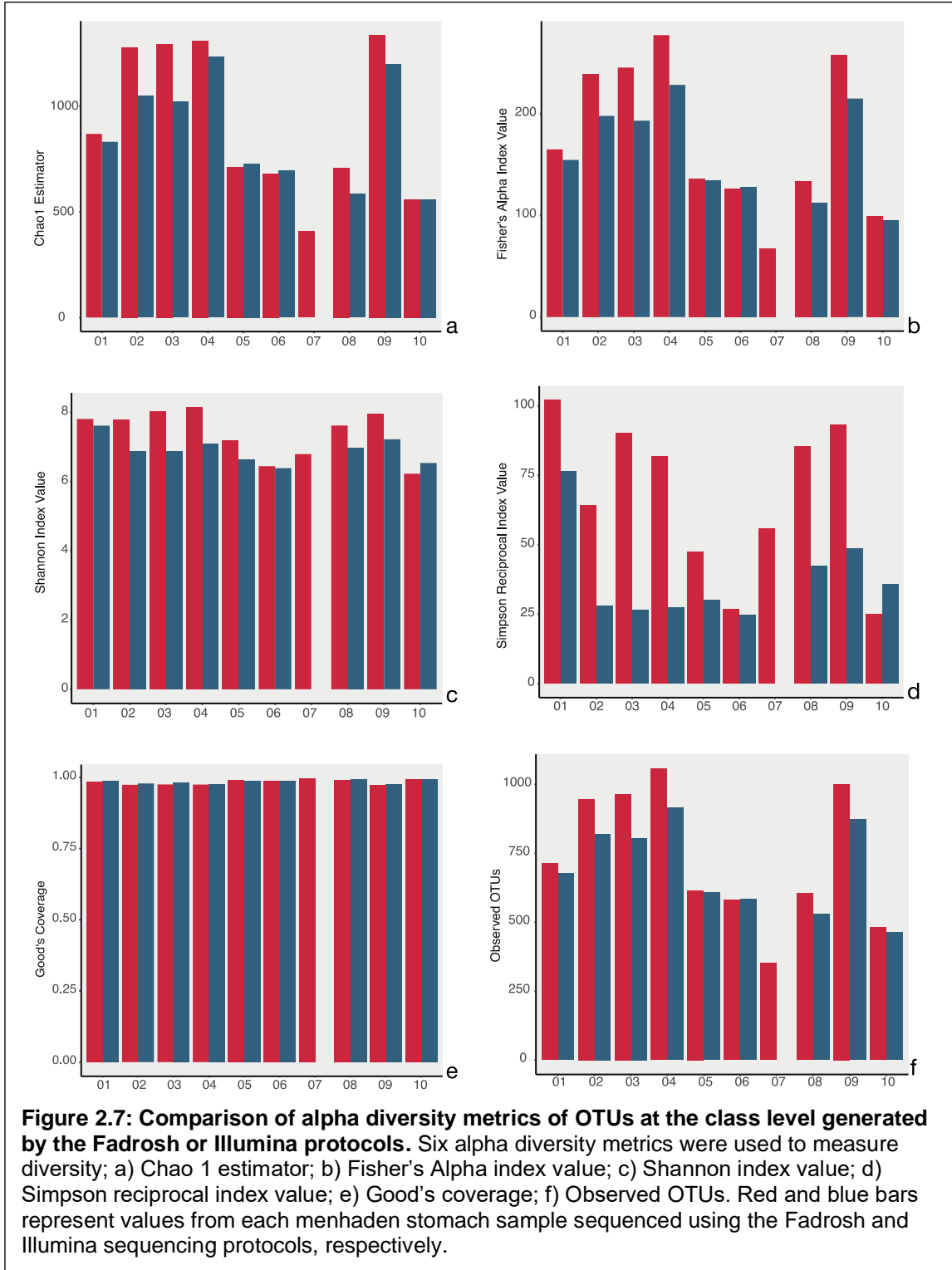
To provide an understanding of the apparent differences in the microbiota suggested from each sequencing method, we looked at several commonly used alpha diversity metrics using OTU abundancies (**Figure 2.7**). We also tested the null hypothesis of differences between sequencing methods using Mann-Whitney

statistical test (**Table 2.4**). Observed OTUs (species), Chao1 estimator, Fisher’s alpha ( $\alpha$ ) indexes are commonly used to measure richness simply defined as the number of different species in a sample. Observed OTUs is the simplest measurement of species richness; it counts the number of different OTUs in a sample. In all but sample 06, Observed OTUs were greater using the Fadrosh sequencing protocol but showed no significant difference between sequencing method ( $p = 0.661$ ). In sample 06, the number of Observed OTUs was higher using the Illumina sequencing protocol, although the difference is very small. Chao1 diversity metric is another estimator of species richness; however, it makes the assumption that all possible species are often not accounted for because of sequencing depth in a sample. The metric attempts to account for this by providing an estimate of “true” species richness. The Chao1 shows a similar result as Observed OTUs where this measure is greater in all but 3 samples (05, 06 and 10) using the Fadrosh sequencing method. However, it does not show a significant difference between the two sequencing methods ( $p = 0.78$ ). Fisher’s ( $\alpha$ ) is used as a sample-size independent estimator to address the issue of OTU richness bias due to sample size (sequencing depth). Similar to

Alpha diversity metric	Mean Fadrosh	Mean Illumina	mannwhitney p-value
Observed OTUs	731.6000	696.8889	0.6607
Chao1	915.6579	878.2551	0.7802
Goods coverage	0.9839836	0.9846812	0.8702
Fisher alpha	175.1648	162.2811	0.6607
Simpson reciprocal	67.16076	37.75702	0.03499
Shannon	7.401079	6.914679	0.1128

**Table 2.4: Mean alpha diversity metrics from 16S rDNA gene sequences.** The mean alpha diversity metric value from sequences derived using the Fadrosh or Illumina sequencing protocols are given, looking at: Observed OTUs; Chao1 estimator; Good’s coverage; Fisher’s alpha index; Simpson reciprocal index; Shannon index; along with the p-value using Mann-Whitney statistical tests.

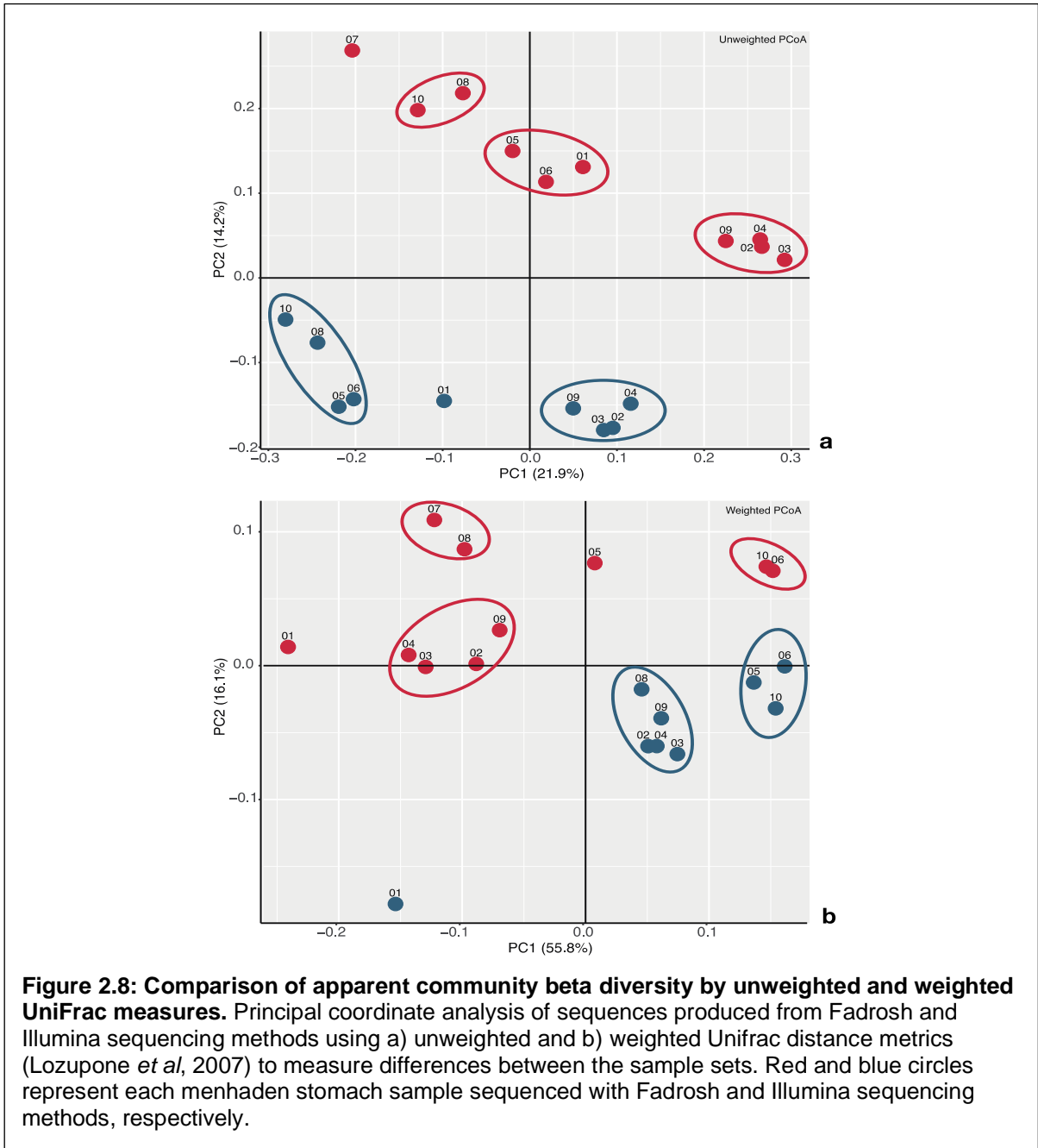
Observed OTUs, Fisher's ( $\alpha$ ) was greater using the Fadrosch sequencing protocol but the difference was not significant ( $p = 0.661$ ).



In sample 06, Fishers ( $\alpha$ ) was greater using the Illumina sequencing method, however the difference is very small. The metric Good's coverage, a method of estimating what percent of the total number of species is represented in a sample, only varied slightly per sample and sequencing protocol and showed no significant difference between the two sequencing protocols ( $p = 0.870$ ). The Shannon index and Simpson reciprocal are diversity metrics that measure richness but take into account the relative abundance or evenness of each group. However, they do so in different ways. The Shannon index gives an equal weight to richness and evenness, thus as richness and evenness increase, so does the Shannon index. The Simpson's index calculates the probability that any two reads randomly sampled from a community will belong to the same taxonomic assignment, thus is more heavily influenced by dominant taxa. Typically, these two metrics compliment well and will follow the same trend. Both Shannon and Simpson reciprocal show that all samples have a higher diversity index using the Fadrosh sequencing method except for sample 10. The difference is shown to be significant only in the Simpson reciprocal index ( $p = 0.035$ ). Taken together the diversity metrics show that the two protocols gave sufficient sampling depth and are similar in species richness. However, the significant difference in the Simpson reciprocal indicates the presence of dominant taxa in samples sequenced with the Illumina sequencing method that reduce the evenness. A possible reason for this is that the samples sequenced using the Illumina sequencing method are dominated by chloroplast reads in comparison to the samples obtained using the Fadrosh sequencing protocol, probably reflecting the lower stringency of the Illumina primers.

In order to see the apparent differences of the bacterial communities between each sample and the apparent differences found between the two sequencing protocols, we used weighted and unweighted Unifrac distance metrics visualized by principal coordinate analysis (PCoA) plots (Lozupone *et al*, 2007) (**Figure 2.8**). Unweighted Unifrac analysis is a qualitative measure. In contrast, weighted Unifrac analysis reveals quantitative community differences due to relative taxon abundance, or in this case apparent differences in relative

taxon abundance. The weighted PCoA plots show a clear separation of samples based on Illumina and Fadrosh sequencing methods. The sample set sequenced with the Illumina sequencing protocol form two clusters with 95 % confidence, with sample 01 being an outlier. In contrast, the sample set sequenced with Fadrosh sequencing method is more scattered but forms three clusters with 95 % confidence, with samples 01 and 05 being outliers. Samples 06 and 10 sequenced by each sequencing method vary only along PC2 (y-axis) in the weighted PCoA plot, indicating that these show the least change between the two sequencing protocols.



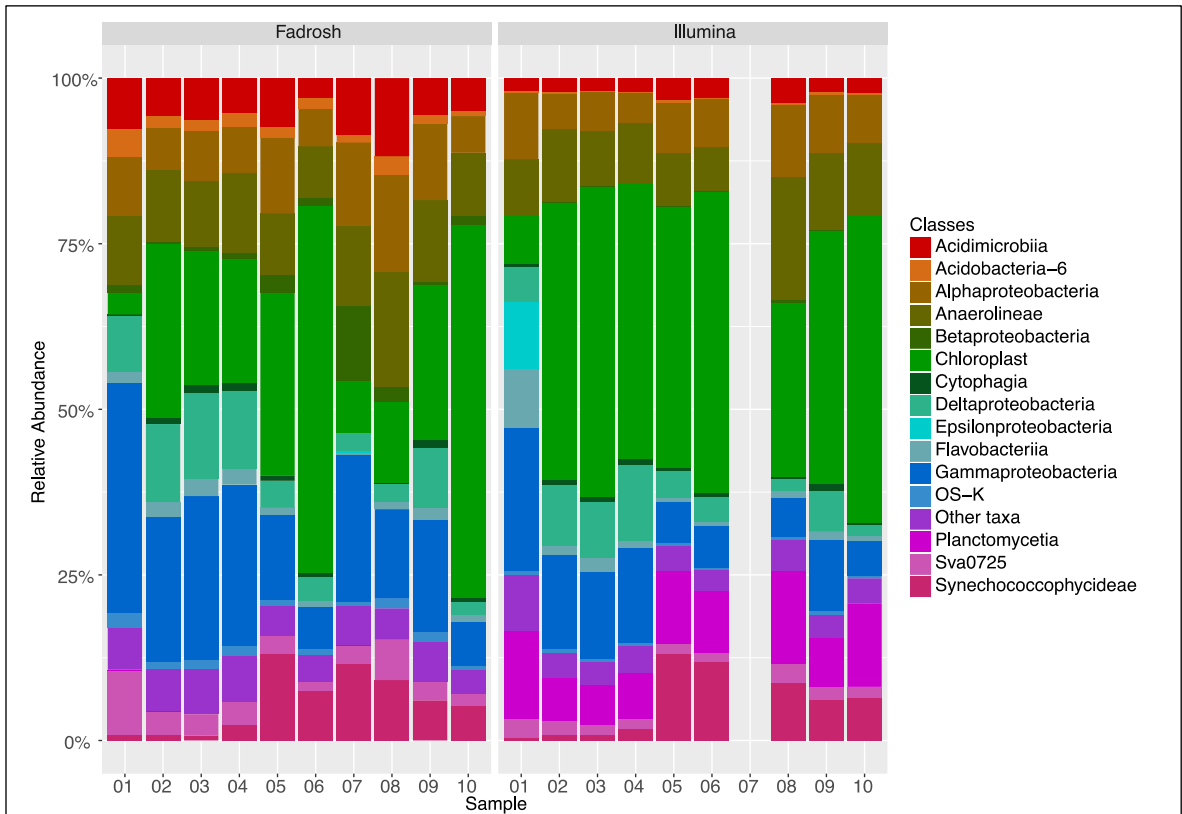
The unweighted PCoA plots also show a clear separation of samples based on sequencing protocol. Again, the Illumina sequencing protocol shows two distinct clusters, with sample 01 again being an outlier. Samples group in the same cluster for both weighted and unweighted analyses, except for sample 08. Overall, it is clear that the sequencing methods do affect overall community



results. However, based on the scale of the axes however, the differences are probably too small to impact our estimate of overall community composition. Overall the microbial community, using either set of sequencing primers, shows the same microbial taxa, although the relative abundances of those taxa differ depending on which sequencing primers are used.

#### *2.4.4 Variation in taxonomic composition in each sample at the class level using the two different primers/sequencing protocols*

The most dominant taxa at the class level consistently across all samples regardless of sequencing primers/method used was from chloroplast 16S rRNA sequences and thus are not representative of prokaryotes (**Figure 2.9**). These sequences accounted for 25.1 % and 37.0 % of the 16S rRNA gene sequences using Fadrosh and Illumina sequencing protocols respectively. Following chloroplast, the next dominant groups were Gammaproteobacteria and Anaerolineae with relative abundance averaging approximately 10.0 % to 18.0 % of the bacteria community respectively. Though these were on average the next dominant groups, there was variability between samples. For example, in sample 05 only, Synechococcophyceae were more dominant than Anaerolineae (**Figure 2.9**). Using MetaStats to test for significant differences between the sequencing methods, there were significant differences in taxa from only six classes; Acidimicrobia, Planctomycetia, Sva0725, Betaproteobacteria, Acidobacter-6 and OS-K (**Table 2.5**). Using MetaStats to test for significant differences between the sequencing methods, there were significant differences in taxa from only six classes; Acidimicrobia, Planctomycetia, Sva0725, Betaproteobacteria, Acidobacter-6 and OS-K (**Table 2.5**). There are decreased numbers of reads of Acidimicrobia, Sva0725, Betaproteobacteria, Acidobacteria-6 and OS-K using the Illumina sequencing method compared with the Fadrosh sequencing method. Only Planctomycetia showed more reads using the Illumina sequencing protocol.



**Figure 2.9: Comparison of the relative abundance of OTUs at the class level derived using Fadrosh and Illumina protocols for 16S rRNA gene sequencing.** The relative abundance of OTUs at the class level from individual fish, caught at different St. Vincent Sound sites, when sequenced using primers for 16S rRNA gene sequences along with the Fadrosh and Illumina sequencing method.

Class	Mean (Fadrosh)	Mean (Illumina)	p-value	q-value	Effect (F:I)
Acidimicrobia	0.066	0.024	<0.001	0.002	decrease
Planctomycetia	0.001	0.097	<0.001	0.002	increase
Sva0725	0.038	0.019	0.008	0.020	decrease
Betaproteobacteria	0.022	0.001	<0.001	0.002	decrease
Acidobacter-6	0.020	0.003	<0.001	0.002	decrease
OS-K	0.012	0.004	<0.001	0.002	decrease

**Table 2.5: Significant differences in mean 16S rRNA reads at the class level using Fadrosch versus Illumina sequencing protocols.** Differences in 16S rRNA reads at the class level using Fadrosch versus Illumina sequencing protocols, along with p-values and q-values for each and showing whether using the Fadrosch protocol gives and increase or decrease in the number of reads belonging to an OTU.

#### 2.4.4 Community assessment of stomach content using universal eukaryotic primers

Given the success of assessing the prokaryotic microbiota using universal prokaryotic primers, we applied this method using universal eukaryotic primers to gain an understanding of prey items. We used metabarcoding targeting a portion of the V4-V5 hypervariable regions of 18S rRNA gene sequences as developed by Hugerth *et al.* (Hugerth *et al.*, 2014).

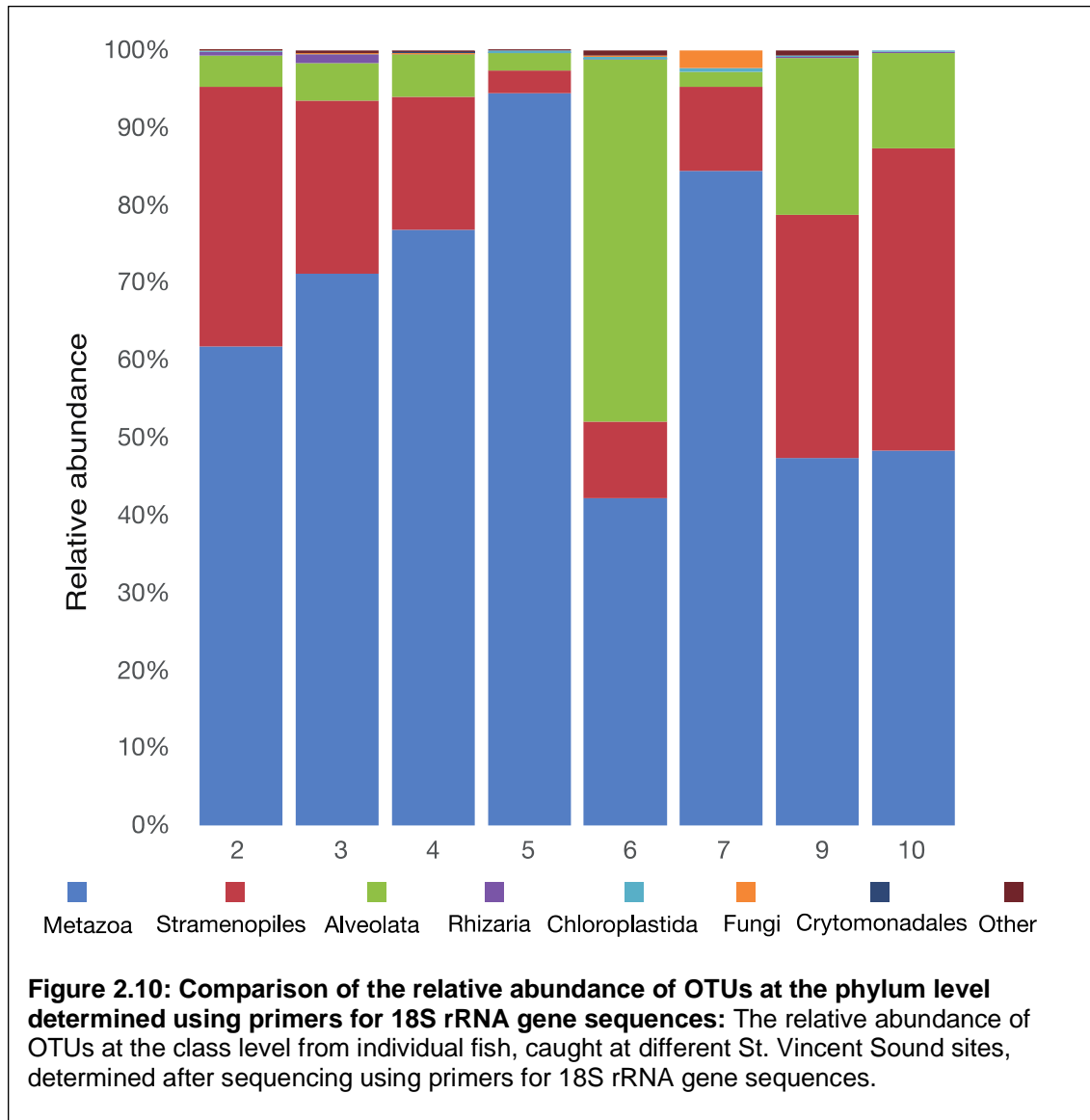
Sample ID	Total OTUs
02	6797
03	6508
04	6217
05	5201
06	10573
07	1573
09	4369
10	544

Sample ID	Raw reads	Post processing
02	54,085	53,989
03	84,779	84,679
04	76,896	76,757
05	15,450	15,423
06	44,303	44,254
07	39,251	39,191
09	30,788	30,743
10	4,671	4,661
Total reads	350,223	349,697
Average reads	43,777.9	43,712.1

**Table 2.6: Number of 18S rRNA reads and OTUs per sample.** a. 18S OTUs per sample: Number of 18S OTUs per sample using 97 % similarity of the SILVA 111 eukaryote only sequence database. OTUs matching menhaden were removed before analysis. b. Comparison of 18S rRNA reads generated by each sample: Number of reads per sample before and after processing pipeline using QIIME.

The total number of raw reads using 18S rRNA gene primers was 350,223, with 349,697 reads generated post processing (Table 2.6). Looking at the 18S rRNA gene

sequencing results, we see that the majority of the reads that were assigned to chloroplasts by the 16S rRNA gene sequencing analysis, belong to the groups Alveolata and Stramenopiles. These are the dominant protists, consisting mainly of ciliates, dinoflagellates and diatoms (**Figure 2.10**). Overall, 18S rRNA gene sequences are dominated by metazoan sequences. In each sample, between 57 % to 97 % of the 18S rRNA gene sequences are assigned to Metazoa (**Figure 2.10**). The four most dominant metazoan phyla represented are Arthropoda, Craniata, Mollusca, and Annelida respectively. The reads from the Arthropoda consisted almost entirely of the copepod, *Acartia tonsa*. The relative abundance of *Acartia* ranged from 12.7 to 86.9 % of metazoan species found. The remainder



of the 18S rRNA reads from Crustacea are from several genera of barnacles, the nauplius and cyprid larvae of which should be retained by juvenile menhaden gill rakers. The reads assigned to the phylum Mollusca are dominated by two organisms. One is a gastropod of the genus *Deroceras*, an air breathing land slug from permanently wet habitats. *D. laeve* is widespread in the Gulf of Mexico coastal region.

Both eggs and adults of *D. laeve* can survive when submerged; the adult is 25-35 mm; eggs 1.7-2 mm. The other mollusk is a bivalve of the rock boring genus *Leiosolenus* which is also found in the Gulf of Mexico with *L. aristata*, dispersed through ballast water. The most unexpected metazoan representatives were from the Craniata. This group included several organisms commonly found in the freshwater aquarium pet trade, such as zebrafish and spotted gar. Given that the adults from these unexpected taxa are inconsistent with the size of menhaden gill rakers, we conclude that their presence may represent captured earlier life stages, degraded cellular material, or eDNA.

#### 2.4.5 Validation of the presence of taxa found by high throughput sequencing with end-point PCR using group specific primers

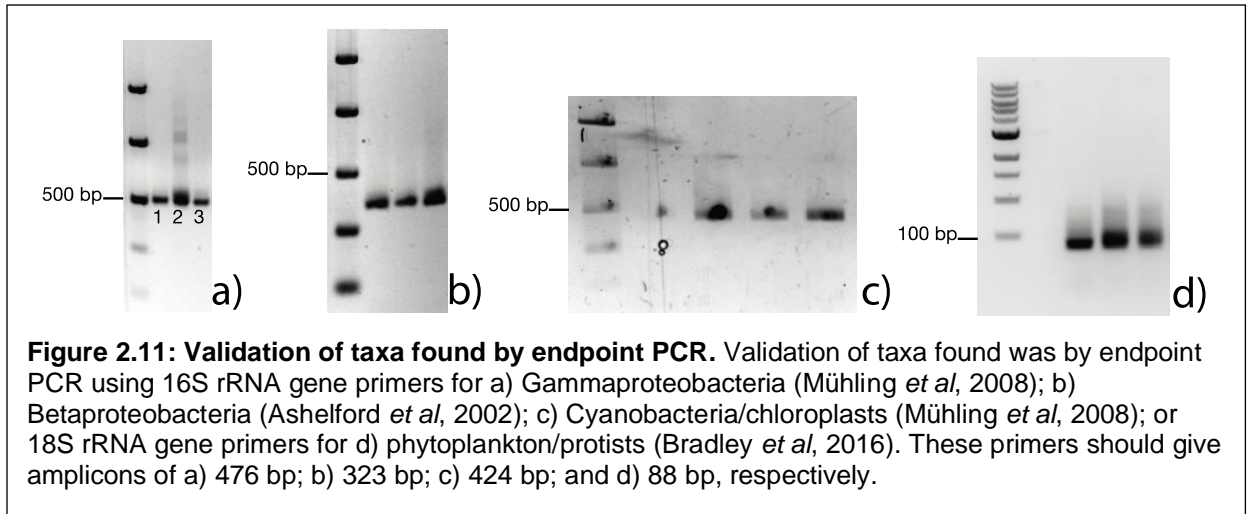
In order to validate the method, we targeted several taxa by amplification using their published taxon specific primers for 16S or 18S rRNA genes

(Ashelford *et al*, 2002; Mühling *et al*, 2008; Bradley *et al*, 2016) (**Table 2.7**)

Primer	Target group	Sequence (5' - 3')	Reference
Beta359f Beta682r	Betaproteobacteria	GGGGAATTTTGGACAATGGG ACGCATTTCACTGCTACACG	Ashelford et al 2002
Gamma395f Gamma871r	Gammaproteobacteria	CMATGCCGCGTGTGTGAA ACTCCCCAGGCGGTCDACTTA	Muhling et al 2008
CYA361f CYA785r	Cyanobacteria, chloroplasts	GGAATTTTCCGCAATGGG GACTACWGGGGTATCTAATCC	Muhling et al 2008
V8f 1510r	microalgae / phytoplankton	ATAACAGGTCTGTGATGCCCT CCTTCYGCAGGTTACCTAC	Bradley et al 2016

**Table 2.7: Taxon-specific primers for validation of sequence assignments.** Published rRNA gene primers for amplifying the 16S and 18S rRNA genes of different taxa were taken from Ashelford *et al* (Ashelford *et al*, 2002) for Betaproteobacteria; Mühling *et al* (Mühling *et al*, 2008), for Gamma proteobacteria and Cyanobacteria/chloroplasts; and Bradley *et al* (Bradley *et al*, 2016) for microalgae/phytoplankton.

For bacteria, primers designed to amplify 16S rRNA gene sequences from Gammaproteobacteria were chosen due to the relatively high abundance of Gammaproteobacteria amongst the samples. Primers to amplify 16S rRNA gene sequences of Betaproteobacteria was chosen to represent a taxon that was not as widely abundant. For eukaryotic taxa, primers that are designed to amplify 18S rRNA gene sequences of phytoplankton were chosen, which also includes a wide variety of protists. The presence of the above taxa was considered to be confirmed if the relevant primers produced an amplicon of expected product length by end-point PCR. Amplicons corresponding to the expected sizes were seen in all cases (**Figure 2.11**).



## 2.5 Discussion

### 2.5.1 Strengths of DNA metabarcoding

DNA metabarcoding, whether for the analysis of microbiota or diet items allows high taxonomic resolution and enables the simultaneous analysis of many samples. The general workflow is well established; extraction of total DNA from the dietary sample, PCR amplification of DNA barcode markers from taxa of interest and then DNA sequencing for taxonomic classification of the recovered sequences. A clearer picture of bacterial diversity will be seen if chloroplast sequences are removed prior to analysis.

Applying DNA techniques to diet identification has recently increased identification resolution, particularly in marine systems (Blankenship & Yayanos, 2005; Riemann *et al*, 2010; Cleary *et al*, 2012; Jakubavičiūtė *et al*, 2017; Waraniak *et al*, 2019). DNA metabarcoding enables the identification of most prey items, even when diets are broad and diverse, as well as the simultaneous analysis of many samples. This work has identified trophic linkages within food webs, as well as predator diet breadth and preference. One of the few investigations examining the potential of using DNA analyses of fish gut contents in the monitoring of ecosystem function is the study of the stomach contents of coral reef fish (Leray *et al*, 2013).

Visual identification has conventionally been used for the analysis of gut contents of fish based on prey morphology (Hyslop, 1980). The method is time consuming and requires taxonomic expertise. However, most studies based on visual inspection, particularly in small filter feeding fish have the following disadvantages: ambiguous prey specimen identification due to extensive digestion, the presence of unidentified partial tissues, and a lack of expert knowledge of identification (at least higher than family or order level) (Baker *et al*, 2014). Stable isotope analysis can allow a sense of trophic level but can only inform of relative proportion of phytoplankton and zooplankton. Because of this, PCR methods for gut-contents analysis have developed rapidly and they now dominate the diagnostic methods used for gut-contents analysis in field-based research (Pompanon *et al*, 2012).

### 2.5.2 *Limitations of DNA metabarcoding*

Problems with DNA barcoding using universal primers do exist. Because it is PCR amplification based, problems of contamination can occur. This includes field-based contamination from nonfood environmental DNA, laboratory contamination (De Barba *et al*, 2014). In addition, misassignment of sequence-to-sample during high throughput sequencing (Schnell *et al*, 2015).

Analysis by 18S rRNA gene metabarcoding is only semi-quantitative producing relative read abundancies (Deagle *et al*, 2019). Although it does not provide quantitative estimates of prey items, it can provide a guide for future targeted study design. The choice of primers can make a big difference; even closely related primer sets can affect apparent representation of some taxa, as shown in this study, even though in this study no differences overall in what dominant OTUs was found. Furthermore, there is no perfect set of universal primers for either 16S or 18S rRNA gene sequences. Even now, 18S rRNA gene sequences are not well represented across the eukaryotic tree of life, particularly for aquatic species (Weigand *et al*, 2019).



### 2.5.3 *Other considerations*

DNA purification methods will continue to improve to allow for amplification of difficult samples. In addition, remedies against PCR inhibitors have been developed. For example, Quantabio (Beverly, MA, US) has developed an engineered Taq DNA polymerase that is combined with high avidity monoclonal antibodies. These antibodies bind the polymerase and keep it inactive prior to the initial PCR denaturation step preventing binding to potential PCR inhibitors. This polymerase, in Quantabio's ToughMix, comes with additives that neutralize PCR inhibitors like polysaccharides, humic acid, and polyphenols to ensure reliable assay performance with a spectrum of starting materials including environmental samples.

Host or predator DNA contamination is a common problem in DNA barcoding research. Because prey samples are collected from the predator's gut, there is a very high probability that predator DNA is included with the prey samples. In this study, we analyzed this directly by assessing what percentage of reads represented menhaden rRNA gene sequences and were able to eliminate samples in which host DNA contamination was too high. This required knowledge of the menhaden 18S rRNA gene sequence which we were able to determine. However, with other tissues such as gill, which is also used for diet studies, the problem of host DNA contamination can be problematic. This can be resolved by stratagems such as blocking the detection of predator DNA with ligase or using a blocking oligonucleotide (Cleary *et al*, 2012; Craig *et al*, 2013), although such methods may reduce depth of coverage.

High-throughput sequencing (HTS) technologies are now increasingly used in fisheries research and are producing ever-increasing quantities of data. While many laboratories and even undergraduate students generate high throughput sequencing data, analyzing these results requires a skill set that is traditionally reserved for bioinformaticians. Learning to program, using languages such as R and Python, and making sense of the vast amounts of available 'omics data have become easier thanks to a multitude of available resources. This can empower

bench researchers to perform more complex computational analyses. Although not used here, KNIME is an example of an accessible entry point for researchers daunted by programming. It is a graphical user interface (GUI) analytics environment that offers a 'point and click' alternative to classical programming (Berthold *et al*, 2009; Fillbrunn *et al*, 2017). These GUI programs are becoming more popular as applications for metabarcoding expand. Since the beginning of this work QIIME has been superseded by QIIME2 which includes a semi-GUI application (Caporaso *et al*, 2010, Bolyen *et al*, 2019). MOTHUR is another example of a semi-GUI HTS analysis platform (Schloss *et al*, 2009). These tools such as KNIME, QIIME2 and MOTHER, together with a growing number of tutorials and courses, have been crucial in providing simple user interfaces to conduct complex analyses, making big data accessible to biologists (Grabowski & Rappsilber, 2019). However, these platforms offer less flexibility for pipeline development compared with programming languages such as R and Python. Furthermore, they can still require hours of set-up and use.

## 2.6 Comments and recommendations

Assessment of menhaden stomach contents is technically challenging because the visible food items are small (5-100  $\mu\text{m}$ ) and menhaden have a gizzard-like stomach that grinds ingested items to an amorphous paste (Friedland *et al*, 1984). The methods reported here shows that DNA metabarcoding can be applied successfully to the study of the stomach microbiota and diet items of *Brevoortia patronus*, the Gulf menhaden. The method allowed characterization of the stomach prokaryotic community, as well as eukaryotic diet items. Neither the diet not the stomach microbiota of Gulf menhaden has been analyzed previously by DNA metabarcoding. The analysis of 18S rRNA sequences has uncovered a greater taxonomic richness than previously described with OTUs (species) per sample ranging from 1500 to over 10,000. The methods described are suitable also for the study of Atlantic menhaden, *Brevoortia tyrannus*. The method can be applied to studying the effects of location, salinity, season and developmental stage on the diet and

stomach microbiota of menhaden and can be applied to investigating the microbiota of different regions of the gut. It could also be used to monitor Gulf menhaden stomach contents over time to tease out effects of climate change. We also anticipate applying the method to other filter feeding species. Increased data management resources and the reduction in cost for high throughput sequencing make DNA metabarcoding an attractive alternative to traditional methods. If routinely included in the investigation of ecosystem function, DNA metabarcoding has the potential to complement other approaches and ultimately enhance ecosystem-based management and biomonitoring (Taberlet *et al*, 2012; Evans *et al*, 2016; Deiner *et al*, 2017; Bohan *et al*, 2017).

### 2.7 Acknowledgments

This work was supported by the National Oceanic and Atmospheric Administration, Educational Partnership Program (NOAA-EPP) award to the Living Marine Resources Cooperative Science Center (LMRCSC) NA11SEC4810002 and NA16SEC4810007. Ammar Hanif was supported in part by MD SeaGrant Graduate Student Fellowship, NA10OAR4170072 SA7528129-D and in part as an LMRCSC Graduate Fellow. Thanks to Sabeena Nazar for her technical expertise in optimizing the high throughput sequencing. Thanks to Dr. Stacy Smith, Delaware State University for coordinating the collection of menhaden, to LMRCSC summer undergraduate intern Malisa Smith and to Centennial High School intern, Eric Sibanda for assistance in determining the menhaden 18S rRNA gene sequence.

## Supplementary Materials

Sample	Site	Salinity ppt	Temp °C	pH	DO	Secchi
S1	SVS02	37.6	28.4	8.0	5.5	0.6
S2	SVS02	37.6	28.4	8.0	5.5	0.6
S3	SVS02	37.6	28.4	8.0	5.5	0.6
S4	SVS03	36.3	27.8	7.9	5.5	0.6
S5	SVS04	36.4	28.7	8.0	5.8	0.4
S6	SVS04	36.4	28.7	8.0	5.8	0.4
S7	SVS04	36.4	28.7	8.0	5.8	0.4
S8	SVS05	36	29	7.8	5.7	0.4
S9	SVS05	36	29	7.8	5.7	0.4
S10	SVS05	36	29	7.8	5.7	0.4

**Supplemental Table 2.1: Characteristics of water at collection sites in Apalachicola Bay**

Water quality measurements of temperature, salinity, pH, dissolved oxygen, and turbidity are recorded at the time and place of sample collection using a YSI 556 multiparameter water quality meter.

## Sequence data

Sequence data for this project can be found at

<https://github.com/Hanif82/gulfmenhadenotutable.git>.

## Chapter 3: The stomach microbiota of juvenile Gulf menhaden (*Brevoortia patronus*) differs with location in Apalachicola Bay, FL

### 3.1 Abstract

As part of a study investigating the ecology of juveniles of the Gulf menhaden (*Brevoortia patronus*) from Apalachicola Bay, FL, the microbiota of the stomach was characterized. As juveniles, Gulf menhaden live in tidal creeks, marsh and open bay areas. In the present study we compared the stomach microbiota of juvenile Gulf menhaden at two different locations within Apalachicola Bay, FL, using DNA metabarcoding. Juvenile menhaden from the same year class were collected in Apalachicola Bay from Two-mile Channel and St. Vincent Sound representing different salinities. MiSeq Illumina high-throughput sequencing was used to analyze DNA amplicons from the V3-V4 region of the 16S ribosomal RNA (rRNA) gene from the stomachs of menhaden at each site. 14,184 unique operational taxonomic units (OTUs) were identified, 14,145 of which were found in samples from both locations. The stomach microbiota of samples from both locations showed a predominance of Proteobacteria, Chloroflexi, Bacteroidetes, Acidobacteria and Actinobacteria, although significant differences in composition at the class level were seen. Ninety-six OTUs were present in all samples from both locations, representing 32.45 % of total reads. Alpha diversity metrics showed that samples from Two-mile Channel showed a higher level of taxonomic richness with thirty-seven OTUs unique to these samples. Beta diversity analysis showed a strong association between the microbiota in menhaden stomachs and sampling location, possibly reflecting differences in water quality. Proteobacteria represent 51 % and 49 % of the total reads in samples from Two-mile Channel and Vincent Sound, respectively. Within the most abundant of these were families with the potential to contribute to menhaden nutrition by the provision of B vitamins or enzymes that could aid in digestion of cellulose and chitin and degrade detrital sulfated polysaccharides. Within the Proteobacteria, only five

OTUs could be assigned at the genus level; *Anaerospira*, *Desulfosarcina*, *Desulfobacca*, *Shewanella* and *Halomonas*. Such a consortium has been linked with the processes of anaerobic hydrocarbon degradation, sulfate reduction, denitrification and/or methanogenesis associated with petroleum biodegradation.

### 3.2 Introduction

The microbiota of fish is among the better characterized among the non-mammalian vertebrates (Colston & Jackson, 2016). Earlier work on the microbial communities of fish has been largely culture based, although, as with studies of the microbiota of other host taxa, culture-independent studies have increased more recently and the patterns of fish microbiota has been reviewed (Clements *et al*, 2014). The stomach is not often included for gut microbial composition analyses in fish. However, fish have a unique and close interaction with their surrounding environment as well as with the microorganisms that co-exist there compared to tetrapods. This is particularly true for filter-feeders that travel with open mouths taking in material in the same proportion as they occur in the surrounding water although the size of the gill rakers can provide some selectivity. Effectively, menhaden can concentrate organisms from the water in which they travel. In the present study we compared the stomach microbiota of juvenile Gulf menhaden (*Brevoortia patronus*) using DNA metabarcoding with primers for the DNA encoding the 16S rRNA gene sequence. DNA metabarcoding has rapidly become the method of choice in characterizing living communities in any environment. This approach provides a comprehensive culture-independent approach that has been used increasingly to obtain a full inventory of the gut microbiota from multiple fish species (Jami *et al*, 2015; Tarnecki *et al*, 2017; Egerton *et al*, 2018). Because of the intimate association of fish with their surrounding environment, we compared the stomach microbiota of Gulf menhaden at two different locations within Apalachicola Bay, Florida, Two Mile Channel (TMC) and St. Vincent's Sound (SVS), representing very different salinities.

*B. patronus* is the main menhaden species in the Gulf of Mexico, ranging from the northern Gulf from Brazos Santiago, Texas, to Tampa Bay, Florida

(Christmas & Gunter, 1960). Young-of-the-year Gulf menhaden are ubiquitous members of the northern Gulf of Mexico estuarine nekton communities and occupy fresh to brackish waters using both marsh and open bay habitats (Deegan *et al* , 1990). After transformation, juveniles remain in low salinity, near-shore areas where they travel about in dense schools, often near the surface (Lassuy, 1983). As juveniles, menhaden live in tidal creeks, marsh and open bay areas where they filter the water column via their gill rakers. The migration pattern of juvenile Gulf menhaden involves the sequential use of marsh creek and open bay areas where they remain until late summer or fall (Deegan *et al*, 1990). This migration pattern coincides with the productivity peaks in tidal creek and open bay areas. Menhaden are omnivorous filter feeders, feeding by straining phytoplankton and zooplankton from water. Their movement as juveniles is related to the availability of food in the water.

The Apalachicola Bay estuary is a highly productive lagoon and barrier island complex on the upper Gulf coast of Florida. It covers about 212 square miles and serves as the interface between the river system and the Gulf of Mexico. Four barrier islands bound the bay: St. Vincent Island, St. George Island, Little St. George Island, and Dog Island. The bay area, including Apalachicola Bay, East Bay, St. George Sound, St. Vincent Sound, Indian Lagoon, and Alligator Harbor, is about 65 km long and 5 to 10 km wide. Apalachicola Bay is a river-dominated system with the major source of freshwater input coming from the Apalachicola River. The alluvial Apalachicola River is the largest Florida river in terms of flow (NWFWM report 2017). Maximum river flows occur during late winter to early spring months and are highly correlated with Georgia rainfalls (Meeter *et al*, 1979). The bulk of seawater flow is from the east entering St. George Sound. The western end of Apalachicola Bay is linked to the Gulf of Mexico by Indian Pass, the narrow channel between St. Vincent Island and the mainland, with a maximum water depth of about 4 m. St. Vincent Sound itself is shallow, with an average depth of little more than 1 m, containing numerous oyster bars. Within the bay's shallow waters, with an average depth of 3 m, are numerous oyster reefs and sandy shoals. The surrounding wooded lowlands consist of saltwater

and freshwater marshes, and freshwater swamps. Two-Mile Channel follows the coastline from the west side of the Apalachicola River estuary for approximately 2 miles.

Given that menhaden sample the environment through filtering, it seemed likely that the stomach microbiota would reflect the microbial communities in the water of the two locations although their contributions to menhaden health or as indicators of environmental quality was unknown. The present study looks at the diversity of prokaryotic microorganisms within menhaden juvenile stomachs with regard to location and water quality, their capability to provide nutritional benefits, as well as the potential of menhaden to function as a biomonitors for coastal and near-shore bays.

### *3.3 Methods*

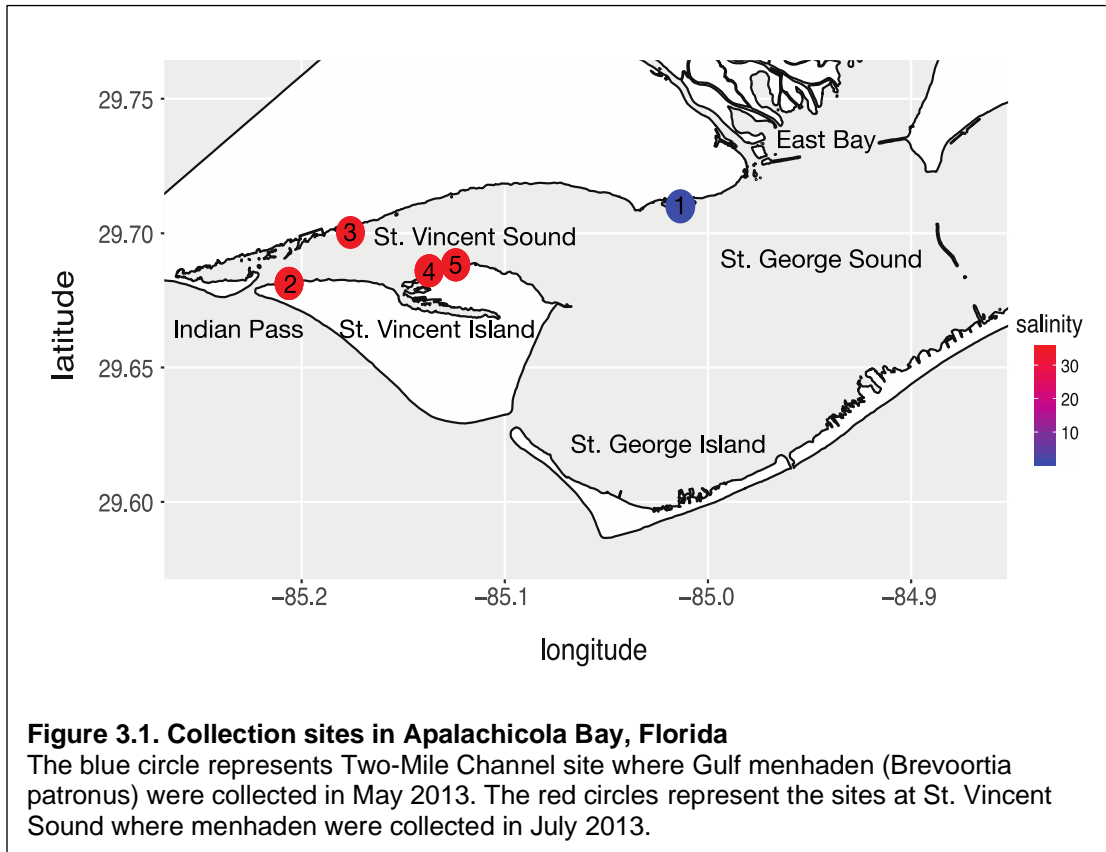
Detailed methodology for DNA metabarcoding using universal primers for 16S rRNA genes has been provided in Chapter 2.

#### *3.3.1 Sample collection*

Colleagues from Delaware State University (DSU) collected samples in May 2013 at Two-Mile Channel (TMC) in Apalachicola Bay, Florida, (TMC01, Lat:29.712467, Long: -85.01525) (**Figure 3.1**). Gulf menhaden samples were collected from this location using a 10-foot cast net. At this time of year and location, the menhaden were concentrated in large, slow-moving, and tightly packed schools in lower salinity water, allowing all fish to be collected at the same site within this location. A second collection was made by Florida Fish and Wildlife in July 2013 from the St Vincent Sound location (SVS) in Apalachicola Bay, Florida. The choice of location was determined by where the menhaden were found. At this time, 07/02/2013, the menhaden were not found in tightly packed schools or in the lower salinity waters of TMC. Because the menhaden were not so tightly grouped, sampling took place at four sites within SVS (SVS02, SVS03, SVS04, SVS05), with latitude and longitude coordinates corresponding to 29.6813, -85.206167 (SVS02); 29.700267, -85.17595 (SVS03); 29.686033, -85.1372 (SVS04) and 29.688233, -85.124167 (SVS05), respectively (**Figure**



**3.1).** Collections at SVS sites were done using a 183 m seine net at 1 m depth. Water quality measurements of temperature, salinity, pH, dissolved oxygen, and turbidity were also recorded at all sites (**Table 3.1**).



Location	Sample	Site	Salinity ppt	Temp °C	pH	DO	Secchi
Two Mile Channel	T1	TMC01	1.8	27.1	7.5	5.7	0.7
	T2	TMC01	1.8	27.1	7.5	5.7	0.7
	T3	TMC01	1.8	27.1	7.5	5.7	0.7
	T4	TMC01	1.8	27.1	7.5	5.7	0.7
	T5	TMC01	1.8	27.1	7.5	5.7	0.7
	T6	TMC01	1.8	27.1	7.5	5.7	0.7
	T7	TMC01	1.8	27.1	7.5	5.7	0.7
	T8	TMC01	1.8	27.1	7.5	5.7	0.7
	T9	TMC01	1.8	27.1	7.5	5.7	0.7
	T10	TMC01	1.8	27.1	7.5	5.7	0.7
St. Vincent Sound	S1	SVS02	37.6	28.4	8.0	5.5	0.6
	S2	SVS02	37.6	28.4	8.0	5.5	0.6
	S3	SVS02	37.6	28.4	8.0	5.5	0.6
	S4	SVS03	36.3	27.8	7.9	5.5	0.6
	S5	SVS04	36.4	28.7	8.0	5.8	0.4
	S6	SVS04	36.4	28.7	8.0	5.8	0.4
	S7	SVS04	36.4	28.7	8.0	5.8	0.4
	S8	SVS05	36	29	7.8	5.7	0.4
	S9	SVS05	36	29	7.8	5.7	0.4
	S10	SVS05	36	29	7.8	5.7	0.4

**Table 3.1: Water quality measurements from each site**

Water quality measurements of temperature, salinity, pH, dissolved oxygen, and turbidity were recorded at the time and place of sample collection using a YSI 556 multiparameter water quality meter. The sites and samples are color coded to be consistent with the colors used in Figures 3.6 and 3.7.

### 3.3.2 Sample preparation, DNA extraction, and estimation of DNA quality

Sample preparation, DNA extraction and estimation of DNA quality was done as described in Chapter 2.

### 3.3.3 MiSeq library preparation and high throughput sequencing

Sequencing of DNA from all twenty stomach samples was performed on the Illumina MiSeq platform located in the BioAnalytical Services Laboratory (BAS Lab) at the University of Maryland Center for Environmental Science-Institute of Marine and Environmental Technology. The library was prepared by Illumina's standard library construction protocol as detailed in the 16S metagenomic sequencing library preparation procedures available at:

[https://support.illumina.com/downloads/16s\\_metagenomic\\_sequencing\\_library\\_p](https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_p)

[reparation.html](#), with the exception that a dual indexing approach was used, as described in Fadrosch *et al* (Fadrosch *et al*, 2014; Hanif *et al*, 2020). PCR was performed using denaturation at 95 °C for 3 min, followed by 8 cycles at 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s, and a final extension step at 72 °C for 5 min. PCR clean-up was done using AMPure XP beads to purify the 16S rRNA gene V3 amplicon away from free primers and primer-dimer species. After PCR cleanup, libraries were quantified, normalized and pooled. A 2 x 300 cycle run was performed in the MiSeq, providing high-quality, paired reads of the V3-V4 16S rRNA gene.

#### 3.3.4 *Post-sequencing pipeline*

Removal of index sequences (called de-multiplexing), base calling and data quality assessment were performed on the MiSeq instrument. MacQIIME (Caporaso *et al*, 2010) was used to process and assess quality of output reads from the sequencing primers. Post sequence processing was done using the recommended QIIME pipeline for Illumina reads (Caporaso *et al*, 2010). The MacQIIME script, `join_paired_ends.py`, was used to join forward and reverse reads. Paired reads were filtered for low quality reads (quality score of <25) and short read length (<200 bp) which were removed from the library using the `split_libraries.py` command. Chimeric sequences were identified *de novo* using the USEARCH61 (v6.1.544) algorithm with the script `identify_chimeric_seqs.py` (Edgar, 2010). This was followed by the removal of PhiX sequences by a BLAST analysis with the scripts `parallel_blast.py` and `filter_fasta.py`. The remaining reads were used for operational taxonomic unit (OTU) picking using the script `pick_otus.py`. A UCLUST *de novo* clustering method within the USEARCH61 (v6.1.544) algorithm was used to pick OTUs for both sequencing methods (Edgar, 2010). Taxonomic assignment was done using the script `assign_taxonomy.py` with the Greengenes `gg_13_8` as a reference database at 90 % similarity. These data were used to construct an OTU table for subsequent relative sequence abundance analysis and diversity analysis.

### 3.3.5 Microbiota data analysis

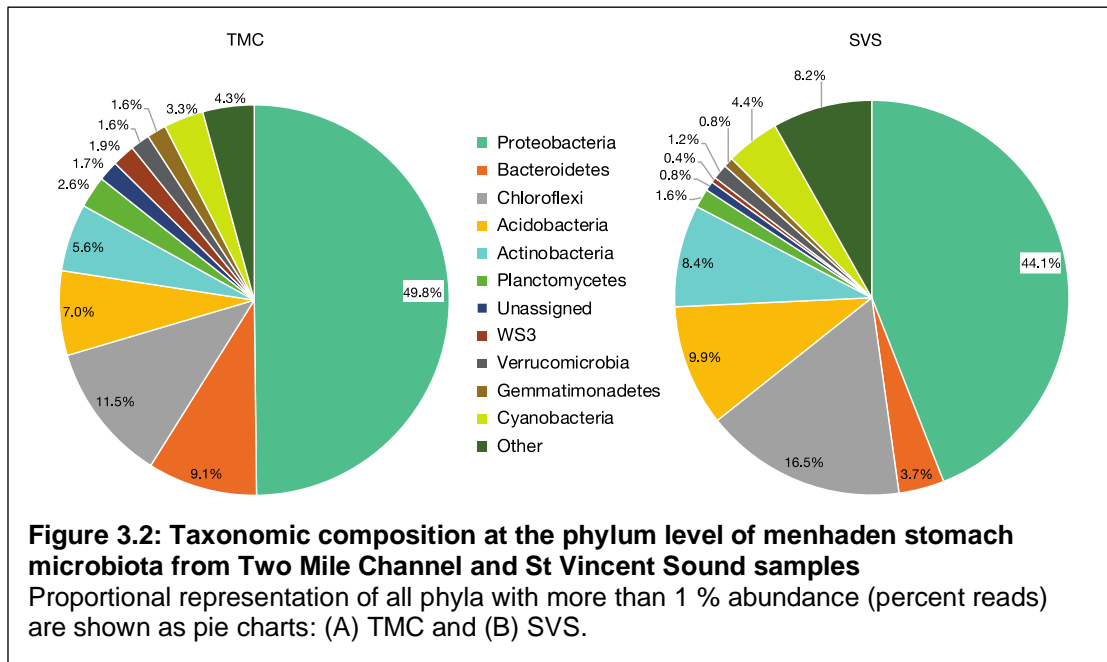
For this study, 16S rRNA sequences assigned as chloroplasts were considered protist prey and removed from this analysis. The package Phyloseq within the statistical software R and QIIME was used to assess OTU richness and evenness using the estimates of alpha-diversity metrics, Observed OTUs (species), Chao 1, Shannon and Simpson index. To analyze how closely the samples were related to each other and to compare observed differences in the microbial community from the two Apalachicola Bay sites, beta diversity analyses were determined based on Bray-Curtis distance metric and visualized using the ordination method principal coordinate analysis (PCoA) into two-dimensional plots.

## 3.4 Results

### 3.4.1 Characterization of stomach bacteria communities from menhaden caught at Two-Mile Channel and St. Vincent Sound

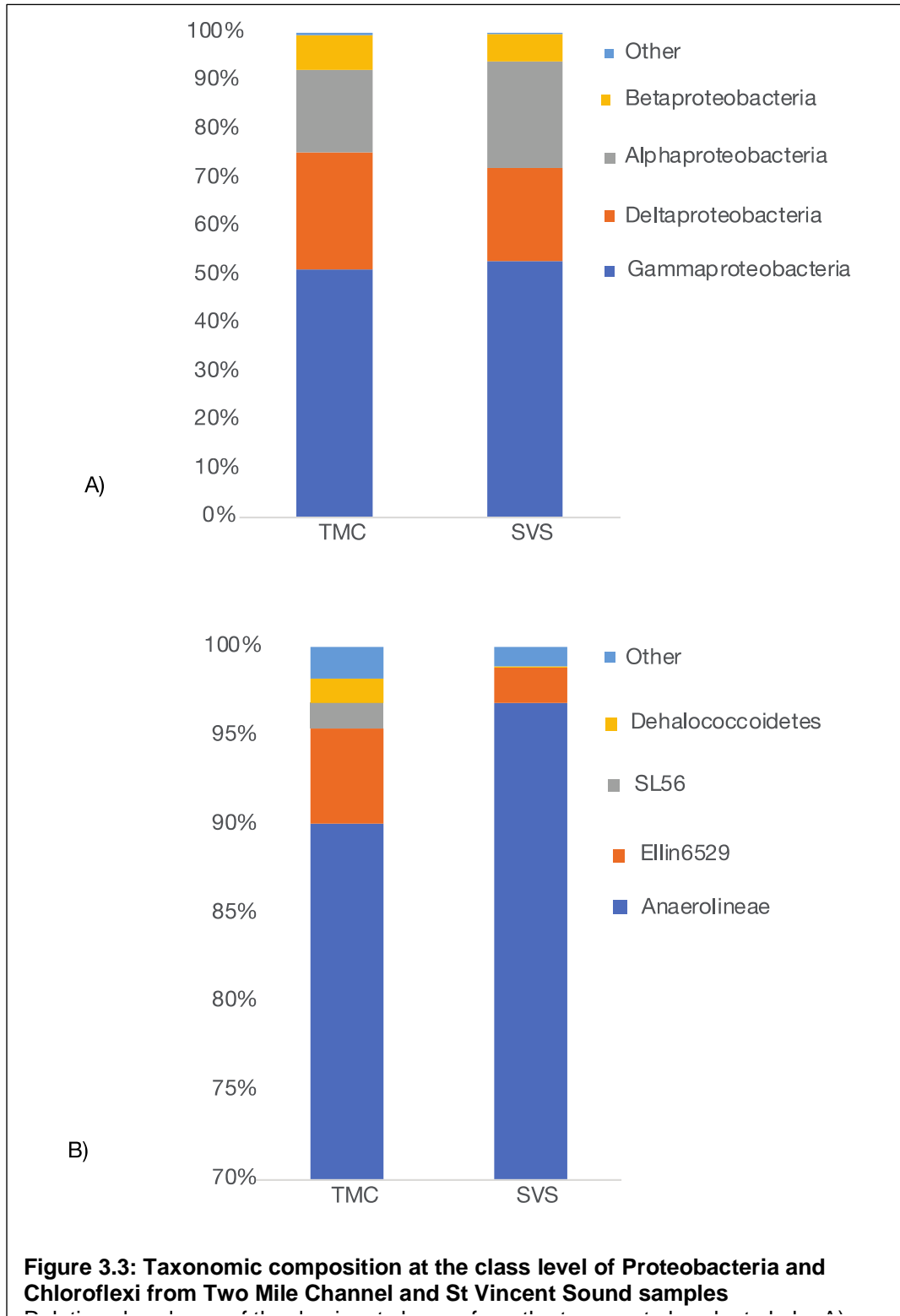
DNA extracted from twenty menhaden stomachs, ten from each sampling location, were sequenced using the Illumina MiSeq high-throughput sequencing platform. A total number of 1,420,063 and 906,483 raw reads were retrieved from TMC and SVS samples respectively. After post sequencing processing these numbers were reduced to 1,411,708 and 898,662 reads from TMC and SVS samples respectively (**Table 3.2**). Binning at 0.03 % divergence resulted in 14,184 total OTU assignments from sixty-five phyla. There were ten phyla that represented greater than 1 % of the total relative read abundance are shown in **Figure 3.2**. All were represented in stomachs from fish caught at both locations although in very different relative abundances (percent reads). The stomach microbial communities from menhaden caught at TMC were dominated (representation greater than 5 %), in descending order, by Proteobacteria, Chloroflexi, Acidobacteria, Actinobacteria and Bacteroidetes, (**Figure 3.2A**). The stomach microbial communities from menhaden caught at SVS were dominated (representation greater than 5 %), in descending order, by Proteobacteria, Chloroflexi, Acidobacteria, Actinobacteria, and Bacteroidetes

(Figure 3.2B). Proteobacteria was the most dominant taxon representing 49.8 % and 44.1 % of the relative abundance for menhaden stomach DNA from TMC and SVS samples respectively. This was followed by Chloroflexi, representing a much lower fraction at 11.5 % and 16.5 % for menhaden from TMC and SVS samples respectively. Due to their abundance I looked more deeply into these two phyla.

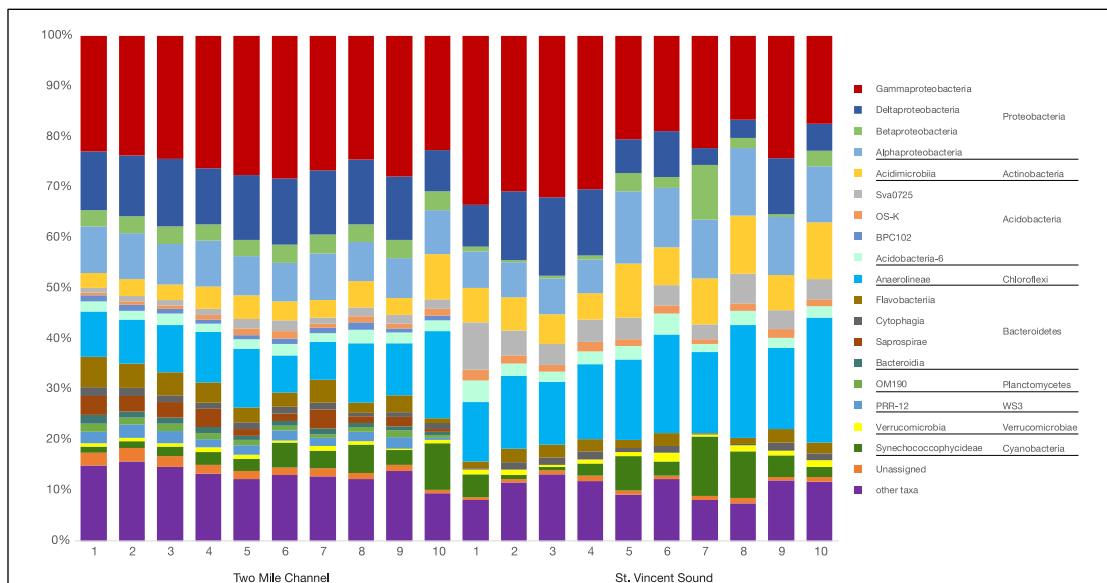


The classes represented in Proteobacteria and Chloroflexi in stomach samples from TMC and SVS can be seen in **Figure 3.3**. In samples from both locations, Proteobacteria were represented by the classes Gammaproteobacteria, Deltaproteobacteria, Alphaproteobacteria, and Betaproteobacteria with Gammaproteobacteria showing the highest representation. The next most represented classes were Deltaproteobacteria in TMC samples and Alphaproteobacteria in SVS samples. Of the four mentioned classes, Betaproteobacteria was the least represented. The four most abundant classes from the phylum Chloroflexi were Anaerolineae, Ellin6529, SL56, and Dehalcoccoidetes, with Anaerolineae having the highest representation in both TMC and SVS samples. Anaerolineae representation was greater in samples from SVS than from TMC samples. The next most represented class was

Ellin6529 in both TMC and SVS samples with a greater representation in TMC samples.



There is much variation between individual samples, even from the same location. **Figure 3.4** shows the variation in the relative abundance of classes between each sample (**Figure 3.4**). These represent classes that have a relative abundance of greater than 1 %. Classes that have a relative abundance of less than 1 % abundance were grouped in Other Taxa. To facilitate comparison between samples, the number of OTUs was normalized to 34,000 per sample. It should be noted that there are five taxa that were less than 1 % of the relative abundance in SVS samples but greater than 1 % of the relative abundance in TMC samples. Two of these were from the phylum Bacteroidetes assigned as Saprospirae and Bacteroidia. The other three were assigned as BPC102, PRR-12, and OM190 from the phyla Acidobacteria, WS3, and Planctomycetes respectively. Nine phyla were represented by eighteen classes. Proteobacteria, Acidobacteria, and Bacteroidetes were represented by four classes each. The remaining phyla were represented by a single class. Chloroflexi was the second most abundant phylum, but it was represented solely by Anaerolineae. The relative abundance other classes of Chloroflexi that are shown in **Figure 3.4** accounted for less than 1 % representation.



**Figure 3.4: Variation in taxonomic composition at the class level of stomach microbiota in each sample from Two Mile Channel and St Vincent Sound**

The relative abundance of OTUs at the class rank from each sample from TMC and SVS sites, as measured by percent reads. The number of OTUs per sample was normalized to 34,000. Classes are grouped by phyla.

Given the high relative abundance of Proteobacteria, I looked further into this phylum. **Table 3.3** shows the most abundant proteobacterial OTUs expressed as the mean number of OTUs and mean number of reads per sample. The lowest classification given to the OTUs is shown here. Not all OTUs were represented in each sample but are reflected in the mean number of OTUs. Overall, the highest representation is of the family Piscirickettsiaceae (class Gammaproteobacteria). Samples from both sampling locations have roughly the same number of OTUs representing this family although the mean number of reads were higher in SVS samples. This is followed by percent representation of the OTUs from the family Marinicellaceae (class Gammaproteobacteria), again with approximately the same number of OTUs from each location and again with higher percent representation in samples from SVS compared to those from TMC. The number of OTUs from the order Chromatiales, another gammaproteobacterial class is similar in samples from both locations although the percent representation in samples from SVS is more than two-fold higher than in samples from TMC. The number of OTUs from the family Rhodobacteraceae (class Alphaproteobacteria) is comparable in samples from both locations, however, representation in SVS samples is more than two-fold higher in comparison to samples from TMC. In TMC samples, although OTUs from the family Piscirickettsiaceae had the highest percent representation, the highest diversity of OTUs was from the order Chromatiales. For all taxa, except OTUs from the Alphaproteobacteria genus, *Anaerospira*, the highest percent representation was seen in samples from SVS. *Halomonas* spp showed the greatest difference between TMC and SVS samples with six-fold higher representation at St. Vincent Sound from a small number of OTUs, probably reflecting the differences in salinity. Though sequences from both sampling locations were dominated by Proteobacteria, the most abundant OTUs in all samples was from *Synechococcus* spp.



Class	Lowest classification	Two Mile Channel		St. Vincent Sound	
		Mean # reads	Mean # OTUs	Mean # reads	Mean # OTUs
Alphaproteobacteria	f:Rhodobacteracea	518	62.7	1282.7	71.8
	g: <i>Anaerospira</i>	65	2.6	40.1	71.8
Betaproteobacteria	f:Oxalobacteraceae	104.2	4.6	576.7	3.1
Deltaproteobacteria	g: <i>Desulfococcus</i>	376.1	44	700.9	40.5
	g: <i>Desulfosarcina</i>	29.4	3.1	60	5.1
	f:Desulfobulbaceae	377.4	47.8	700.9	31.9
Gammaproteobacteria	g: <i>Shewanella</i>	31.9	2.7	98	2.4
	o:Chromatiales	534.7	88.1	1155.7	72.3
	f:HTCC2089	67.4	11.6	149.3	9.7
	g: <i>Halomonas</i>	61.7	2.2	429	3.3
	f:Thiohalorhabdales	29.9	2.9	299.4	4.7
	f:Piscirickettsiaceae	1202	81.1	1831.4	79.9
	f:Marinicellaceae	626.2	43.4	914.4	46

**Table 3.3: Representation of the most abundant proteobacteria from Two Mile Channel and St Vincent Sound samples**

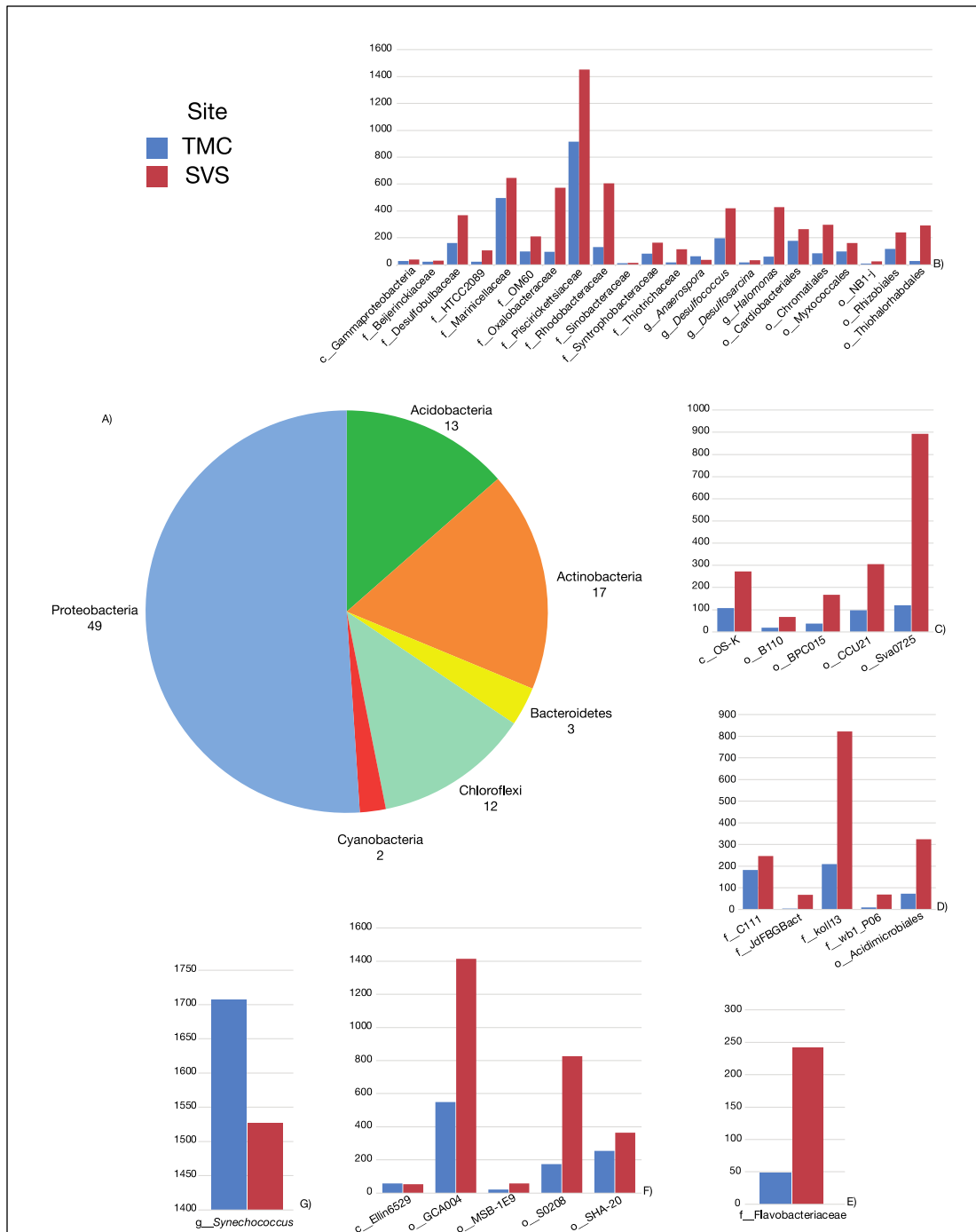
Representation of the most abundant proteobacteria by the mean number of reads and mean number of OTUs from all samples from TMC and SVS. The data from each site was normalized to 34,000 reads/sample. The lowest assigned taxonomic levels are shown.

### 3.4.2 Shared and unique OTUs

There are sixty classes of bacteria not represented in SVS samples, however only 44 of these could be attributed to a known taxonomic assignment. Most of the taxonomic assignments were given a generic identifier, for example class B5-096 of the phylum Fibrobacteres. The unidentified taxa were of low abundance (less than 1 % relative abundance) and most belonged to the phylum Planctomycetes. In contrast there were only seven classes that were not represented in TMC samples five of which could not be given a named taxonomic assignment. These were also of low abundance.

*Shared OTUs:* In order to get an idea of which bacteria could be considered candidates of a resident core microbiota in menhaden stomachs, we looked at OTUs that were present in all samples regardless of sampling site or location (**Figure 3.5** and **Table 3.4**). These shared OTUs, could indicate a common menhaden stomach microbiome for further study, or could just reflect the taxa present in the water from all sites. Ninety-six OTUs from six phyla were observed in all samples, representing 32.45 % of the total reads. The number of OTUs and their lowest classification can be seen in **Table 3.2**. The highest number of OTUs

were assigned to the phylum Proteobacteria represented by forty-nine OTUs (**Figure 3.5A** and **Table 3.4**). This was followed by the phylum Actinobacteria, Acidobacteria, Chloroflexi, Bacteroidetes and Cyanobacteria which had seventeen, thirteen, twelve, three, and two OTUs, respectively. **Figure 3.5**, panels **B-G** shows that the number of reads per OTU varied greatly between each OTU as well as between samples from the TMC and SVS locations. The number of reads from OTUs having the same lowest classification were summed. For example, Cyanobacteria had only two OTUs both of which had the lowest classification of g:*Synechococcus*. The number of reads for each OTU were summed to represent the number of OTUs for *Synechococcus*. Samples from SVS had more reads for all OTUs, samples except for those assigned to the genera *Anaerospora* and *Synechococcus*.



**Figure 3.5: Taxonomic composition of the shared OTUs from all samples**  
 Shared OTUs, defined as OTUs found in every sample at both sampling locations. A) Pie chart depicting the number of shared OTUs per phylum. B-G) Bar charts depicting read number (y-axis) of each shared OTU classified to its lowest taxonomic level from each phylum. B) Proteobacteria; C) Acidobacteria; D) Actinobacteria; E) Bacteroidetes; F) Chloroflexi; G) Cyanobacteria. Blue and red bars represent TMC and SVS samples respectively.

Phylum	# of OTUs	lowest classification	# of OTUs
Acidobacteria	13	o__Sva0725	3
		c__OS-K	3
		o__BPC015	3
		o__CCU21	3
		o__B110	1
Actinobacteria	17	o__Acidimicrobiales	2
		f__C111	2
		f__koll13	10
		f__wb1_P06	1
		o__Acidimicrobiales	1
Bacteroidetes	3	f__JdFBGBact	1
		f__Flavobacteriaceae	3
Chloroflexi	12	o__GCA004	5
		o__S0208	3
		o__SHA-20	2
		c__Ellin6529	1
		o__MSB-1E9	1
Cyanobacteria	2	g__Synechococcus	2
Proteobacteria	49	f__Piscirickettsiaceae	6
		f__Marinicellaceae	5
		f__Oxalobacteraceae	1
		o__Cardiobacteriales	1
		o__Rhizobiales	1
		f__Desulfobulbaceae	2
		o__Chromatiales	5
		f__Rhodobacteraceae	6
		g__Halomonas	2
		o__Myxococcales	1
		f__Syntrophobacteraceae	1
		g__Desulfococcus	2
		o__Thiohalorhabdales	2
		f__OM60	3
		g__Desulfococcus	1
		f__Thiotrichaceae	1
		f__HTCC2089	2
		c__Gammaproteobacteria	1
		f__Hyphomicrobiaceae	1
		g__Anaerospora	1
		f__Beijerinckiaceae	1
g__Desulfosarcina	1		
o__NB1-j	1		
f__Sinobacteraceae	1		

**Table 3.4: Taxonomic composition of the shared OTUs from all samples**

The taxonomic composition of the shared OTUS are shown from the phylum level to the lowest assigned taxonomic levels. The number of OTUs are given for each of the lowest assigned taxonomic levels.

The greatest taxonomic resolution of shared taxa was seen at the order level with twenty-four orders in total being represented. The highest diversity belonged to Gammaproteobacteria with nine known orders represented. This was followed by Deltaproteobacteria, Alphaproteobacteria, and Betaproteobacteria with, five, two and one order(s) being represented, respectively. The order with the highest representation was Thiotrichales from the Gammaproteobacteria (**Table 3.4**). At the family level this order was represented by Piscirickettsiaceae and Thiotrichaceae, with most reads coming from the former. The order Desulfobacterales, from the Deltaproteobacteria, contained the second largest number of reads, represented by the families Desulfobacteraceae and Desulfobulbaceae (genus *Desulfococcus*), with the most reads being assigned to the former of the two. The order with the highest representation in the phylum Acidobacteria was Sva0725. All OTUs from the Actinobacteria were from order Acidimicrobiales, with the highest representation from the family koll13. From the phylum Bacteroidetes, all shared OTUs were from the family Flavobacteriaceae. The orders GCA004 and SO208 from the phylum Chloroflexi had the highest number of reads. Despite Proteobacteria having the greatest number of both OTUs and shared OTUs, the shared OTUs with the greatest number of reads from Cyanobacteria, the two *Synechoccus spp.*

*Unique OTUs:* In terms of unique taxa, there are thirty-six OTUs represented solely in TMC samples and one OTU solely represented in SVS samples (**Table 3.5**). Unique taxa are defined here as taxa represented in all samples from one sampling location with no representation in the other sampling location. Amongst the OTUs unique to TMC samples, there were eight belonging to the phylum Cyanobacteria, all of which were assigned at the genus level as either *Synechococcus* or *Pseudanabaena*. There were six OTUs assigned to the phylum Bacteroidetes and further classified to the families Saprospiraceae, Flavobacteriaceae, and Cryomorphaceae and one OTU assigned only to the order Flavobacteriales. The phylum Proteobacteria represented the greatest number of unique OTUs at thirteen. These were further classified as g:*Rhodobacter*; o:Myxococcales; f:Acetobacteraceae; o:Rickettsiales;

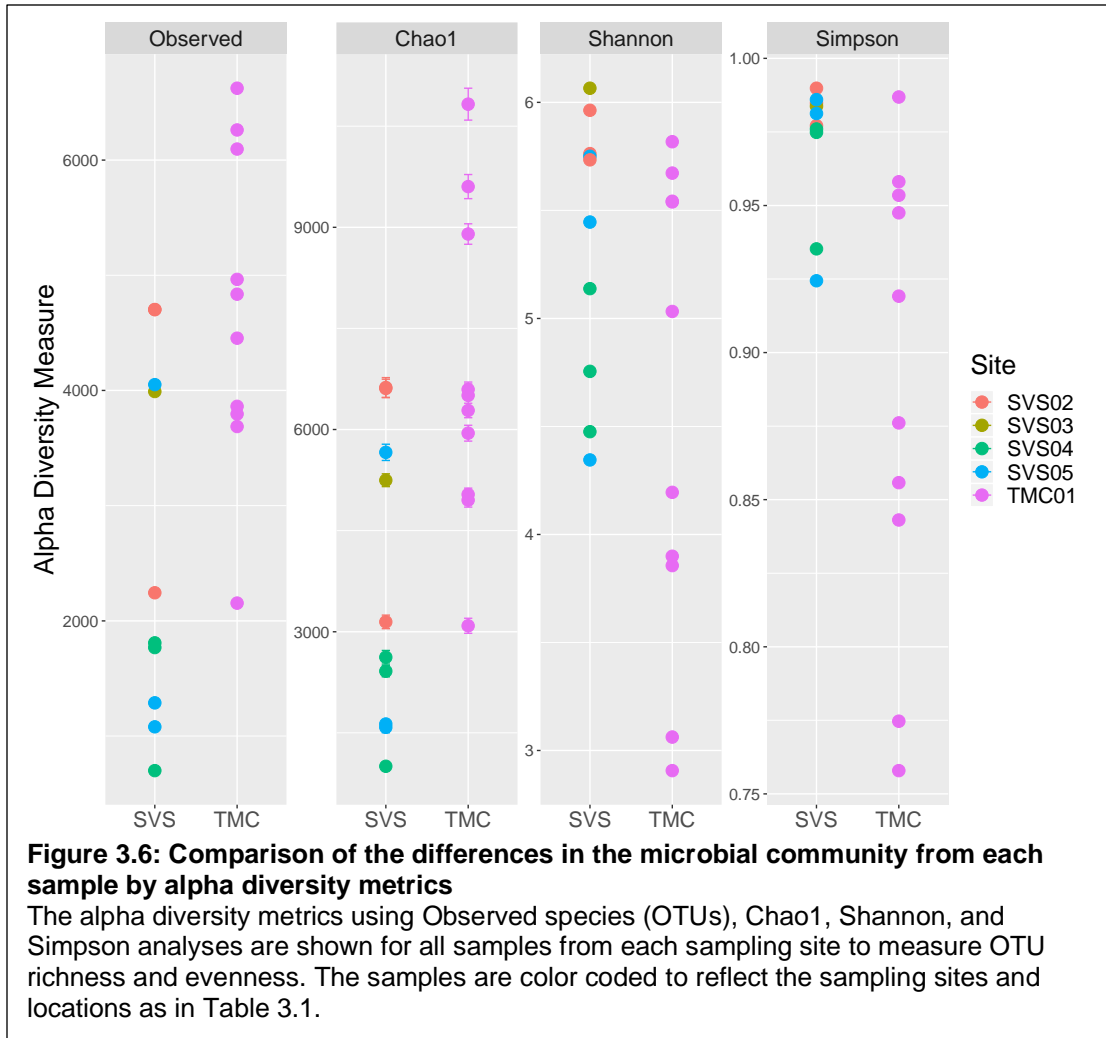
f:Xanthomonadaceae; and f:Alcaligenaceae, all of which were only represented by a single OTU, as well as f:Sinobacteraceae and f:Comamonadaceae which were represented by two and three OTUs respectively. The phylum Actinobacteria and Planctomycetes were represented by two and three OTUs respectively. f:C111 and g:*Candidatus* represented Actinobacteria. f:Gemmataceae; o:Phycisphaerales; and o:CL500-15 represented Planctomycetes. The remaining phyla Acidobacteria, Chloroflexi, Gemmatimonadetes, and WS3 were each represented by a single OTU, classified as c:AT-s54; f:A4b; c:Gemm-2; and o:Sediment, respectively. The single OTU unique to SVS samples was from the phylum Cyanobacteria with the lowest classification as g:*Synechococcus*. Overall, the most abundant of the unique OTUs were from the genus *Synechococcus*. However, unique OTUs accounted for less than 1 % of the relative representation.

Phylum	Lowest Classification	TMC % representation	SVS % representation
Cyanobacteria	<i>g:Pseudanabaena</i>	0.262	
	<i>g:Pseudanabaena</i>	0.018	
	<i>g:Pseudanabaena</i>	0.017	
	<i>g:Synechococcus</i>	0.011	
	<i>g:Synechococcus</i>	0.010	
	<i>g:Synechococcus</i>	0.008	
	<i>g:Synechococcus</i>	0.007	
	<i>g:Synechococcus</i>	0.005	
	<b><i>g:Synechococcus</i></b>		
Bacteroidetes	f:Saprospiraceae	0.047	
	f:Saprospiraceae	0.019	
	f:Flavobacterioceae	0.011	
	f:Cryomorphaceae	0.011	
	f:Cryomorphaceae	0.010	
	o:Flavobacteriales	0.005	
Chloroflexi	F:A4b	0.047	
Proteobacteria	o:Myxococcales	0.025	
	<i>g:Rhodobacter</i>	0.031	
	f:Acetobacteraceae	0.023	
	o:Rickettsiales	0.019	
	f:Comamonadaceae	0.014	
	f:Xanthomonadaceae	0.014	
	f:Sinobacteraceae	0.012	
	f:Alcaligenaceae	0.009	
	o:PHOS-HD298108	0.009	
	f:Aeromonadaceae	0.008	
	f:Comamondaceae	0.008	
	f:Sinobacteraceae	0.005	
	f:Comamondaceae	0.004	
Planctomycetes	o:Phycisphaerales	0.029	
	f:Gemmataceae	0.008	
	o:CL500-15	0.005	
Gemmatimonadetes	c:Gemm-2	0.023	
WS3	o-Sediment-1	0.016	
Actinobacteria	f:C111	0.015	
	<i>g:Candidatus</i>	0.009	
Unassigned	-	0.009	
Acidobacteria	c:AT-s54	0.005	

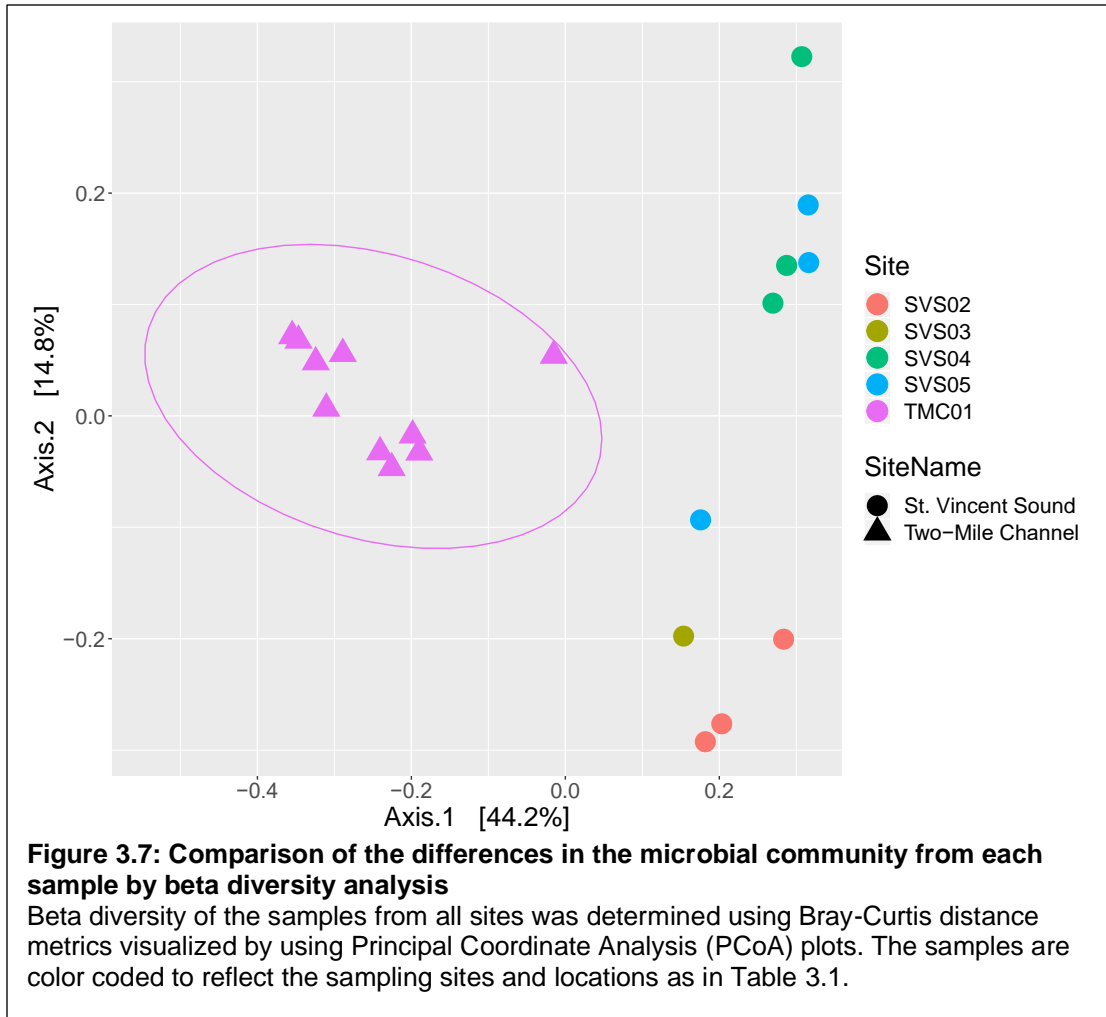
### 3.4.3 *Alpha and beta diversity analysis*

*Samples from TMC show higher diversity:* Alpha diversity was investigated using Observed OTUs, Chao 1, Shannon and Simpson indices (**Figure 3.6**). The combined samples from TMC on average have more unique OTUs in comparison to the combined samples from SVS based on both Observed species and Chao1 estimates, both common measures of species richness. This indicates that the combined TMC samples have more unique OTUs in comparison to SVS samples, in most cases double the number of unique OTUs. Both Shannon and Simpson diversity metrics look at species richness and evenness but places more emphasis on richness and evenness respectively. However, both diversity metrics indicate higher diversity for SVS samples. Furthermore, TMC samples have a greater range of values for both diversity metrics. This may indicate that the microbial communities of SVS samples more closely resemble each other in comparison to TMC samples. In samples from both locations, there are three samples that appear to have a higher number of rare or unique OTUs. This could indicate that such OTUs may provide adaptability to environmental changes, since higher numbers of rare OTUs often indicate a redundancy in the environmental microbial processes. Within the TMC samples, there is higher diversity between samples, even though all samples from this location were from the same site. In contrast, the SVS samples seem to have a similar diversity to each other. Furthermore, the relative abundance of the OTUs in the SVS samples are more like each other than OTUs in the TMC samples.





*Beta diversity analyses shows differences in samples from TMC and SVS: A hierarchical clustering of samples based on Bray-Curtis analysis shows cluster*



relationships between samples, indicating differences in bacterial composition between samples from Two Mile Channel and St. Vincent Sound (**Figure 3.7**). PCoA plots shows a clear distinction between TMC and SVS samples that correlates with the differences in salinity at the two sites. One sample from site SVS05 and one sample from site SVS04 more closely resemble each other than other samples from the same sites. It should be noted that water from the four SVS sites differ from each other with respect to turbidity and dissolved oxygen (**Table 3.1**). The samples from SV04 and SV05 that come from water with higher dissolved oxygen and lower turbidity cluster away from the samples at SV02 and SV03.

### 3.5 Discussion

For fish in general, it remains unclear how diversity of the microbiota is partitioned among niches, either between stomach and lower gut or at other body sites like gill. While a great diversity of gut microorganisms has been uncovered across fish species, most communities have been dominated by the Proteobacteria (Roeselers *et al*, 2011; Sullam *et al*, 2012; Givens *et al*, 2015; Ghanbari *et al*, 2015; Song *et al*, 2016). Consistent with this, in Gulf menhaden from the TMC and SVS locations in Apalachicola Bay, Proteobacteria had the highest representation in terms of total reads, shared OTUs and unique OTUs in samples. Although samples from both sampling locations were dominated by Proteobacteria, the most abundant OTUs in all samples was from *Synechococcus* spp. Comparison of 16S rRNA gene sequences in Gulf menhaden from TMC and SVS suggest the stomach microbiota to be dynamic, correlating with water characteristics at the different sites. Unfortunately, although water physical characteristics are known, water samples at each site were not taken for comparison. However, in general the stomach microbiota is expected to change with location, reflecting the different physical conditions and therefore microbial community of the surrounding water. Salinity can affect the microbial community of water (Egerton *et al*, 2018). Similarly, turbidity and dissolved oxygen can also influence the microbial community. Peck, as early as 1893, hypothesized that menhaden stomach contents reflected the plankton in the surface waters they occupied (Peck, 1893). However, although differences in composition of the microbial community at the class level from the two locations reflected differences in water characteristics, a predominance of the same bacterial phyla; Proteobacteria, Chloroflexi, Bacteroidetes, Acidobacteria and Actinobacteria was seen at each site in each location. Ninety-six OTUs were present in all samples, representing twenty-four orders and comprising 32.45 % of the total reads. These are candidates for a core menhaden stomach microbiome.

Whether or not some of the stomach microbiota constitute a needed core microbiome, their role in the menhaden should be considered and their possible roles in the digestion of food or as food items themselves should be investigated. Conclusive evidence for a metabolically active gastric microbiota is difficult to provide from 16S rRNA gene sequencing-based microbiota analysis alone. However, consideration of highly represented and shared taxa should be deliberated in the context of the menhaden feeding apparatus.

In the similarly sized allopatric species, Atlantic menhaden (*B. tyrannus*), the plankton filtering capability of juveniles was calculated from the experimental determination of clearing rates of uniformly sized particles. For 14-cm fork-length fish, a size representing the upper limit of juvenile size ranges, the fish can filter uniformly sized particles in a range of 7–9  $\mu\text{m}$  in diameter at filtration efficiencies of approximately 10 % (Friedland *et al*, 1984). Based on this, the effective pore size of the gill rakers of menhaden juveniles typical of spring and summer nursery residents (7.5-cm fork-length fish) is 60 % of this, or 4.2-5.4  $\mu\text{m}$  diameter. However, for fish in the wild, particle retention at these sizes is enhanced when a background of other particles, typical of a water column are present, such as detritus and other plankton particles. Detritus has certainly been observed in the esophagus of juveniles (Durbin & Durbin, 1975). Coupled with this, scanning electron microscopy and histology revealed complex hydrodynamical arrangements that utilize both the structure of the gill rakers and the fluid dynamics of mucous to trap, translocate, and consume extremely small food particles (Friedland *et al*, 2006). From this, it was suggested that juvenile menhaden retain organisms smaller than their effective filtering capability, and it has been anticipated that juveniles can retain picoplankton including bacterioplankton. Epifluorescent microscopy of 0.2  $\mu\text{m}$ -filtered material showed that autofluorescing coccoid cyanobacterial cells, 1-2  $\mu\text{m}$  in diameter, probably *Synechococcus* spp, could be seen in the esophagus consistent with the findings here that *Synechococcus* spp are found in all stomach samples (Friedland *et al*, 2006). However, the same cells could also be seen intact and fluorescing in the

region immediately before the anal vent suggesting they can survive gut passage and so are unlikely to have been digested and used as food.

There is a range of opinions on the importance of detritus as a food source. In juvenile Atlantic menhaden in estuarine creeks up to 80 % of the stomach content of has been reported to be amorphous matter/detritus and up to 50 % in adults in coastal waters. It has been argued that phytoplankton and zooplankton production are not sufficient to support observed levels of juvenile menhaden populations and that marsh detritus be an important food source (Peters & Schaaf, 1981; Lewis & Peters, 1994). This was confirmed for Gulf menhaden using a combination of  $\delta^{13}\text{C}$ ,  $\delta^{15}\text{N}$ , and  $\delta^{34}\text{S}$  stable isotope analysis to trace the flow of organic matter (Deegan *et al*, 1990). These studies showed that juvenile Gulf menhaden caught near salt marshes derived approximately 30 % of their nutrition from the detritus. If this is the case, the stomach microbiota could play a role in utilization of such a food source. The OTUs from Rhodobacteraceae in menhaden juvenile stomachs could be contributing to this. There are seven OTUs shared in all samples from the Rhodobacteraceae and OTUs from this family are among the most abundant reads with 1.5 % representation in samples from TMC and 3.8 % in samples from SVS. Marine Rhodobacteraceae encode arylsulphatase genes required for cleaving sulphate from breakdown products of the polysaccharide fucoidan, a component of diatom extracellular matrix and the means by which diatoms can bind to detrital matter (Wustman *et al*, 1998). The arylsulfatase activity in stomach Rhodobacteraceae has the potential to make sulfated polysaccharides in the amorphous material available for digestion. Furthermore, the high abundance of Rhodobacteraceae in the stomach microbiota of juvenile menhaden could reflect their association with macroalgae or detrital particles. Fucoidan desulphonation is also important in the Eastern oyster (*Crassostrea virginica*). The ability to use sulfated polysaccharides is thought to come from sulfohydrolases from Planctomycete OTUs in the stomach microbiota that help make sulfated polysaccharides available for digestion (King *et al*, 2012). Marine Rhodobacteraceae are also major vitamin suppliers for B<sub>12</sub>-auxotrophic prokaryotes and eukaryotic primary producers, such as

chlorophytes, diatoms, dinoflagellates, coccolithophores and brown algae and in this capacity, could also play a role in menhaden nutrition. Flavobacteriales and Rhodobacter encode chitinases which could assist in the digestion of copepods and aeromonads which have cellulase systems could assist in the digestion of microalgal cell walls (Munoz *et al*, 2014).

*Halomonas* species are also among the most abundant Proteobacteria in samples from both TMC and SVS and are among the shared OTUs. *Halomonas* spp have been found in a broad variety of saline environments, including estuaries, the ocean, and saline lakes and grow over the range of NaCl from 5 to 25 %. *Halomonas* spp have the capacity to oxidize sucrose, glycerol and hydrolyze the disaccharide, cellobiose, as well as producing halotolerant alkaline protease, giving them a potential role in digestion.

Where oil is naturally present, for example, from seeps associated with oil-containing strata on the floor of the Gulf of Mexico, the community of microbes that collectively feeds on all the different compounds contained in the oil is well established and diverse. Even where the background levels of oil are low, a few microbes with the capability of degrading oil always seem to be present. There are a large number of hydrocarbon-degrading bacterial species in oil-rich environments, such as oil spill areas and oil reservoirs and that their abundance and quantity are closely related to the types of petroleum hydrocarbons and the surrounding environmental factors. Within the most abundant Proteobacteria and the shared OTUs in our samples from TMC and SVS are five genera from the Proteobacteria, *Anaerospira*, *Desulfococcus*, *Desulfosarcina*, *Shewanella* and *Halomonas*, that have been linked with the processes of anaerobic hydrocarbon degradation, sulfate reduction, denitrification and/or methanogenesis associated with petroleum biodegradation (Kostka *et al*, 2011; Koo *et al*, 2015).

*Desulfosarcina* is a hydrocarbon degrader. It is found in oil contaminated marine sediments can function as a chemoorganotroph or chemoautotroph, using formate, acetate, propionate, butyrate, higher fatty acids, other organic acids, alcohols, and benzoate or similar aromatic compounds as electron donors

for carbon sources produced by other members of an oil degrading consortia. *Halomonas* spp are recognized for producing exopolysaccharides (EPS) with amphiphilic properties that bind to hydrophobic substrates, such as hydrocarbons, increasing the solubilization of aromatic hydrocarbons and enhancing their biodegradation the microbial community. It was associated with oil-contaminated surface waters collected during the active phase of the Deepwater Horizon oil spill (Gutierrez *et al*, 2013; Kostka *et al*, 2011) and has been found associated with bacterial communities in crude oil contaminated marine sediments (Bargiela *et al*, 2015). *Halomonas* can also participate in the conversion of uric acid to ammonia stimulating growth of hydrocarbonoplastic bacteria (Gertler *et al*, 2015). Rhodobacteraceae can be used as sentinels for the later stages of degradation when more recalcitrant oil hydrocarbon compounds such as polycyclic aromatic hydrocarbons predominate. Bacteria with the potential to degrade PAHs express genes for oxygenases or peroxidases, such as those found in *Rhodococcus* (*alkB1* and *alkB2*) (Di Gennaro *et al*, 2010; Koo *et al*, 2015). A similar line-up of key players in oil degradation can be seen in Atlantic cod in the gut microbiota of fish in the waters close to the oil fields of southern Norway (Walter *et al*, 2019). These included Gammaproteobacteria from the orders Oceanospirales and Alteromonadales (*Shewanella*), Alphaproteobacteria from the order Rhodobacteraceae (f: Rhodobacteraceae) and Deltaproteobacteria from the order Desulfobacterales (*Desulfosarcina*).

In summary, analysis of 16S rRNA gene sequences in the stomachs of Gulf menhaden from TMC and SVS were dominated by Proteobacteria with the overall composition reflecting differences in water characteristics at the different sites, mainly salinity. This is consistent with the idea that the stomach microbiota to a large extent reflect the microbial constituents of the surface waters they occupy through their filter-feeding lifestyle. Reflecting this picture of the menhaden microbiota functioning as a biomonitor, the most abundant Proteobacteria contain genera that have been associated with petroleum biodegradation in keeping with the location of Apalachicola Bay in the northern Gulf of Mexico. Beyond mirroring the bacterial consortia in the surface waters,

some of the ninety-six OTUs present in all samples suggest that a core stomach microbiota may be providing a contribution to the nutrition of these rapidly growing juveniles. Among this consortium are bacteria that could provide a source of B-vitamins, and others that have the capacity to oxidize sucrose and glycerol and hydrolyze cellobiose, as well as those having chitinase and cellulase activities. In addition, arylsulfatases from the Rhodobacteraceae could make sulfated polysaccharides from diatoms available for digestion. This would be particularly valuable for juveniles feeding in estuarine waters in early summer in which up to 80 % of their stomach contents consist of detritus from particulate organic matter, mainly from the Apalachicola River and surrounding freshwater wetlands, and coastal marshes.

### *3.6 Acknowledgements*

This work was supported by the National Oceanic and Atmospheric Administration, Educational Partnership Program award to the Living Marine Resources Cooperative Science Center (LMRCSC) NA11SEC4810002 and NA16SEC4810007. Ammar Hanif was supported in part by MD SeaGrant Graduate Student Fellowship, NA10OAR4170072 SA7528129-D and in part as an LMRCSC Graduate Fellow. Thanks to Dr. Stacy Smith, Delaware State University for coordinating the menhaden collection.



## Chapter 4: Diet of juvenile Gulf Menhaden (*Brevoortia patronus*) using DNA metabarcoding

### 4.1 Abstract

The Gulf menhaden, *Brevoortia patronus*, is a key forage fish species that serves as a trophic link between the plankton and piscivorous predators in the Gulf of Mexico. As juveniles, in early spring, these obligate filter-feeders forage on plankton and plant detritus in the lower salinity tidal creeks and near-shore marshes of Apalachicola Bay. Later in the summer, they are found in the open bay in areas like St. Vincent Sound. Diet assessment is difficult because the food items are small and easily digested. Indirect assessment of diet by stable isotope analysis is complicated by system-specific processes. Here we show that DNA metabarcoding, can provide an alternative method for assessing diet as defined by eukaryotic prey. The MiSeq Illumina high-throughput sequencing platform was used to sequence the V4-V5 region of the 18S ribosomal (rRNA) gene of the stomach contents from juvenile menhaden collected at two locations within Apalachicola Bay. Fish were collected in early spring (May 2013) at Two Mile Channel close to the Apalachicola River, and six weeks later in summer at St. Vincent Sound, in the enclosed Apalachicola Bay. Of the water characteristics measured, salinity was observed to make the greatest difference to organisms found in the stomach. 1048 unique OTUs (species) were identified, 1035 of which were found in samples from both locations. Members of the Stramenopile/Alveolalata/Rhizaria (SAR) clade account for 66 % representation in samples from Two Mile Channel, dominated by the diatoms *Cyclotella* and *Skeletonema*, as well as the ciliate *Oligotrichia*. In contrast, Metazoa (zooplankton) dominate in samples from St. Vincent Sound, accounting for 83.77 % of the reads. These are mainly *Acartia* copepods. Not only does the diet shift from SAR to Metazoa at these two locations, but significant differences in the species represented within these taxa occur. This dietary shift is indicative of a difference in trophic level in fish caught at SVS, supported by an increase in  $\delta^{15}\text{N}$  in the muscle of fish from SVS. The results from 18S rRNA gene sequences

have provided a more complete description of the diet of menhaden from the two locations as the fish move away from the estuary to the enclosed bay, pointing to an early trophic shift in juveniles of this ecologically and economically important fish.

#### 4.2 Introduction

Forage fish species are exceptionally important to the structure and functioning of marine ecosystems, serving as the main conduit of energy flow from lower to upper trophic levels (Pikitch *et al*, 2014). They can exact either top-down control on plankton or bottom-up control on predators (Cury *et al*, 2000). Gulf menhaden (*Brevoortia patronus*) is a key forage fish species that serves as a trophic link between the plankton and piscivorous predators in the Gulf of Mexico. Numerous ecosystem models have shown the overall importance of Gulf menhaden (*Brevoortia patronus*) to the ecosystem (Geers *et al*, 2016; O'Farrell *et al*, 2017). It is an important forage fish species along the Gulf coast providing an important food source for fish, predatory seabirds and marine mammals (Vaughan *et al*, 2007). Many of the commercially and recreationally harvested fish species along the Gulf coast rely on the abundant schools of menhaden, including king mackerel (*Scomberomorus cavalla*), Spanish mackerel (*Scomberomorus maculates*), dorado (*Coryphaena hippurus*), crevalle jack (*Caranx hippos*), tarpon (*Megalops atlanticus*), red drum and bonito (*Sarda sarda*) (Vaughan *et al*, 2007; Franklin, 2011; Sagarese *et al*, 2016). Among other species, the diets of the blacktip shark (*Carcharhinus limbatus*) and the brown pelican (*Pelecanus occidentalis*), Louisiana's state bird, consist of over 95 % menhaden (Franklin, 2011).

Gulf menhaden along with shrimp support the largest fisheries by landings and by revenue in the Gulf of Mexico (Vaughan *et al*, 2007; O'Farrell *et al*, 2017). Gulf menhaden also support a large directed reduction fishery. Gulf menhaden are reduced to fish oil and meal used in livestock feed, aquaculture feed, menhaden oil, high in omega-3 fatty acids, has been used for pharma- and nutraceuticals, cosmetics and other consumer products (Olsen *et al*, 2014;

SEDAR, 2013; Menhaden Advisory Commission, 2015). Assessments of the fishery have concluded that Gulf menhaden are not overfished or undergoing overfishing (Vaughan *et al*, 2007). However, this assessment does not fully consider the ecological role of the species and its ecological importance has not been adequately quantified suggesting that the potential exists for this large fishery to yield an ecological impact (Olsen *et al*, 2014; Sagarese *et al*, 2016).

Despite the importance of Gulf menhaden as a dominant prey fish and its economic importance in fisheries, there is a dearth of specific dietary data for this species. Most of the menhaden dietary studies have been on the allopatric species, the Atlantic menhaden, *Brevoortia tyrannus*, which does not overlap with *B. patronus*. In view of this, a more complete understanding of the diet of Gulf menhaden is needed. Juvenile and adult Gulf menhaden are obligate filter feeders. After menhaden spawn in coastal waters and bays, their larvae are transported into various estuarine habitats where they metamorphose into juveniles. After transformation, juveniles remain in nearshore areas where they travel in dense schools, often near the surface (Lassuy, 1983). The morphological changes associated with transformation from the larval form are accompanied by a change in feeding behavior. The larvae are carnivorous particulate feeders, the juveniles switch to omnivorous filter feeders, reported to consume phytoplankton, zooplankton, and detritus (Deegan *et al*, 1990). Gulf menhaden are obligate filter feeders and as such are constantly sampling the environment as they feed. As juveniles they live in the mesohaline and oligohaline zones of estuaries where they filter the water column via their gill rakers. Friedland *et al* (1984) calculated the plankton filtering capability of Atlantic menhaden juveniles by measuring filtration efficiencies of uniformly sized particles in captive fish (Friedland *et al*, 1984). For 14-cm fork-length fish, a size representing the upper limit of juvenile size ranges for this species, the fish can filter uniformly sized particles in a range of 7–9  $\mu\text{m}$  in diameter at filtration efficiencies of approximately 10 %, although maximum filtration efficiency is for particles approximately 100  $\mu\text{m}$  (Friedland *et al*, 1984). Based on this, the effective pore size of the gill rakers of menhaden juveniles typical of spring and

summer nursery residents (7.5-cm fork-length fish) was estimated to be 60 % of this, or 4.2-5.4  $\mu\text{m}$  diameter. However, for fish in the wild, particle retention is enhanced when a background of other particles is present, as is typical in estuarine water columns. It has been proposed that detrital material plays a role in the retention of these smaller particles into the stomach (Friedland *et al*, 1984). Detritus is primarily plant-related structural material that washes out of the estuaries and coastal marshes and can also include bacteria, fungi, microalgae and protozoa. Its makeup and quantity in the menhaden stomach reflect the constituents of the water filtered (Deegan *et al*, 1990; Lewis & Peters, 1994). There is a range of opinions on the importance of detritus as a food source. However, analysis of stable isotope evidence has supported its contribution to the diet of Gulf menhaden (Deegan *et al*, 1990; Olsen *et al*, 2014).

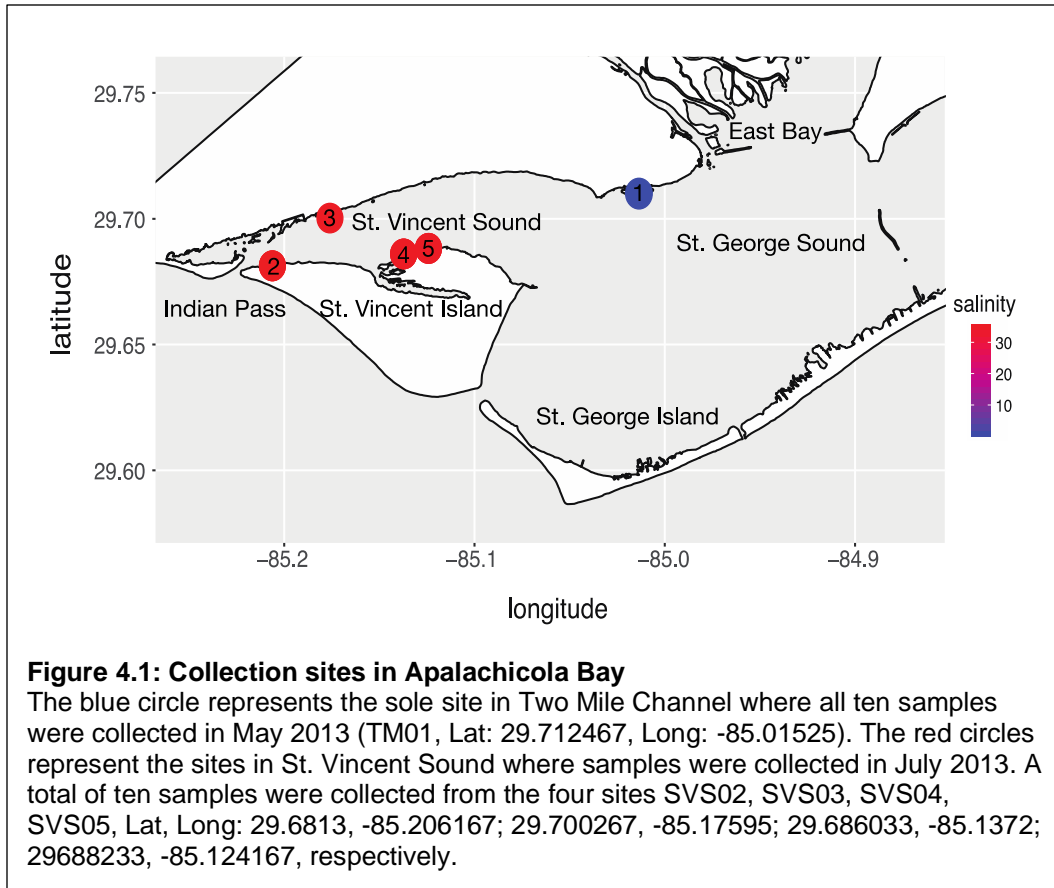
Visual identification has conventionally been used for the analysis of gut contents of fish based on prey morphology (Hyslop, 1980). The method is time consuming and requires expertise in taxonomic classification. In addition, menhaden have a gizzard-like stomach that grinds ingested items to an amorphous paste making microscopy techniques extremely difficult to use for identification (Lewis & Peters, 1994). Coupled with extensive digestion in the stomach, this allows only partial or ambiguous prey specimen identification (Baker *et al*, 2014). To understand food web dynamics, the whole dietary breadth needs to be described. Through DNA metabarcoding methods, organisms in the stomach content of filter feeders can now be assessed even when diagnostic taxonomic features are lacking (Pompanon *et al*, 2012). DNA metabarcoding has increasingly become the method of choice in characterizing diets of marine animals (Symondson, 2002; Jarman *et al*, 2002; Deagle *et al*, 2009; Pompanon *et al*, 2012). Applying DNA techniques to diet identification has increased identification resolution, particularly in marine systems (Blankenship & Yayanos, 2005; Riemann *et al*, 2010; Cleary *et al*, 2012; Jakubavičiūtė *et al*, 2017; Waraniak *et al*, 2019). Such work has identified trophic linkages within food webs and determined predator diet breadth and preference. Hanif *et al* (2020) has successfully developed a DNA metabarcoding method to characterize the

stomach contents in Gulf menhaden using 18S rRNA gene sequences (Hanif *et al*, 2020). In the present study, we apply this method to analyze and compare the eukaryotic prey items of juvenile Gulf menhaden collected from two different locations within Apalachicola Bay, Florida.

### 4.3 Methods

#### *4.3.1 Sample collection*

Delaware State University (DSU) collaborators collected samples on May 23, 2013 at Two-Mile Channel in Apalachicola Bay, Florida, (TMC01, Lat:29.712467, Long: -85.01525) (**Figure 4.1**). Gulf menhaden samples were collected from this location using a 10-foot cast net. Ten samples from this low salinity site (T1-10) were used for this study. At that time of year and location, the menhaden were concentrated in large, slow-moving, and tightly packed schools, allowing all fish to be collected from the same site within this location. Six weeks later, July 2, 2013, a second collection was made by Florida Fish and Wildlife from the St Vincent Sound location (SVS). At that time and location, the menhaden were not so tightly grouped. Fewer fish were collected per cast such that ten fish were collected from different sites at this location. This resulted in three fish (S1-3) being collected at site SVS02, one fish (S4) collected at SVS03, three fish (S5-7) collected at SVS04 and three fish (S8-10) collected at SVS05. These four sites corresponded to latitude and longitude coordinates 29.6813, -85.206167 (SVS02); 29.700267, -85.17595 (SVS03); 29.686033, -85.1372 (SVS04) and 29.688233, -85.124167 (SVS05), respectively (**Figure 4.1**). Collections were done at 1 m depth. Water quality measurements of temperature, salinity, pH, dissolved oxygen, and Secchi depth were recorded at all sites (**Table 4.1**).



Location	Sample	Site	Salinity ppt	Temp °C	pH	DO	Secchi
Two Mile Channel	T1	TMC01	1.8	27.1	7.5	5.7	0.7
	T2	TMC01	1.8	27.1	7.5	5.7	0.7
	T3	TMC01	1.8	27.1	7.5	5.7	0.7
	T4	TMC01	1.8	27.1	7.5	5.7	0.7
	T5	TMC01	1.8	27.1	7.5	5.7	0.7
	T6	TMC01	1.8	27.1	7.5	5.7	0.7
	T7	TMC01	1.8	27.1	7.5	5.7	0.7
	T8	TMC01	1.8	27.1	7.5	5.7	0.7
	T9	TMC01	1.8	27.1	7.5	5.7	0.7
	T10	TMC01	1.8	27.1	7.5	5.7	0.7
St. Vincent Sound	S1	SVS02	37.6	28.4	8.0	5.5	0.6
	S2	SVS02	37.6	28.4	8.0	5.5	0.6
	S3	SVS02	37.6	28.4	8.0	5.5	0.6
	S4	SVS03	36.3	27.8	7.9	5.5	0.6
	S5	SVS04	36.4	28.7	8.0	5.8	0.4
	S6	SVS04	36.4	28.7	8.0	5.8	0.4
	S7	SVS04	36.4	28.7	8.0	5.8	0.4
	S8	SVS05	36	29	7.8	5.7	0.4
	S9	SVS05	36	29	7.8	5.7	0.4
	S10	SVS05	36	29	7.8	5.7	0.4

**Table 4.1: Water quality measurements from each sample site in Apalachicola Bay.**

Water quality measurements of temperature, salinity, pH, dissolved oxygen, and turbidity were recorded at the time and place of sample collection using a YSI 556 multiparameter water quality meter. The sites and samples are color coded to be consistent with the colors used in Figures 4.6 and 4.7.

#### 4.3.2 Sample preparation and DNA extraction

Sample preparation, DNA extraction from menhaden stomach contents and estimation of DNA quality was done as described by Hanif *et al* (2020) (Chapter 2).

#### 4.3.3 High throughput sequencing

High throughput sequencing was performed on the Illumina MiSeq platform located in the BioAnalytical Services Laboratory (BAS Lab) at the University of

Maryland Center for Environmental Science, Institute of Marine and Environmental Technology as described by Hanif *et al* (2020), using the protocol recommended in the Nextera DNA Library Prep Reference Guide (Illumina Document #15027987v1). with the exception of sequencing primers. The sequencing primer set used were 574\*F (CGGTAAYTCCAGCTCYV) and 1132R (CCGTCAATTHCTTYAART) targeting the V4-V5 region of the 18S rRNA gene as developed by Hugerth *et al* (2014).

#### 4.3.4 *Post sequencing pipeline*

Removal of index sequences (called de-multiplexing), base calling and data quality assessment were performed on the MiSeq instrument. MacQIIME was used to process and assess the quality of output reads using the recommended QIIME pipeline for Illumina reads. The MacQIIME script, `join_paired_ends.py`, was used to join forward and reverse reads. Paired reads were then filtered for low quality reads (quality score of <25) and short read length (<200 bp) to be removed from the library using the `split_libraries.py` command. Chimeric sequences were identified *de novo* using Usearch61 with the script `identify_chimeric_seqs.py`. This was followed by identifying sequences containing PhiX sequences by a BLAST analysis of the library with PhiX sequence. The resulting file after the removal of chimeric and PhiX sequences was used for operational taxonomic unit (OTU) (species) picking. OTUs generated from 18S rRNA gene sequencing were picked and taxonomy was assigned using the UCLUST method against the Silva 111 Eukaryote-only database (Edgar 2010; Quast *et al*, 2013). This was used to construct an OTU table for subsequent analysis.

#### 4.3.5 *Data analysis*

Observed OTUs, Shannon index, and Inverse Simpson index metrics were used to assess alpha diversity. A Bray-Curtis matrix was generated and visualized using principal coordinate analysis to assess beta diversity. DESeq2 was used to test for significant differentially abundant taxa. All analyses were done in R primarily using the packages Phyloseq and DESeq2.



#### 4.3.6 *Stable isotope analysis*

Menhaden collected from Apalachicola Bay were immediately placed on dry ice and frozen until samples were transported back to the lab. Once back at the laboratory, all menhaden were stored at -80 °C until preparation for isotopic analysis. Using a stainless X-Acto knife, slivers of menhaden muscle were placed in tins and oven-dried at 60 °C for 6 h. The tissue was pulverized using a mortar and pestle. Approximately 1-1.3 mg of muscle was removed from each sample and weighed. The prepared samples were sent to the Central Appalachian Laboratory Stable Isotope Facility (CASIF) of the University of Maryland Center for Environmental Sciences for  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  analysis.

Given the low sample size, the Anderson-Darling normality test was used to test for normality. This was followed by a test to determine if the variance between the TMC and SVS sample groups was similar. A student's t-test was used to examine significant differences of the means of the  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values between TMC and SVS samples. Results were deemed significant at  $\alpha \leq 0.05$ .

### *4.4 Results*

#### *4.4.1 Characterization of stomach 18S rRNA gene sequences from menhaden caught at Two-Mile Channel and St. Vincent Sound*

DNA from all TMC samples generated useable reads, however only eight SVS samples produced reads. Table 4.2 gives a summary of reads per sample. A total number of 315,874 and 350,223 raw reads were retrieved from TMC and SVS samples respectively. After post sequencing processing in QIIME, these numbers were reduced to 315,072 and 349,697 with an average of 31,072 and 34,697 reads per sample. Two of the SVS samples did not produce reads of sufficient quality and were not included in this calculation or further analysis.

Sample ID	Raw reads	Post processing
T1	60,588	60,382
T2	85,537	85,243
T3	19,840	19,797
T4	7,764	7,742
T5	14,932	14,892
T6	15,546	15,515
T7	16,297	16,271
T8	32,447	32,404
T9	45,379	45,304
T10	17,544	17,522
S2	54,085	53,989
S3	84,779	84,679
S4	76,896	76,757
S5	15,450	15,423
S6	44,303	44,254
S7	39,251	39,191
S9	30,788	30,743
S10	4,671	4,661

1a

Total reads	666,097	664,769
TMC	315,847	315,072
SVS	350,223	349,697
Average reads	37,005.4	36,931.6
TMC	31,587.4	31,507.2
SVS	43,777.9	43,712.1

1b

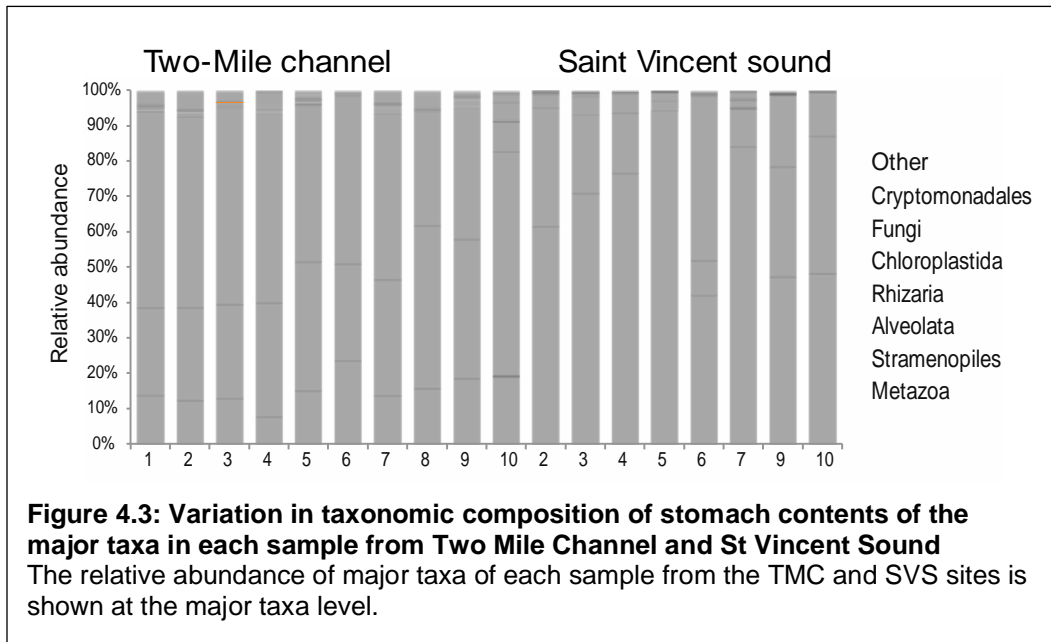
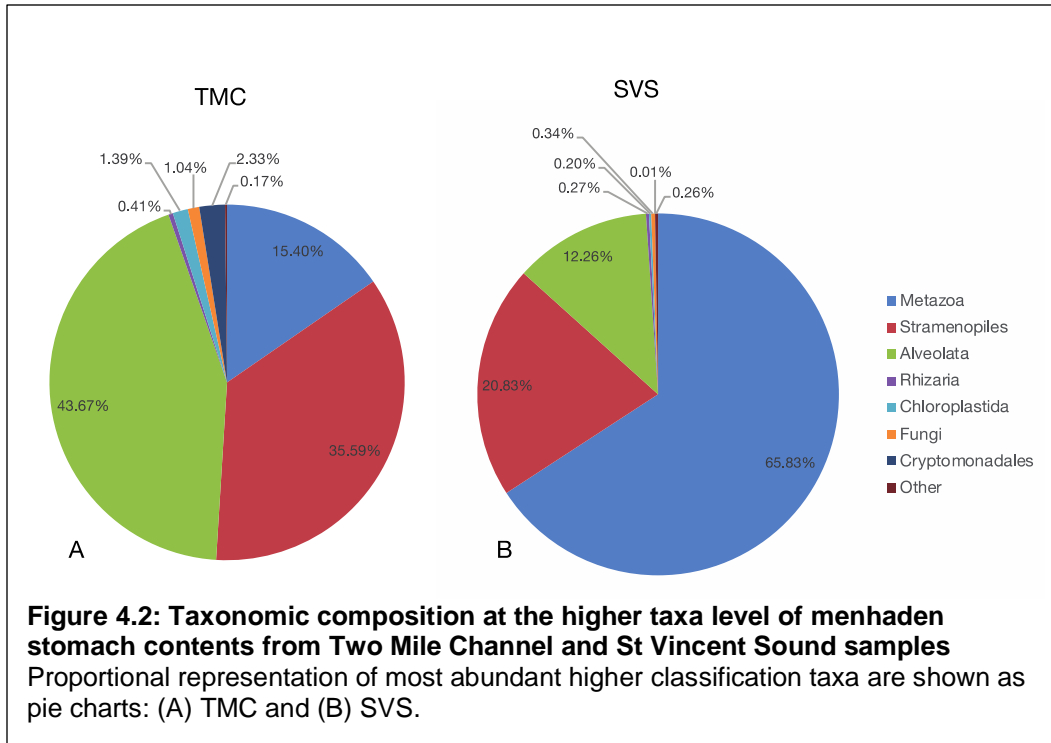
**Table 4.2: Comparison of the number of the number of raw reads and post-processing reads from Two Mile Channel and St Vincent Sound samples**

A region of approximately 550 bp encompassing the V4-V5 hypervariable regions of the 18S rRNA genes was targeted for sequencing using the protocol developed by Hugerth (Hugerth *et al* 2014). The raw sequencing and post-processing reads from each were expressed as reads per sample, mean number of reads per sample and total reads.

There was a high percentage of reads that matched the teleost fish *Tenualosa reevesii* after using the SILVA 111 Eukaryote only database for taxonomic assignment (Hanif *et al*, 2020). We reasoned that these sequences represented the 18S rRNA sequence of the menhaden and removed them from our analysis. Subsequent sequencing of menhaden 18S rRNA (accession #

MN335200) showed it to be 99 % identical to that of *Tenualosa reevesii*. The relative abundance of this OTU was sample dependent and ranged from 81.6 to 0.3 %. Excluding the reads assigned to *T. reevesii* did not change the overall relative abundance pattern of the remaining taxa, except for sample T10.

Binning at 0.03 % divergence resulted in total of 1048 unique OTUs assignments, with 1035 OTUs shared in samples from each location, 10 unique to samples from TMC and 3 unique to samples from SVS. There were seven taxa with greater than 0.36 % representation in the total relative abundance (**Figure 4.2, Suppl Table 4.1**). All these taxa were represented in the stomachs of fish caught at both locations, although at very different relative abundances (percent reads). In samples from both locations the dominant taxa were from the Stramenopile/Alveolalata/Rhizaria (SAR) clade and the kingdom Metazoa. In samples from TMC, OTUs were dominated, in descending order, by Alveolata (alveolates) with a mean representation of 38.5 %, Metazoa 30 % and Stramenopiles (diatoms) at 27.5 % (**Figure 4.2 A, Suppl Table 4.1**). DNA from the stomach samples collected from SVS sites were dominated, in descending order, by Metazoa at 83.77 %, Stramenopiles (diatoms) at 8.53 % and Alveolata (alveolates) at 7.2 % (**Figure 4.2 B**). Representation of Rhizaria was negligible in samples from both locations. The remaining most abundant OTUs were assigned to Chloroplastida, Fungi, Cryptomonadales and Foraminifera. Representatives of all mentioned taxa were present in at least one sample per site. However, the reads from Cryptomonadales and Fungi were more prevalent in samples from the TMC than the SVS location. **Figure 4.3** shows the variation in the relative abundance of the dominant taxa in each sample.

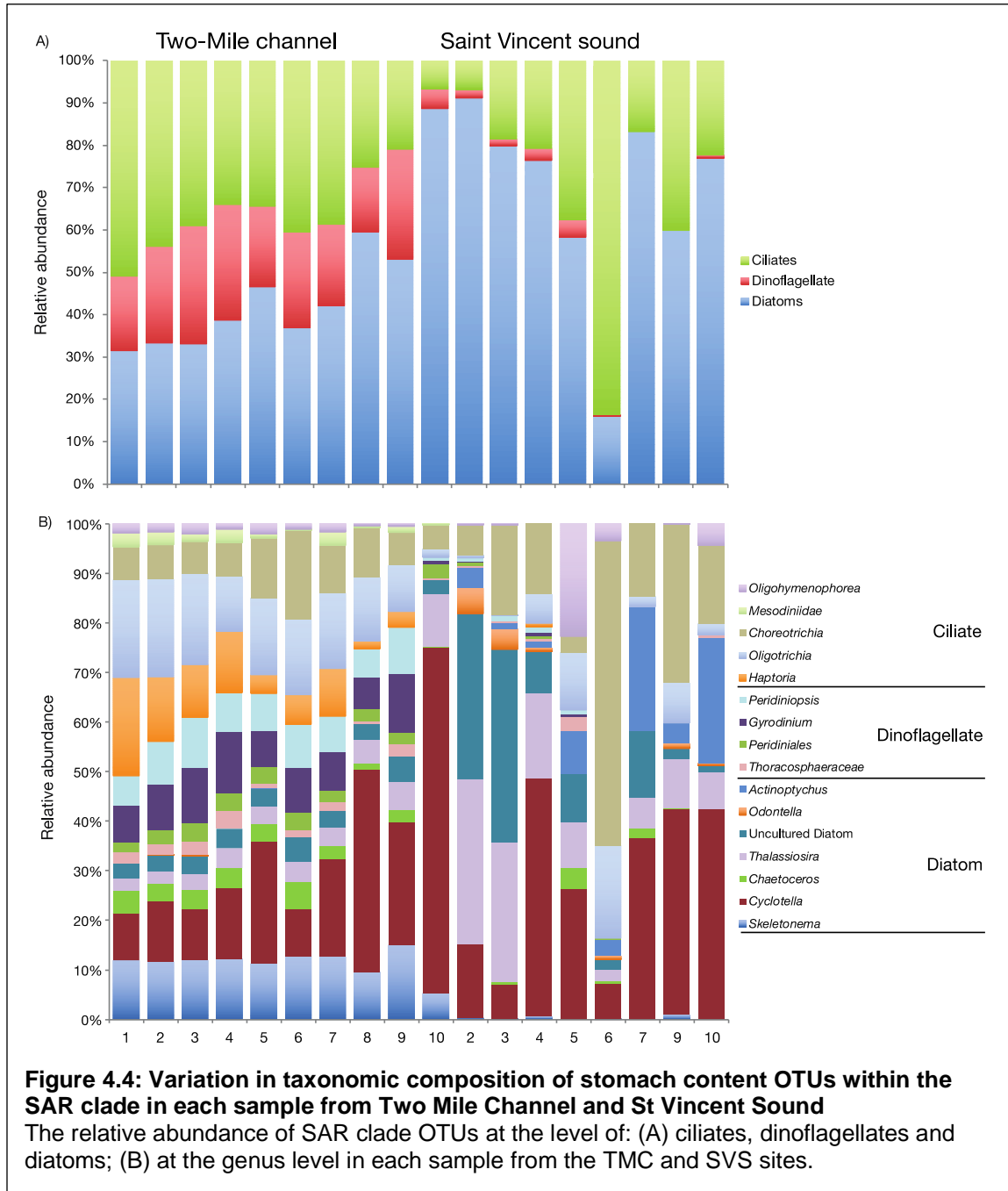


Given that the menhaden stomach contents contain mainly OTUs from the SAR clade and the kingdom Metazoa, which together reflect 96.44 % and >99 % of the total reads from TMC and SVS, respectively, I focused my analysis further to look at the abundances of these taxa at lower classifications. Looking only at OTUs within the SAR clade that account for 66 % and 15.82 % of reads from the

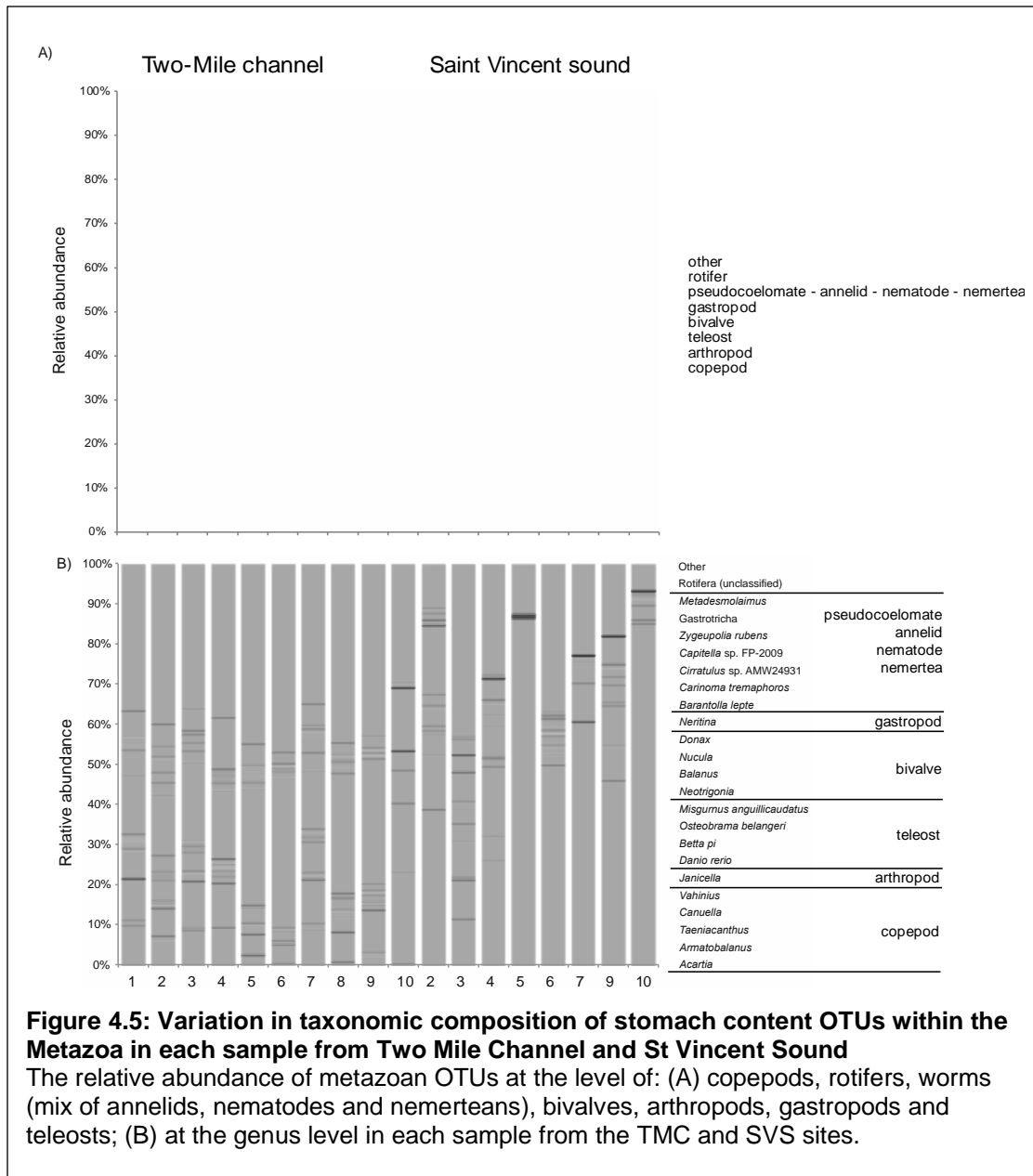
samples from TMC and SVS, respectively, revealed these OTUs belonged to three taxa; the Stramenopile phylum Bacillariophyta (diatoms), the alveolate class Dinophyceae (dinoflagellates) and the alveolate phylum Ciliophora (ciliates) (**Figure 4.4**). Although dinoflagellates and diatoms are considered to be phytoplankton, ciliates are non-photosynthetic and are considered to be microzooplankton. The overall relative abundance from these three taxa is shown in **Suppl Table 4.2** and as relative abundance per sample in **Figure 4.4A**. In samples from TMC, ciliates, dinoflagellates and diatoms overall represent 30.46, 18.41 and 42.64 % of the representation from SAR clade OTUs. This means that overall, phytoplankton have 60 % representation in stomach samples from TMC. Representatives of all three taxa were present in all samples except S7 and S9 in which dinoflagellates were absent. In samples from TMC, diatoms were the largest represented SAR taxon for samples T4, T5, T7, T8, T9, and T10 (**Figure 4.4A**). For the remainder of the TMC samples, the ciliates were the largest represented SAR taxon. In samples from SVS, ciliates, dinoflagellates and diatoms overall represent 27.9 %, 1.23 % and 36.62 % of the representation from SAR clade OTUs, respectively. Diatoms were the most abundant SAR taxon in SVS samples except for S6 which was dominated by ciliates (**Figure 4.4A**).

**Figure 4.4B** shows the representation of OTUs from the SAR clade assigned at the genus level. There was a higher diversity of SAR OTUs in TMC samples than in SVS samples, except for T10 which was dominated by diatoms. The three most abundant SAR species in TMC samples were the diatoms *Cyclotella*, *Skeletonema*, and the ciliate *Oligotrichia* accounting for 22 %, 10.41 %, and 12.63 % relative abundance respectively. The ciliates *Choreotrichia* and *Haptoria* had overall relative abundances of 8 % and 7.31 %, respectively and the dominant dinoflagellates were *Peridinales* and *Gyrodinium* at 6.53 % and 7.54 % relative abundances, respectively. The diatom *Odontella* was the least represented species in TMC samples only appearing in two samples from this site versus six samples from SVS. As in samples from TMC, *Cyclotella* was also the most abundant SAR species in SVS samples with 25 % relative abundance. In SVS samples however, this was followed by the ciliate *Choreotrichia*, the

diatoms *Thalassiosira* and *Actinoptychus*, and an unassigned diatom species with relative abundances of 19 %, 12 %, 8.5 % and 11 % respectively. Ciliates and diatoms have similar representation within the SAR clade in stomach samples from TMC and SVS but are different species.



Metazoa account for 83.77 % of reads from SVS samples and approximately 70 % of these were copepods (**Figure 4.5A, Suppl Table 4.3**). If the ciliate microzooplankton, are included with the mesooplanktonic copepods, zooplankton have over 90 % representation in stomach samples from SVS. There were six copepod species identified, although the copepod *Vahinius* was only found in one sample from SVS (**Figure 4.5B**). Stomach samples from SVS were dominated by *Acartia* which represented a mean relative abundance of 51 %. Metazoa accounted for only 30 % mean representation in samples from TMC and copepods only represented 21 % of these, with 7.3 % mean relative abundance for *Acartia* and 12 % for *Canuella*. In TMC samples, the most highly represented metazoan was the polychaete worm, *Barantolla*, from the phylum Annelida, with 20 % relative abundance. This is not apparent in **Figure 4.5A** because it is included with other polychaete worms as well nemerteans and a nematode, *Metadesmolaimus*. Overall polychaete worms have a 24.13 % relative abundance and the nemertean worms 4.4 %. A bivalve, *Neotrigonia*, is present in TMC samples with a 5.4 % relative abundance. The adult annelid and molluscan species found had size ranges inconsistent with the size of menhaden gill rakers so their presence in menhaden stomach are likely to represent eggs or juveniles or detritus. Also, in stomach samples from TMC was an unassigned rotifer with 8 % overall relative abundance.



The relative abundance of the remaining OTUs was low ranging from 3 - 0.01 % except for teleost DNA. Teleost DNA was represented in all samples except for T1 in which *Betta pi*, *Osteobrama belangeri*, and *Misgurnus anguillicaudatus* were absent. The mean relative abundance of teleosts was 10 % for both TMC and SVS samples. However, for TMC samples this was inflated by the high relative abundance of teleost DNA in sample T10. Teleost DNA in this sample showed a very different relative abundance compared to that in other



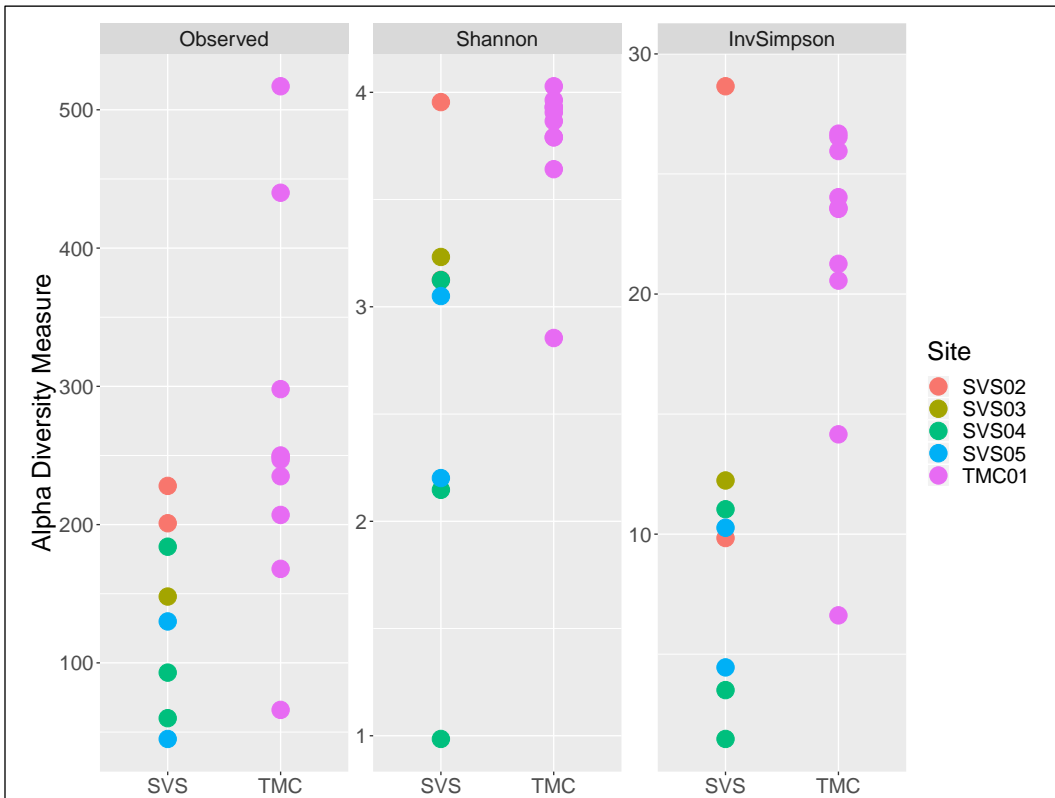
TMC samples. The closest matches to the teleost DNA was of species that are all freshwater from S. Asia and commonly found in the freshwater aquarium pet trade. However, they could be from species without sequences in the SILVA database.

There were ten OTUs represented in the TMC samples that were undetected in the SVS samples. Four of these were unassigned eukaryotes. The remainder were assigned as *Goniomonas*, a genus of cryptomonads, Discosea, a class of Amebozoa, Prymnesiophyceae, a class of Haptocyta and *Katablepharis*, *Leucocryptos* and *Roombia*, all genera of the katablepharid Cryptista. Though unique to the TMC samples their relative abundance was extremely low, ranging between 1e-05 to 1e-06 %. In contrast only three OTUs were solely represented in the SVS samples. One was an unclassified Cryptophyceae, one an assigned member of a haptophyte family, Pavlovophyceae, and the last was an uncultured Rhodophyte (red alga). Similarly, all had extremely low relative abundance ranging between 1e-05 to 1e-06 %.

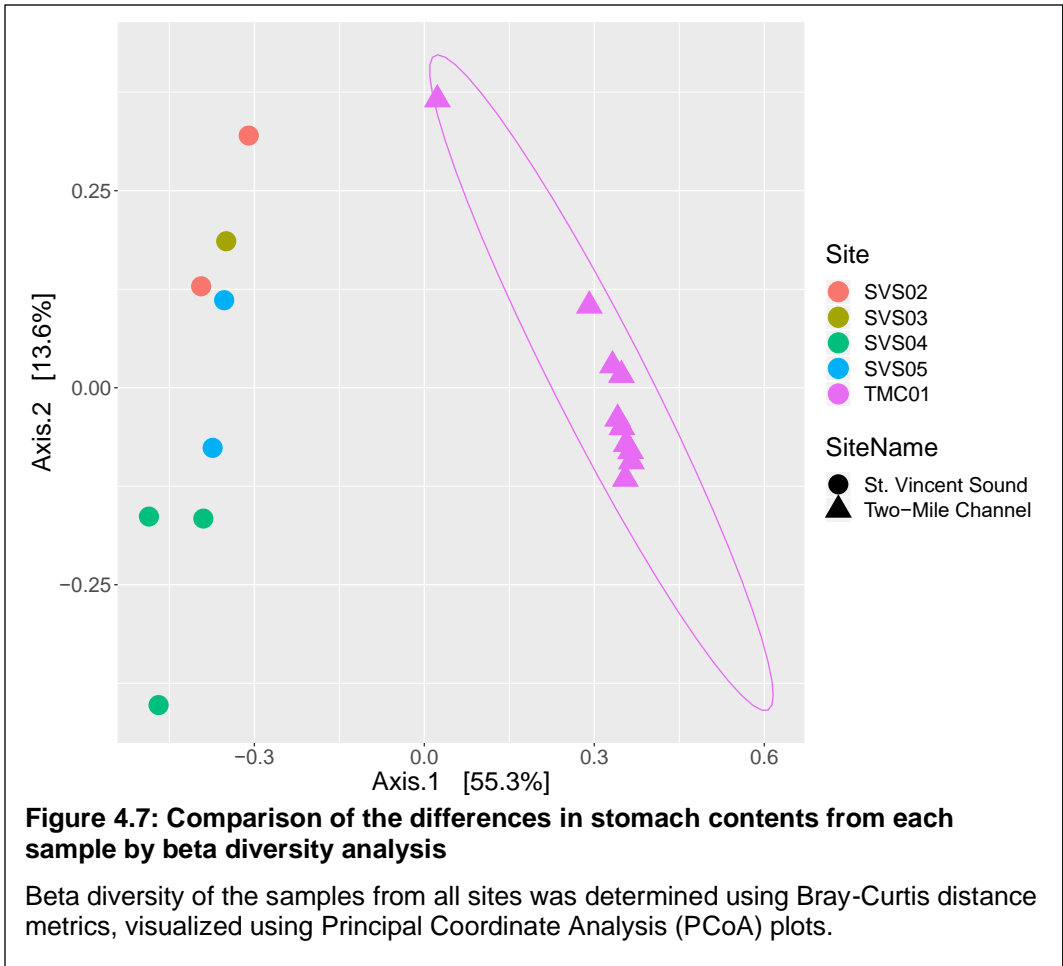
#### 4.4.2 *Alpha and beta diversity analyses show that stomach samples from TMC are more diverse than those from SVS*

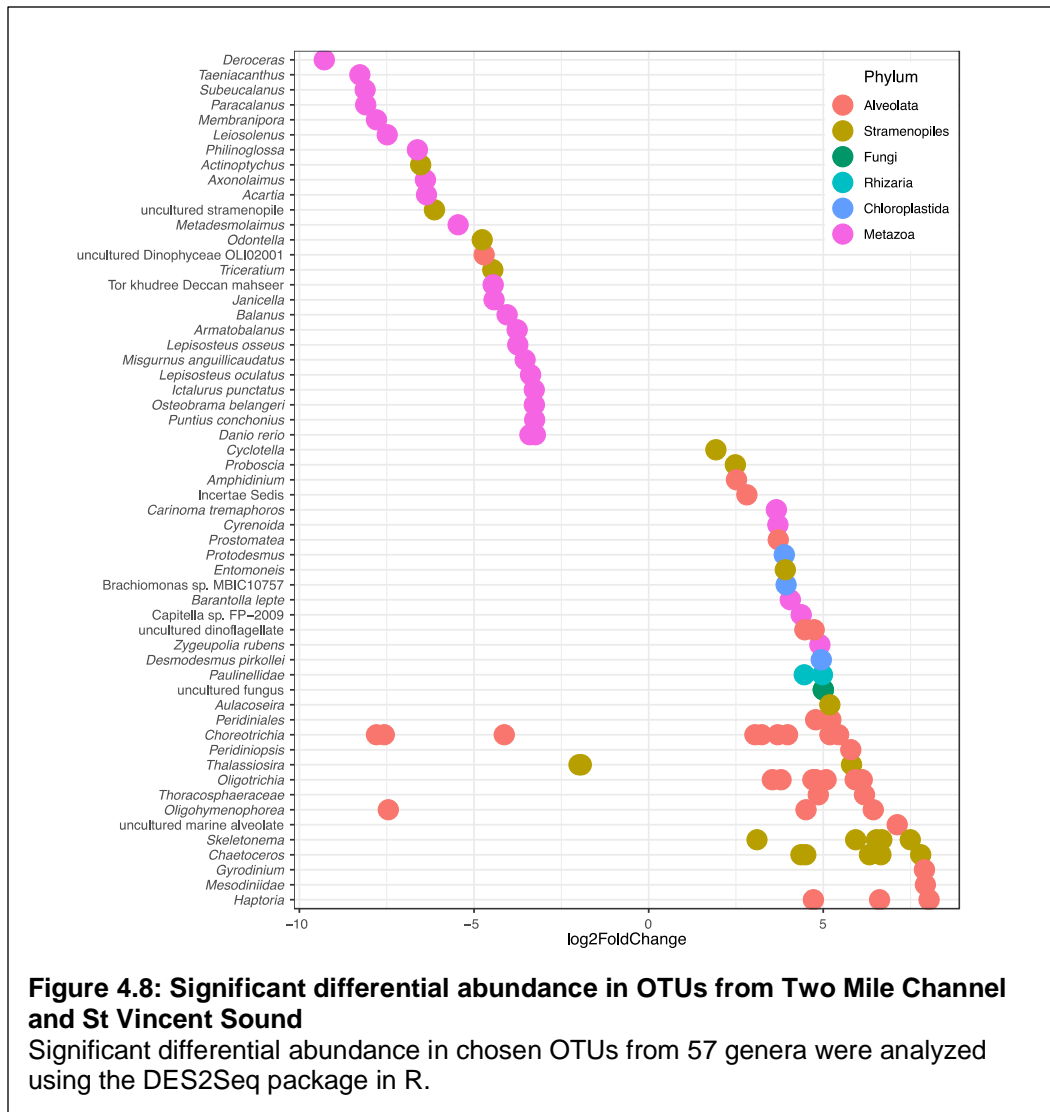
The alpha diversity metrics, Observed Species, Shannon and Simpson inverse measures all show that DNA in stomach samples from TMC was more diverse than that from SVS samples with respect to both richness and evenness (**Figure 4.6**). Furthermore, each metric was shown to be significantly different as measured by Mann-Whitney test (p-value > 0.05). The Bray-Curtis dissimilarity test was visualized by principal coordinate analysis (PCoA) (**Figure 4.7**). There is a clear separation of samples by sampling location, most likely reflecting the difference in salinity. TMC samples group closely together with the exception of one sample. This is in contrast to SVS samples which do not group as closely together, even for those samples taken at the same site. Testing for significant differential abundance was done using the DES2Seq package in R. Overall there were fifty-seven genera whose abundances were found to be significantly different between samples from the different sampling locations (**Figure 4.8**).

Focusing on the most dominant taxa as shown in **Figures 4.4** and **4.5**, the phylum Metazoa, three copepod genera (*Taeniacanthus*, *Acartia*, *Armatobalanus*), three teleosts (*Misgurnus anguillicaudatus*, *Osteobrama belangeri*, *Danio rerio*), the annelid *Barantolla*, the nematode, *Metadesmolaimus*, the ribbon worm, *Carinoma tremaphoros*, the single decapod (*Janicella*), and a bivalve (*Balanus*) had significantly different abundances at the two locations. Amongst the Stramenopiles there were five diatoms (*Odontella*, *Cyclotella*, *Thalassiosira*, *Skeletonema*, *Chaetoceros*) whose abundances were found to be significantly different. For the Alveolata, the abundances of all ciliates and dinoflagellates mentioned in **Figure 4.3** were found to be significantly different between locations.



**Figure 4.6: Comparison of the differences in stomach contents from each sample by alpha diversity metrics**  
 The alpha diversity metrics Observed Species (OTUs), Shannon, and Simpson inverse measures are shown for all samples from each sampling site to measure OTU



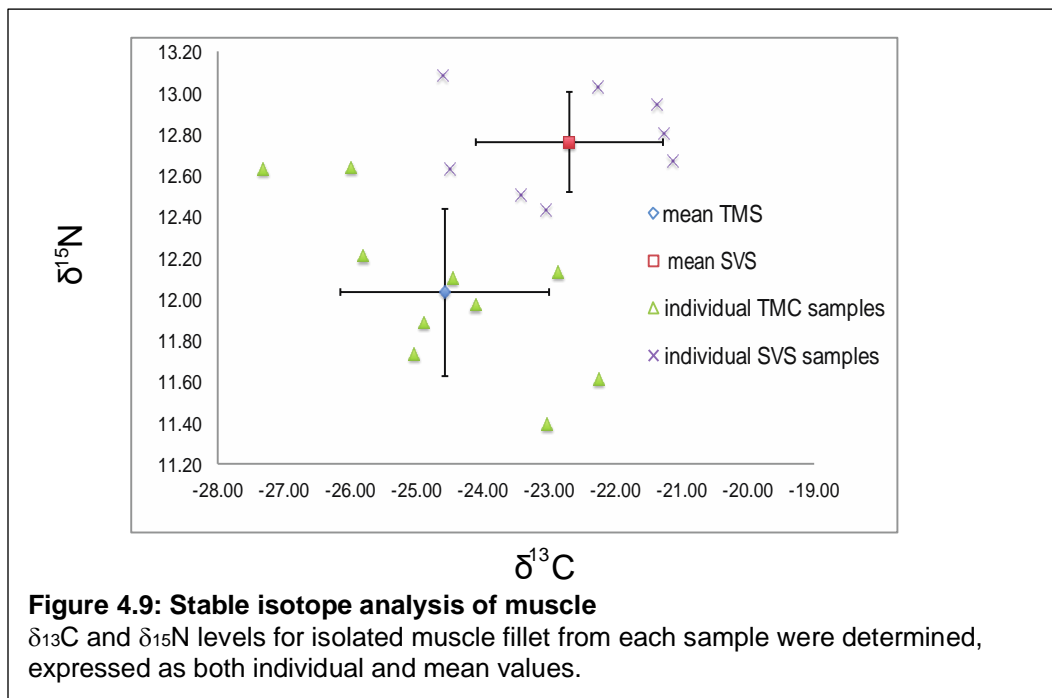


#### 4.4.3 Stable isotope analysis

Carbon and nitrogen stable isotopes ( $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$ ) can be used to evaluate the relative contributions of different food sources and the trophic position of fish (Anderson *et al* 1987; Lochman and Phillips 1996; Gu *et al* 1996a,b; Gamboa-Delgado *et al* 2008). Because isotope compositions reflect the organic compounds that have been incorporated into the bodies of consumers, the measurements of  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  provide information on the dietary component assimilated by consumers. Isotope enrichment in consumers takes place during the assimilation of carbon and nitrogen from the diet (Post, 2002). The average isotope enrichment during each trophic transfer is considered to be

0.5% for  $\delta^{13}\text{C}$  and 3.4% for  $\delta^{15}\text{N}$ . However, many unknown and uncontrolled factors such as food sources and differences in growth rates may affect the magnitude of isotope fractionation giving large variations in both stable isotopes (Post, 2002).

Samples from TMC were lower in both  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  in comparison to those from SVS. TMC samples  $\delta^{13}\text{C}$  values ranged from -27.31 to -22.25 and  $\delta^{15}\text{N}$  values ranged from 11.39 to 12.64. SVS samples  $\delta^{13}\text{C}$  values ranged from -24.60 to -21.13 and  $\delta^{15}\text{N}$  values ranged from 12.43 to 13.08. The range of both  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values was greater for TMC samples than SVS samples. This resulted in the mean values of TMC and SVS for  $\delta^{13}\text{C}$  -24.57 and -22.70 and  $\delta^{15}\text{N}$  12.03 and 12.76 respectively. Though the means are close, the means of the  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  were shown to be significantly different for both TMC and SVS samples as determined by t-test (p-value=0.01722 and p-value=0.0002654 respectively, (Figure 4.9). The increase in  $\delta^{15}\text{N}$  in the muscle of fish from SVS is consistent a trophic shift. However, the difference in stable isotope composition is not definitive by itself since system-specific processes can alter local isotope values and values for plankton and debris from the different sites are not available for comparison.



#### 4.5 Discussion

In the present study, we have compared the stomach contents of juvenile Gulf menhaden collected from two different locations within Apalachicola Bay, FL using a DNA metabarcoding method we previously developed (Hanif *et al* 2020). MiSeq Illumina high-throughput sequencing was used to analyze DNA amplicons from the V4-V5 region of the 18S ribosomal (rRNA) gene from the stomachs of juvenile menhaden. The fish were collected in May 2013 at Two Mile Channel (TMC), a low salinity estuarine site close to the Apalachicola River estuary, and six weeks later at St Vincent Sound (SVS), a high salinity site at the western end of the enclosed Apalachicola Bay. Although water at the sites varied little with regard to temperature, dissolved oxygen, pH, Secchi measurements, the salinities were very different; 1.8 ppt at TMC and between 36-37.6 ppt at four different sites in SVS. I identified 1048 unique OTUs (species) in the stomach contents, 1035 of which were found in samples from both locations. However, the stomach contents from fish caught at the two locations were very distinct. In stomach samples from Two Mile Channel, members of the Stramenopile/Alveolate/ Rhizaria (SAR) lineage account for 66 % representation and Metazoa at 30 %. The SAR OTUs were dominated by the diatoms *Cyclotella* and *Skeletonema*, as well as the ciliate *Oligotrichia*. In contrast, stomach samples from St. Vincent Sound, members of the Metazoa account for 83.77 % representation and SAR at 15 %. The metazoans are mainly *Acartia* copepods. Since ciliates are considered to be microzooplankton, this means that phytoplankton have just over 60 % representation in samples from TMC and zooplankton have over 90 % representation in samples from SVS. However, it must be remembered that numbers of reads for any species do not correspond to numbers of organisms for a variety of reasons including copy number of 18S rRNA genes which vary across species.

Some of our findings match what is known about plankton communities in Apalachicola Bay. Diatoms are abundant year-round and species that represent a large fraction of net phytoplankton include *Thalassiosira spp*, *Cyclotella* and

*Skeletonema costatum* (Estabrook, 1973). Copepods are the main constituents of the plankton accounting for 80 % of the plankton population (Putland, 2005). Thirty-six species of copepods have been identified from the Apalachicola Bay system with *Acartia tonsa* being the dominant species in every area. *Acartia tonsa* densities averaged over 5,500 numbers per cubic meter copepod naupliar stages found are generally six to 16 times greater than the number of adults. Some species expected were not seen. For instance, forty-two species of fish larvae and thirteen species of planktonic fish eggs have been identified in ichthyoplankton surveys (Blanchet, 1979), although the only fish sequences we found matched most closely with exotic cyprinids. The most abundant species found was the bay anchovy, accounting for over 75 % of all larvae identified and 92 % of all fish eggs collected. Bay anchovy larvae peak during the months of this collection.

Overall, identification of stomach contents by DNA metabarcoding has shown that menhaden sampled at TMC had a diet of mainly phytoplankton while those sampled at SVS had a diet of mainly zooplankton. The fish were an average of 54.1 mm (TL) and 63.4 mm (SL) at TMC and an average of 85.3 mm (TL) and 100.2 mm (SL) (**Supplemental Table 4.4**). Juvenile menhaden are obligate filter feeders. As juveniles they live in tidal creeks, marsh and open bay areas where they filter the water column via their gill rakers. Apalachicola Bay is a productive estuary located in the northern Gulf of Mexico. The high productivity is a result of the Apalachicola River delivering freshwater and nutrients to the bay (Livingston 1984, Mortazavi *et al* 2000a,b, 2001). Apalachicola Bay is a river-dominated system with the major source of freshwater input coming from the Apalachicola River. Maximum river flows occur during late winter to early spring months and are highly correlated with Georgia rainfalls (Meeter 1979). Nutrient input supports high levels of phytoplankton productivity (Mortazavi *et al* 2000b) which in turn supports the Bay's secondary productivity (Chanton & Lewis 2002). The migration pattern of juvenile Gulf menhaden in Apalachicola Bay involves the sequential use of tidal and marsh creeks in early spring, followed by the open bay areas like SVS later in the summer. Food availability increases early in the

year in tidal creeks. High productivity comes from the influx of nutrients with high spring river flow, the flushing of detritus from the river mouth, coupled with warmer temperatures than the open bay area. It has been shown that detrital material can also be used as a food source in juvenile Gulf menhaden using stable isotope evidence (Deegan *et al*, 1990). This results in rapid growth and high survival. However, in mid-summer, when food availability in the tidal creeks begins to decline, the fish move to the open bay area where phytoplankton and zooplankton are increasing.

Prior to our study, the best evidence for dietary shifts in Gulf menhaden came from stable isotope analyses, although such studies cannot identify specific diet items. It has been shown that  $^{13}\text{C}$  enrichment can be used to determine the carbon source in the food web and  $^{15}\text{N}$  enrichment can be used to identify the foraging trophic level (Fry 1988, Vander Zanden *et al*, 1999). The source of the isotopes, therefore, provides insight into temporal, spatial, and ontogenetic variation of the consumer in the local environment. Olsen (Olsen *et al*, 2014) examined stable carbon ( $^{13}\text{C}$ ) and nitrogen ( $^{15}\text{N}$ ) isotope ratios traced through coastal food webs to investigate the trophic level of Gulf menhaden and their role along the Mississippi coast in the northern Gulf of Mexico ecosystem. On the basis of this, Olsen *et al* concluded that the most important dietary item for juvenile (<100 mm total length) fish was phytoplankton (74.0 % dietary composition), while that of subadults (100–200 mm) and adults (>200 mm) was zooplankton (61.6 % for sub-adults and 52.4 % for adults). Juvenile fish also utilized a larger component of terrestrial-based detritus as the source carbon than older sub-adult fish farther from the lower parts of the estuary and offshore. The authors suggested that juvenile menhaden are ‘trophically balanced’ between a phytoplanktivore and zooplanktivore with an opportunistic feeding strategy based on the available food sources but did proportionally consume two to three times more phytoplankton than larger menhaden. Our results suggest a somewhat different picture for juveniles that are essentially opportunists. Gulf menhaden juveniles consume high proportions of zooplankton in mid-summer if they are available.



The Olsen studies noted that these trophic level calculations based on stable isotope analysis are very sensitive to the  $\delta^{15}\text{N}$  baselines used and were careful to compare stable isotope levels of menhaden stomach contents to that of four different size fractions of plankton, as well as stable isotope composition of black needle rush *Juncus roemerianus*, the dominant marsh grass. They found that isotope values for both  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  were significantly different between plankton size fractions. Changes in stable isotope tracers commonly occur across estuarine salinity gradients from freshwater to the sea (Fry, 2002). The tracer gradients reflect the different geochemistries and mixing of freshwater and seawater, and these bottom-up geochemical influences are recorded in estuarine food webs in the isotopic compositions of animals. Watershed-level inputs of freshwater and nutrients can exert strong influences on isotopic values of estuarine consumers, especially filter feeders that largely depend on phytoplankton production. Deviations from conservative isotope mixing can especially with inputs of non-phytoplankton foods such as macrophyte detritus. This means that the measuring of consumer isotopes may only reflect watershed nutrient loading. Consistent with this, Chanton & Lewis (2002) using stable isotope analysis, showed that the estuary consumer diets appeared to vary depending on their position within the estuary and that this was associated with riverine inflows.

Our results show that the menhaden juvenile diet is mainly phytoplankton in fish from TMC in spring and almost entirely zooplankton in fish from SVS in mid-summer. Furthermore, significant abundance analysis shows significant differences in the species represented within many taxa at the two locations. These differences suggest that fish caught at SVS are feeding at a higher trophic level in. The stable isotope analysis shows an increase in  $\delta^{15}\text{N}$  in the muscle of fish from SVS, also supporting a difference in trophic level. However, although suggestive of a trophic shift, the difference in stable isotope composition is not definitive by itself since system-specific processes can alter local isotope values without the corresponding local baseline  $\delta^{15}\text{N}$  values. Local dissolved inorganic nitrogen sources can vary substantially with some upstream sources such as

artificial fertilizers being quite light and some local producers like nitrogen-fixing cyanobacteria being anomalously light as well. However, the results from 18S rRNA gene sequences have provided a more complete description of the diet of fish from the two locations as they move away from the estuary to the enclosed bay, suggesting an early trophic shift in juveniles of this ecologically and economically important fish.

*4.6 Acknowledgements*

This work was supported by the National Oceanic and Atmospheric Administration, Educational Partnership Program award to the Living Marine Resources Cooperative Science Center (LMRCSC) NA11SEC4810002 and NA16SEC4810007. I was supported in part by MD SeaGrant Graduate Student Fellowship, NA10OAR4170072 SA7528129-D and in part as an LMRCSC Graduate Fellow. Thanks, are extended to Dr. Ryan Woodland for discussions on stable isotope analysis.

	<b>TMC (% representation)</b>	<b>SVS (% representation)</b>
Metazoa	30.08	83.77
Stramenopiles	27.5	8.53
Alveolata	38.5	7.2
Rhizaria	0.36	0.09
<b>Total SAR</b>	<b>66.36</b>	<b>15.82</b>
Chloroplastida	0.85	1.05
Fungi	0.74	0.078
Cryptomonadales	2.12	-

**Supplemental Table 4.1: Representation of the most abundant higher taxa from Two Mile Channel and St Vincent Sound samples**

Mean proportional representation of the most abundant higher taxa, are shown as mean percentage reads from TMC and SVS.

SAR	ID	TMC (% representation)	SVS (% representation)
<i>Oligohymenophora</i>	ciliate	1.09	3.31
<i>Mesodiniidae</i>	ciliate	1.42	-
<i>Choreotrichia</i>	ciliate	8.01	18.95
<i>Oligotrichia</i>	ciliate	12.63	5.59
<i>Haptoria</i>	ciliate	7.31	0.079
All ciliates		30.46	27.9
<i>Peridiniopsis</i>	dino	2.7	0.12
<i>Gyrodinium</i>	dino	7.54	0.15
<i>Peridiniales</i>	dino	2.7	0.12
<i>Thoracosphaeraceae</i>	dino	1.64	0.53
All dinos		14.58	0.92
All alveolates		45	28.82
<i>Actinoptychus</i>	diatom	0.039	8.5
<i>Odontella</i>	diatom	0.012	1.25
<i>Uncultured diatom</i>	diatom	3.3	11.27
<i>Thalassiosira</i>	diatom	4.12	12.03
<i>Chaetoceros</i>	diatom	2.85	0.79
<i>Cyclotella</i>	diatom	21.89	25.47
<i>Skeletonema</i>	diatom	10.41	0.23
All diatoms		42.62	59

**Supplemental Table 4.2: Taxonomic composition of the SAR clade at the genus level of stomach contents from Two Mile Channel and St Vincent Sound**

The mean proportional representation of SAR OTUs (species) at the genus level from TMC and SVS sites, as measured by mean percentage reads. Genera are grouped as ciliates, dinoflagellates and diatoms.

Metazoa	ID	TMC (% representation)	SVS (% representation)
<i>Acartia</i>	copepod	7.3	51.2
<i>Balanus</i>	copepod	0.95	4.99
<i>Armatobalanus</i>	copepod		5.2
<i>Taeniacanthus</i>	copepod		4.4
<i>Canuella</i>	copepod	12.22	
<i>Janicella</i>	copepod		
all copepods		21.2	66.4
shrimp		0.03	0.045
teleosts		9.9	10
<i>Neotrigonia</i>	bivalve	7.2	-
<i>Neritina</i>	gastropod	1	-
<i>Barantolla</i>	Annelid:polychaete	20	0.8
<i>Capitella</i>	annelid:polychaete	1.72	0.011
<i>Cirratulus</i>	Annelid:polychaete	2.41	0.028
all polychaetes		24.13	0.839
<i>Carinioma</i>	nemertean	3.34	0.03
<i>Zygeupolia</i>	nemertean	1.05	-
<i>Gastrotricha</i>	nemertean	0.019	0.88
all nemertean worms		4.4 %	0.91 %
<i>Mesodesmo</i>	nematode	0.019	0.39
all rotifers		7.98	0.08
<i>Danio rerio</i>	FW cyprinid from S. Asia	4.73	5.25
<i>Betta pi</i>	FW perciform from Thailand & Malaysia	2.85	2.06
<i>Osteobrama belangeri</i>	FW cyprinid from India and Myanmar	1.53	1.71
<i>Misgurnus anguillicaudatus</i>	FW cyprinid from India & Myanmar	0.86	1.02
all fish		9.97	10.04
other		32.69	22.9

**Supplemental Table 4.3: Taxonomic composition of Metazoa at the genus level of stomach contents from Two Mile Channel and St Vincent Sound**

The mean proportional representation of metazoan OTUs (species) at the genus level from TMC and SVS sites, as measured by mean percentage reads. Genera are grouped as copepods, rotifers, annelids, nematodes and nemerteans, bivalves, arthropods, gastropods and teleosts.

Location	Site	Sample	Date	TL (mm)	SL (mm)	
Two Mile Channel	TMC01	T1	May 2013	48	54	
		T2		55	61	
		T3		61	70	
		T4		55	60	
		T5		53	65	
		T6		50	62	
		T7		54	65	
		T8		49	58	
		T9		71	84	
		T10		45	55	
St. Vincent Sound	SVS02	S1	July 2013	83	103	
		S2		86	97	
		S3		100	106	
	SVS03	S4		106	122	
		SVS04		S5	85	105
				S6	74	88
	S7			70	82	
	SVS05	S8		84	99	
		S9		85	100	
		S10		80	100	

**Supplemental Table 4.4: Length of fish caught at TMC and SVS:**

TMC avg TL = 54.1 mm; TMC avg SL = 63.4 mm;

SVS avg TL = 85.3 mm; SVS avg SL = 100.2 mm

## Chapter 5: Future steps and final thoughts

### *5.1 Limitations*

I have identified over a thousand potential diet items in the stomachs of Gulf menhaden using DNA metabarcoding without even considering any of the prokaryotic species as a food source. This seems like a major leap from describing the menhaden diet as "mainly phytoplankton" or "mainly zooplankton". An ongoing challenge for ecological studies has been the collection of data with high precision and accuracy at a sufficient scale to detect effects relevant to management of critical global change processes (like climate change). If we know that a forage fish like Gulf menhaden operate at the "wasp-waist" of an ecosystem, we need to know what they are eating, but also everything that is above and below. Biased observations of stomach content looking only at a narrow number of species expected to be present or drawing inferences from the small portion of the stomach contents that can be identified visually, are unable to describe the biodiversity of the diet (Lindenmayer & Likens, 2011).

The current study had its limitations. For instance, although I had suitable fish from an interesting, productive estuary, I did not have samples of the water they came from to verify that these filter-feeding fish really did "sample the environment". A collaborator on the "megaproject" helpfully provided me with stable isotope analysis on muscle fillets from the very fish I used. However, I did not have the corresponding data on different size fractions of plankton, or detritus from the river mouth or off the marshes. Although this project on Gulf menhaden represented a large collaboration of interested parties, it foundered somewhat on adequate experimental design. Unfortunately, neither I nor my advisor had been able to give our input on the experimental design. Nevertheless, it is clear that the difference in stomach content in fish from TMC and those from SVS is startling and the percent representation of zooplankton in stomach samples from TMC is unexpectedly compared with older studies. Zooplankton have over 90 % representation in samples from SVS when the microzooplankton is folded in. In the samples from TMC, total zooplankton representation is almost 40 %. Beyond

this, within each major taxa, whatever the representation, the species composition is quite different in the stomachs of fish from the two locations. Overall this is very different from the picture of juveniles consuming two to three times more phytoplankton than larger juveniles that have left the immediate tidal estuary areas as deduced from Olsen's impressive stable isotope analysis (Olsen *et al*, 2014). I am left instead seeing juvenile menhaden as opportunistic feeders and this would seem to be a great strategy for a fish that spends its summer months in an enclosed bay with the possibility of plankton blooms and crashes dependent on factors like the inflow of water from the river.

Apalachicola Bay, as a National Estuarine Research Reserve has been the subject of many research studies from both federal scientist and the faculty and students from surrounding academic institutions. I imagined that in such a well-studied estuary there might have been a comprehensive catalog of picoplankton as well as micro and mesozooplankton with information on seasonal variations. However, such in depth descriptions of the composition of planktonic species over the years and seasons do not seem to be available for Apalachicola Bay as they are for other important water bodies like the Great Lakes.

### 5.2 How should we convert sequence reads to dietary data?

When high throughput sequencing first became available, the potential applications in diet studies were clear and the methods were quickly embraced by the community (Deagle *et al*, 2009; Valentini *et al*, 2009). Studies reported by Deagle *et al* (2019) indicate that using relative read abundance (RRA) can provide an accurate overview of population-level diet, although using read counts as an indication of biomass in samples is more controversial. Relative read abundances are sensitive to DNA recovery biases, differential affinity of some sequences for the "universal" primers used and differences in gene copy number between species (Alberdi *et al*, 2018).

### 5.3 Studying ecosystems with DNA metabarcoding

A major hurdle for many identification workflows is the time-consuming and challenging process of sorting and identifying organisms. The rapid development

of DNA metabarcoding as a biodiversity observation tool provided a potential solution. Due to the high resolution and prey detection capacity DNA metabarcoding has been increasingly used to address ecological questions grounded in dietary relationships. One of the advantages of DNA-based techniques for prey identification purposes is that successful amplification can be achieved in samples that are usually not in optimal condition (i.e. feces and gut contents) as it only requires a small amount of tissue for DNA extraction (Teletchea, 2009). However, although the approach is certainly feasible, the "big data" generated is actually REALLY BIG such that adequate, thoughtful analysis can take a long time. However, to develop a picture of the whole ecosystem and how it can change over the years and seasons, accurate and reliable estimates of biodiversity are essential. This can feed into successful ecosystem management as well as shape environmental policy (Hooper *et al* 2005; Rees *et al* 2004). Monitoring biodiversity using DNA metabarcoding as a long term undertaking by a dedicated team could go a long way to providing this. An estuary such as Apalachicola Bay would be ideal site for this. Estuaries are highly productive systems that provide food, shelter and nursery habitats for greater density, survival rates and growth of juvenile fish (Beck *et al*, 2001; Kraus & Secor, 2005). Such a project could help provide data on the effects of climate change and perhaps help to project its impacts. It could monitor the effects of environmental catastrophes such as the Deepwater Horizon oil spill or follow the impacts of diverting water from the Apalachicola Rivers. It would allow us to ask questions such as: will climate-driven changes in planktonic species composition act as bottom-up regulators of productivity?

#### 5.4 Use of filter feeders as environmental biomonitors in environmental metabarcoding studies

Environmental DNA metabarcoding and high throughput sequencing focuses on detecting organisms from the DNA trace they leave in the environment is a molecular biodiversity assessment method (Taberlet, *et al*, 2012; Thomsen *et al*, 2012). In recent years the use of characterizing this environmental DNA (eDNA)



using high throughput sequencing technology to study marine ecology has greatly increased. Various uses of this method and technology include monitoring marine fish biodiversity, detection of invasive species, effects of introducing or slowing conservation effort, planktonic identification and biomonitoring for rare taxa (Thomsen *et al*, 2012; Bohmann *et al* 2014, Cristescu *et al*, 2018).

There is an increasing amount of work showing that the metabarcoding of eDNA samples is more effective in determining local aquatic community structure than traditional net surveys (Siegenthaler *et al*, 2019; Boussarie *et al*, 2018; Thomsen *et al*, 2012; Valentini *et al*, 2016). Some advantages to this method are the identification of small cryptic or decomposed organisms, reduced cost and effort for analysis, and independence from the developmental stage of organisms (Chariton *et al*, 2015; Hajibabaei *et al*, 2011; Lejzerowicz *et al*, 2015; Leray & Knowlton, 2015). However, one of the drawbacks of identification through DNA metabarcoding can be the inability to determine whether a DNA sequence comes from an egg, larva, juvenile or adult (Valentini *et al* 2016).

Environmental DNA extracted from water samples usually integrates information over large spatial scales but has a low temporal resolution due to the high dispersion and the low persistence of DNA in sea water (Barnes & Turner, 2016; Thomsen *et al.*, 2012). There has been increasing use of filter feeding species as environmental samplers constituting a valuable and effective method for biomonitoring. Leeches and carrion flies have been used as biodiversity sampling tools in studies examining mammal biodiversity in terrestrial habitats (Calvignac-Spencer *et al*, 2013; Schnell *et al*, 2015, 2012). In the marine environment shrimp and sponges have been used to determine marine biodiversity (Siegenthaler *et al*, 2019; Mariani *et al*, 2019). In our study looking at the 18S rRNA gene sequencing results in Gulf menhaden, we find organisms that have not been identified in previous studies. For example, the shrimp, bivalve, and worm genera have not previously been reported in the stomach contents of menhaden. Given that the adult organisms are too large to be filtered via menhaden gill rakers, we conclude that some evidence of them (i.e., eggs,

larvae, or decaying body part) is present in the water as part of marine snow or detritus. For this reason, we propose that menhaden should be added to the list of environmental biomonitors, using stomach contents as a proxy for ecosystem health. However, more research with menhaden is needed to verify that their stomach content accurately reflects what is in the water column. The area “sampled” by a school of menhaden will likely provide information on a larger spatial scale than acquired by a water samples since the fish actively move around and shows seasonal movements. The “sampled biodiversity” is naturally encapsulated in the stomachs of menhaden, from netting the fish all the way to DNA extraction in the laboratory. It represents a significant way to streamline and by-pass many of the fastidious steps required to reduce degradation and contamination when sampling water; a fact that is often understated in eDNA research.

#### *5.5 Accounting for functional ecological importance/significance of the stomach microbiota*

Microbial ecology using DNA metabarcoding focuses on identification of bacterial species but does not, by itself, provide insight of their function. There are even fewer studies on the diversity of microeukaryotic organisms such as diatoms, dinoflagellates and ciliates and their function. In the face of climate change more work needs to be done to understand how changing environments will affect the marine microbiota and how that will reflect marine processes, as well as how changes in the stomach microbiota of fish and other aquatic organisms. Will this result in a loss or gain in function? How will organisms cope with new microbial residence? Could new microbial residences aid with the adjustment in a new climate in terms of heat stress, disease, salinity changes, pH changes, etc.

Given that fish have a more intimate interaction with the surrounding microbiota than land animals, interactions between the aquatic microbial community and the fish microbiota need further attention. My study saw differences in the stomach microbiota with respect to sampling location. Although

profiling phylogenetic marker genes, such as the 16S rRNA gene, is a key tool for studies of microbial communities, it does not provide direct evidence of metabolic or functional capabilities of a microbial community. However, there is a computational approach to predict the functional composition of a microbial community from the OTUs found and a database of reference genomes. PICRUSt uses an extended ancestral-state reconstruction algorithm to predict which gene families are present and can combine gene families to estimate the composite functional capacity of a community (Langille *et al*, 2013; Douglas *et al*, 2018). Phylogenetic trees based on 16S rRNA sequences closely resemble clusters obtained based on shared gene content, and researchers often infer properties of uncultured organisms from cultured relatives. This 'predictive metagenomic' approach could provide useful insights into the thousands of uncultivated microbial communities for which only marker gene surveys are currently available.

#### 5.6 The evolution of DNA metabarcoding analysis methods

One limitation to DNA metabarcoding is simply the amount of data needed to be analyzed from a high throughput sequencing run. Depending on the number of samples reads can easily be in the order of  $1 \times 10^6$ . This poses many challenges in trying to process this data. This is further complicated by the unconventional type of data in that often is highly dimensional with the number of taxa much greater than the number of samples. This leads to many debates on best practices when analyzing such data and what inferences can be made from the subsequent results, for example the use of model-based for abundance analysis, when to rarefy, the use of normalization and differential abundance strategies, and taxonomic assignment (Tsilimigras & Fodor 2016; McMurdie & Holmes 2014; Weis *et al*, 2017; Hui 2016; White *et al*, 2009; Angiuoli *et al*, 2011, Caporaso *et al*, 2010). As new methods for analysis are constantly being developed it may be difficult for the analyzer to keep up. For example, during the course of this study several updates and methods for post sequencing analysis as well improved taxonomic assignment became available (Callahan *et al*, 2016).

As these new and improved methods become available more insight can be gained from the same data. Therefore, it is imperative that sequence data be securely stored with services such as those offered by NCBI.

A continuing limitation of high throughput sequencing is taxonomic database curation. Given that taxonomic assignment is based on a curated database the accuracy of the assignment is only as good as the database itself. Two issues arise when using a database. The first is how the taxa are ranked and the other is the completeness of the database. The taxonomic ranks of organisms can change with new research of that particular organism. This can pose an issue when looking at previous studies. For example, the SILVA taxonomic database used to assign taxonomy was updated during this study. In doing so, new taxa were added and some the taxonomic ranks were adjusted. This update led to the ability to determine the sequence of menhaden 18S rRNA which has been deposited in NCBI's GeneBank. This in turn allowed me to determine the amount of menhaden DNA contamination in the stomach samples. It is important to maintain a well curated taxonomic database as the use of DNA metabarcoding in biodiversity studies are going to increase. Perhaps for the investigator one way around this would be to develop a database specific to his/her study, an approach not without its own problems.

## Bibliography

- Able, KW. 2005. A re-examination of fish estuarine dependence: evidence for connectivity between estuarine and ocean habitats. *Estuarine, Coastal and Shelf Science* **64**: 5–17.
- Ahrenholz, DW. 1981. Recruitment and exploitation of Gulf menhaden, *Brevoortia patronus*. *Fish Bull US* **79**: 325–36.
- Ahrenholz, DW. 1991. Recruitment and exploitation of Gulf menhaden, *Brevoortia patronus*. *Fishery Bulletin-United States*, **53**, 3-19.
- Alder, J, Campbell, B, Karpouzi, V, Kaschner, K, & Pauly, D. 2008. Forage fish: from ecosystems to markets. *Annual Rev Environ & Resources* **33**: 153.
- Amann, RI, Ludwig, W, & Schleifer, KH 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143–69.
- Angiuoli et al 2011
- Anderson, C, & Cabana, G. 2006. Does  $\delta^{15}\text{N}$  in river food webs reflect the intensity and origin of N loads from the watershed. *Sci Total Environ*, **367**(2-3), 968-978.
- Anderson, JD. 2007. Systematics of the North American menhadens: molecular evolutionary reconstructions in the genus *Brevoortia* (Clupeiformes: Clupeidae). *Fishery Bulletin* **105**: 368–78.
- Arthur, SC 1931. The birds of Louisiana. Department of Conservation.
- Ashelford, KE, Weightman, AJ & Fry, JC. 2002. PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. *Nucleic Acids Res* **30**: 3481-3489.
- Baker, R, Buckland, A, & Sheaves, M. 2014. Fish gut content analysis: robust measures of diet composition. *Fish and Fisheries* **15**: 170–77.
- Bakun, A, Babcock, EA, & Santora, C. 2009. Regulating a complex adaptive system via its wasp-waist: grappling with ecosystem-based management of the New England herring fishery. *ICES J Mar Sci* **66**: 1768–75.
- Bargiela R, Gertler, C, Magagnini, M., Mapelli, F, Chen, J, Daffonchio, D, Golyshin, PN, & Ferrer, M. 2015. Degradation network reconstruction in uric acid & ammonium amendments in oil-degrading marine microcosms guided by metagenomic data. *Front Microbiol* **6**: 1270.
- Barnes, MA, & Turner, CR. 2016. The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, **17**, 1–17.
- Beck, MW, Heck, KI, Able, KW, Childers, DI, Eggleston, DB, Gillanders, B, Weinstein, MP (2001). The identification, conservation, and management of

estuarine and marine nurseries for fish and invertebrates. *BioScience*, 51, 633–641.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., & Kauserud, H. 2010. ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases. *BMC Microbiol* 10: 1.

Berthold, M. R., Cebren, N., Dill, F., Gabriel, T. R., Kotter, T., Meinl, T., Ohl, P., Thiel, K., & Wisdel, B. 2009. KNIME: the Konstanz information miner. *SIGKDD Explor Newsl* 11: 26–31.

Blanchet, RH 1978. The distribution and abundance of ichthyoplankton in the Apalachicola Bay, Florida, area. M.S. Thesis. Florida State University, Tallahassee.

Blankenship, LE, & Yayanos, AA. 2005. Universal primers and PCR of gut contents to study marine invertebrate diets. *Mol Ecol* 14: 891–99.

Blankenship, L., & Levin, LA. (2007). Extreme food webs: Foraging strategies and diets of scavenging amphipods from the ocean's deepest 5 kilometers. *Limnology and Oceanography*, 52(4), 1685-1697.

Bohmann, K., Monadjem, A., Lehmkuhl Noer, C., Rasmussen, M., Zeale, M. R., Clare, E. *et al.* (2011). Molecular diet analysis of two African free-tailed bats (Molossididae) using high throughput sequencing. *PLoS One*, 6, e21441.

Bohmann, K, Evans, A, Gilbert, P, Thomas, P, Carvalho, GR, Creer, S, Knapp, D, Yu, W, Bruyn, M. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution* 29: 358-367.

Bohmann *et al* 2017. Next-Generation Global Biomonitoring: Large-scale, 472 Automated Reconstruction of Ecological Networks. *Trends in Ecology & Evolution*.

Bolyen E, Rideout, JR, Dillon, MR, Bokulich NA, Abnet, CC, Al-Ghalith, GA, Alexander, H, Alm, EJ, Arumugam, M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhoer C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciulek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Lofffield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS 2nd, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, UI-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Iber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R

- Boom, R., Sol, C. J., Salimans, M. M., Jansen, C. L., Irtheim-van Dillen, P. M., & van der Noordaa, J. 1990. Rapid and simple method for purification of nucleic acids. *J Clin Microbiol* **28**: 495–503.
- Bradley, I. M., A. J. Pinto, and J. S. Guest. 2016. Design and evaluation of Illumina MiSeq-Compatible, 18S rRNA gene-specific primers for improved characterization of mixed phototrophic communities. *Appl Environ Microbiol* **82**: 5878-5891.
- Brown-Peterson, N. J., Leaf, R. T., Schueller, A. M., & Andres, M. J. 2017. Reproductive dynamics of gulf menhaden (*Brevoortia patronus*) in the northern Gulf of Mexico: effects on stock assessments. *Fishery Bulletin* **115**: 284–99.
- Bundy, A., Boldt, J. L., & Cook, A. M. 2012. Relative importance of fisheries, trophodynamic and environmental drivers in a series of marine ecosystems. *Mar. Ecol.*
- Bush, A, Compson, ZG, Monk, W, Porter, TM, Steves, R, Emilson, E, Gagne, N, Hajlbaael, M, Roy, M & Baird, DJ. 2019 Studying ecosystems with DNA metabarcoding: lessons from aquatic macroinvertebrates. *Front Ecol*, 10.3389. <https://www.frontiersin.org/articles/10.3389/fevo.2019.00434/full>
- Caporaso JG. 2019 Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**:852-857.
- Callahan, J. Benjamin, McMurdie, J. Paul, Rosen, J. Michael, Han, W. Andrew, Johnson, Jo. A. Amy, Holmes, P. Susan. 2016. DADA2: High-Resolution sample inference from Illumina amplicon data. *Nature Methods* **13**: 281-583.
- Calvignac-Spencer, Leendertz, FH, Gilbert, MT, & Schubert, G. 2013. An invertebrate stomach's view on vertebrate ecology: Certain invertebrates could be used as “vertebrate samplers” and deliver DNA-based information on many aspects of vertebrate ecology. *BioEssays*, **35**: 1004-1013.
- Caporaso JG, Kuczynski, J, Stombaugh, J, Bittinger, K, Bushman, FD, Costello, EK, Fierer, N, Pena, AG, Goodrich, JK, & Gordon, JI. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335–36.
- Carreon-Martinez, L, & Heath, DD. 2010. Revolution in food web analysis and trophic ecology: diet analysis by DNA and stable isotope analysis. *Mol Ecol* **19**: 25–2.
- Carreon-Martinez, L, Johnson, TB, Ludsins, SA, & Heath, DD. 2011. Utilization of stomach content DNA to determine diet diversity in piscivorous fishes. *J Fish Biol* **78**: 1170–82.
- Chanton, JP. & Lewis, FG. 1999. Plankton and dissolved inorganic carbon isotopic composition in a river-dominated estuary: Appalachicola Bay, Florida. *Estuaries* **22**:575-583.

- Chariton, AA, Stephenson, S, Morgan, MJ, Steven, AD, Colloff, MJ, Court, LN & Hardy, CM. 2015. Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. *Env Pollution*, 203: 165-174.
- Christmas, J. Y., & Gunter, G. 1960. Distribution of menhaden, genus *Brevoortia*, in the Gulf of Mexico. *Transactions of the American Fisheries Society* **89**: 338-43.
- Christmas, JY., McBee, JT., Waller, RS., and Sutter, FC. 1982. Habitat Suitability Index Models: Gulf Menhaden. U.S. Department of the Interior Fish and Wildlife Service, FWS/OBS-82/10.23. p. 31.
- Christmas, J., & Waller, RS. 1975. Location and time of menhaden spawning in the Gulf of Mexico. Gulf Coast Research Laboratory.
- Claassen, S., duToit, E., Kaba, M., Moodley, C., Zar, H. J., AND Nicol, M. P. 2013. A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples. *Journal of microbiological methods* **94**: 103–10.
- Cleary, AC, Durbin, EG, & Ryneerson, TA. 2012. Krill feeding on sediment in the Gulf of Maine (North Atlantic). *Mar Ecol Progr Ser* **455**: 157–72.
- Clements, K. D., Angert, E. R., Montgomery, W. L., AND Choat, J. H. 2014. Intestinal microbiota in fishes: what's known and what's not. *Molecular ecology* **23**: 1891–98.
- Clooney AG, Fouhy F, Sleator RD, O' Driscoll A, Stanton C, Cotter PD, Claesson MJ. 2016 Comparing Apples and Oranges: Next Generation Sequencing and Its Impact on Microbiome Analysis *PLoS One*, e0148028.
- Colston TJ, & Jackson, CR. 2016. Microbiome evolution along divergent branches of the vertebrate tree of life: what is known & unknown. *Molecular ecology* **25**: 3776–800.
- Combs, R. M. 1969. Embryogenesis, Histology, and Organology of the Ovary of *Brevoortia patronus*. *Gulf and Caribbean Research* **2**: 333–434.
- Corse, E., Megléc, E., Archambaud, G., Ardisson, M., Martin, J. F., Tougard, C., Chappaz, R., & Dubut, V. 2017. A from-benchttop-to-desktop workflow for validating HTS data and for taxonomic identification in diet metabarcoding studies. *Mol Ecol Resour* **17**: e146–59.
- Craig, C, Kimmerer, WJ, Cohen, CS. 2013. A DNA-based method for investigating feeding by copepod nauplii *J. Plankton Res.* (2014) **36**(1): 271–275.
- Cristescu, Melania E. and Hebert, Paul D. N., Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation 2018. *Annual Review of Ecology, Evolution, and Systematics*. 49:209-230.
- Cury, P., Bakun, A., Crawford, R. J. M., Jarre, A., Quiñones, R. A., Shannon, L. J., & Verheye, H. M. 2000. Small pelagics in upwelling systems: patterns of interaction and structural changes in wasp-waist ecosystems. *ICES Journal of Marine Science: Journal du Conseil* **57**: 603–18.



- Dahlberg, M. D. 1970. Atlantic and Gulf of Mexico menhadens, genus *Brevoortia* (Pisces: Clupeidae). Ph.D. thesis, University of Florida.
- Dailey et al, 2008
- Deagle, B. E., Tollit, D. J., Jarman, S. N., Hindell, M. A., Trites, A. W., & Gales, N. J. (2005). Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. *Mol Ecol*, *14*(6), 1831-1842.
- Deagle, B. E., Gales, N. J., Evans, K., Jarman, S. N., Robinson, S., Trebilco, R. et al. (2007). Studying seabird diet through genetic analysis of faeces: a case study on macaroni penguins (*Eudyptes chrysolophus*). *PLoS One*, *2*(9), e831.
- Deagle, B. E., Kirkwood, R., & Jarman, S. N. (2009). Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Mol Ecol*, *18*(9), 2022-2038.
- Deagle, B. E., Chiaradia, A., McInnes, J., & Jarman, S. N. (2010). Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out. *Conservation Genetics*, *11*(5), 2039-2048.
- Deagle, B. E., Thomas, A. C., Shaffer, A. K., Trites, A. W., & Jarman, S. N. (2013). Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count. *Mol Ecol Resour*, *13*(4), 620-633.
- Deagle, BE, Jarman, SN, Coissac, E, Pompanon, F, & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biol Lett*, *10*(9).
- Deagle, B. E., Clarke, L. J., Kitchener, J. A., Polanowski, A. M., & Davidson, A. T. (2018). Genetic monitoring of open ocean biodiversity: An evaluation of DNA metabarcoding for processing continuous plankton recorder samples. *Mol Ecol Resour*, *18*(3), 391-406.
- Deagle, BE, Thomas, AC, McInnes, JC, Clarke, LJ., Vesterinen, EJ., Clare, EL. et al. (2019). Counting with DNA in metabarcoding studies: How should I convert sequence reads to dietary data. *Mol Ecol*, *28*(2), 391-406.
- Deegan, LA. 1986. Changes in body composition and morphology of young-of-the-year gulf menhaden, *Brevoortia patronus* Goode, in Fourleague Bay, Louisiana. *Journal of Fish Biology* **29**: 403–15.
- Deegan, LA 1993. Nutrient and energy transport between estuaries and coastal marine ecosystems by fish migration. *Canadian Journal of Fisheries and Aquatic Sciences* **50**: 74–79.
- Deegan, LA, Peterson, BJ, & Portier, R. 1990. Stable isotopes & cellulase activity as evidence for detritus as a food source for juvenile Gulf menhaden. *Estuaries* **13**: 14–19.
- Deegan, LA, & Thompson, BA. 1987. Growth rate and life history events of Young-of-the-Year Gulf menhaden as determined from otoliths. *Transacs American Fish Soc*, *116*: 663-667.

- Degnan, PH, & Ochman, H. 2016. Illumina-based analysis of microbial community diversity. *ISME J* **6**: 183–94.
- De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. 2014. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Mol Ecol Resour* **14**: 306–23.
- Di Gennaro P., Terreni, P., Masi, G., Botti, S., De Ferra, F., & Bestetti, G. 2010. Identification & characterization of genes involved in naphthalene degradation in *Rhodococcus opacus* R7. *Applied Microbiol & Biotechnol* **87**: 297–308.
- Djurhuus, A., K. Pitz, N. A. Sawaya, J. Rojas-Márquez, B. Michaud, E. Montes, F. Muller-Karger, & M. Breitbart. 2018. Evaluation of marine zooplankton community structure through environmental DNA metabarcoding. *Limnol Oceanogr Methods* **16**: 209-221.
- Douglas, GM, Belko, RG, Langille, MG. 2018. Predicting the Functional Potential of the Microbiome from Marker Genes Using PICRUSt. *Methods Mol Biol*, 1849169-177.
- Dunshea, G., Barros, N. B., Ills, R. S., Gales, N. J., Hindell, M. A., & Jarman, S. N. (2008). Pseudogenes and DNA-based diet analyses: a cautionary tale from a relatively ill sampled predator-prey system. *Bull Entomol Res*, *98*(3), 239-248.
- Durbin A. G., & Durbin, E. G. 1975. Grazing rates of the Atlantic menhaden, *Brevoortia tyrannus*, as a function of particle size & concentration. *Mar Biol* **33**: 265–77.
- Edgar, R. C. 2010. Search & clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–61.
- Egerton S., Culloty, S., Whooley, J., Stanton, C., & Ross, R. P. 2018. The Gut Microbiota of Marine Fish. *Front Microbiol* **9**: 873.
- Eichmiller JJ, Miller LM, Sorensen PW. 2016. Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Mol Ecol Resour*. *16*:56-68.
- Essington, T. E., Moriarty, P. E., Froehlich, H. E., Hodgson, E. E., Koehn, L. E., Oken, K. L., Siple, M. C., AND Stawitz, C. C. 2015. Fishing amplifies forage fish population collapses. *Proc Natl Acad Sci U S A* **112**: 6648–52.
- Estabrook, R. H. (1973). *Phytoplankton of Apalachicola Bay*. Ph.D thesis, Florida State University.
- Etzold, D. J., and J. Y. Christmas. 1979. A Mississippi marine finfish management plan. MS-AL Sea Grant Consortium MASGP-78-046 1–12.
- Fadrosh D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* **2**: 6.

- Felske, A., & Akkermans, A. D. L. (1998). Spatial homogeneity of abundant bacterial 16S rRNA molecules in grassland soils. *Microbial Ecology*, 36(1), 31-36.
- Fierer, EK, Pena, N, Goodrich, AG, & Gordon, J. I. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7: 335–36.
- Fillbrunn, A., Dietz, C., Pfeuffer, J., Rahn, R., Landrum, G. A., & Berthold, M. R. 2017. KNIME for reproducible cross-domain analysis of life science data. *J Biotechnol* 261: 149–56.
- Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3: 294–99.
- Fore, P. L. 1970. Oceanic distribution of eggs and larvae of the Gulf menhaden.
- Forin-Wiart, M. A., Poulle, M. L., Piry, S., Cosson, J. F., Larose, C., & Galan, M. 2018. Evaluating metabarcoding to analyse diet composition of species foraging in anthropogenic landscapes using Ion Torrent and Illumina sequencing. *Sci Rep* 8: 17091.
- Franklin, 2007
- Franklin, H. B. 2011. The most important fish in the sea: Menhaden and America. Island Press, 280 pp.
- Friedland, K. D., Haas, L. W., & Merriner, J. V. 1984. Filtering rates of the juvenile Atlantic menhaden *Brevoortia tyrannus* (Pisces: Clupeidae), with consideration of the effects of detritus & swimming speed. *Marine Biology* 84: 109–17.
- Friedland, K. D., AND Haas, L. W. 1988. Emigration of Juvenile Atlantic Menhaden, *Brevoortia tyrannus* (Pisces, Clupeidae), from the York River Estuary. *Estuaries* 11: 45–50.
- Friedland, K. D., Ahrenholz, D. W., Smith, J. W., Manning, M., & Ryan, J. 2006. Sieving functional morphology of the gill raker feeding apparatus of Atlantic menhaden. *J Exp Zool A Comp Exp Biol* 305: 974–85.
- Fry, B. 1988. Food web structure on Georges Bank from stable C, N, and S isotopic compositions. *Limnology and Oceanography* 33:1182–1190.
- Fry, B 2002 Conservative Mixing of Stable Isotopes Across Estuarine Salinity Gradients: A Conceptual Framework for Monitoring Watershed Influences on Downstream Fisheries Production. *Estuaries*, 25: 264-271.
- Gamboa-Delgado, J., Canavate, J. P., Zerolo, R., & Vay, L. (2008). Natural carbon stable isotope ratios as indicators of the relative contribution of live and inert diets to growth in larval Senegalese sole (*Solea senegalensis*). *Aquaculture*.
- Geers, T. M., Pikitch, E. K., AND Frisk, M. G. 2016. An original model of the northern Gulf of Mexico using Ecopath with Ecosim and its implications for the effects of fishing on ecosystem structure and maturity. *Deep-Sea Res II* 129: 319–31.

- Gertler C., Bargiela, R., Mapelli, F., Han, X., & Chen, J. 2015. Conversion of uric acid into ammonium in oil-degrading marine microbial communities: a possible role of Halomonads. *Microbial Ecol* **70**: 724–40.
- Ghanbari M., Kneifel, W., & Domig, K. J. 2015. A new view of the fish gut microbiome: advances from next-generation sequencing. *Aquaculture* **448**: 464–75.
- Gläsner, A. H., Reischl, J., AND Gessner, U. 2015. Analyses of Intestinal Microbiota: Culture versus Sequencing. *ILAR J* **56**: 228–40.
- Givens C. E., Ransom, B., Bano, N., & Hollibaugh, J. T. 2015. Comparison of the gut microbiomes of 12 bony fish & 3 shark species. *Mar Ecol Progr* **518**: 209–23.
- Gołębiewski, M., AND Tretyn, A. 2019. Generating amplicon reads for microbial community assessment with next-generation sequencing. *J Appl Microbiol* doi: 10.1111.
- Goode, G.B. 1878. A revision of the American species of the genus *Brevoortia*, with a description of a new species from the Gulf of Mexico. *Proceedings of the U.S. National Museum* 1:30-42.
- Gossling, J., & Slack, J. M. 1974. Predominant gram-positive bacteria in human feces: numbers, variety, and persistence. *Infect Immun* **9**: 719–29.
- Grabowski, P., & Rappsilber, J. 2019. A Primer on Data Analytics in Functional Genomics: How to move from data to insight. *Trends Biochem Sci* **44**: 21–32.
- Greenstone, M. H., Iber, D. C., Coudron, T. A., Payton, M. E., AND Hu, J. S. 2012. Removing external DNA contamination from arthropod predators destined for molecular gut-content analysis. *Mol Ecol Resour* **12**: 464–69.
- Gu, B., Schell, D. M., & Huang, X. (1996). Stable isotope evidence for dietary overlap between two planktivorous fishes in aquaculture ponds. *Can J.*
- Gutierrez T., Singleton, D. R., Berry, D., Yang, T., Aitken, M. D., & Teske, A. 2013. Hydrocarbon-degrading bacteria enriched by the Deepwater Horizon oil spill identified by cultivation & DNA-SIP. *The ISME journal* **7**: 2091.
- Gunter, G. 1945. *Studies on marine fishes of Texas*. Austin, Publications of the Institute of Marine Science, University of Texas, Austin, Texas. 1(1):1-190.
- Hajibabaei *et al*, 2011. Assessing biodiversity of a freshwater benthic 568 macroinvertebrate community through non-destructive environmental barcoding of DNA 569 from preservative ethanol.
- Hanif A., White, J., Place, A. R., & Jagus, R. 2020. Methodology for the identification of stomach contents in the filter feeding fish (*Brevoortia patronus*) using DNA metabarcoding. *Limnol & Oceano Methods* submitted
- Han, Z., Sun, J., Lv, A., & Wang, A. 2019. Biases from different DNA extraction methods in intestine microbiome research based on 16S rDNA sequencing: a case in the koi carp, *Cyprinus carpio* var. Koi. *Microbiology*, **8**: e00626.

- Harms-Tuohy, C. A., Schizas, N. V., & Appeldoorn, R. S. (2016). Use of DNA metabarcoding for stomach content analysis in the invasive lionfish *Pterois volitans* in Puerto Rico. *Marine Ecology Progress Series*, 558, 181-191.
- Harper, G. L., Sheppard, S. K., Harwood, J. D., Read, D. S., Glen, D. M., Bruford, M. W. et al. (2006). Evaluation of temperature gradient gel electrophoresis for the analysis of prey DNA within the guts of invertebrate predators. *Bulletin of Entomological Research*, 96(3), 295-304.
- Harris, T. W., Lee, R., Schwarz, E., Bradnam, K., Lawson, D., Chen, W. et al. (2003). WormBase: a cross-species database for comparative genomics. *Nucleic Acids Res*, 31(1), 133-137.
- Hiergeist, A., Gläsner, J., Reischl, U., & Gessner, A. (2015). Analyses of intestinal microbiota: culture versus sequencing. *ILAR journal*, 56(2), 228-240.
- Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G. D., Lewis, J. D., & Bushman, F. D. 2013. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. *PLoS One* 8: e66019.
- Holechek, J. L., Vavra, M., AND Pieper, R. D. 1982. Botanical composition determination of range herbivore diets: a review. *Journal of Range Management* 309–15.
- Hooper et al, 2005
- Houde, E. D., & Fore, P. L. 1973 Guide to the identity of eggs and larvae of some Gulf of Mexico clupeid fishes. Florida Department of Natural Resources, Marine Research Laboratory Leaflet Series 4, (Part1, No. 23), 14 pp.
- Hugerth, L. W., Muller, E. E. L., Hu, Y. O. O., Lebrun, L. A. M., Roume, H., Lundin, D., Wilmes, P., & Andersson, A. F. 2014. Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. *PLoS One* 9: e95567.
- Hui 2016
- Hyslop, E. J. 1980. Stomach contents analysis-a review of methods and their application. *J Fish Biol* 17: 411–29.
- Ingerson-Mahar, J. 2002. Relating diet and morphology in adult carabid beetles. *The Agroecology of Carabid Beetles* 111–36.
- Illumina Document #15027987v1
- Jakubavičiūtė, E., Bergström, U., Eklöf, J. S., Haenel, Q., & Bourlat, S. J. 2017. DNA metabarcoding reveals diverse diet of the three-spined stickleback in a coastal ecosystem. *PLoS One* 12: e0186929.
- Jami M., Ghanbari, M., Kneifel, W., & Domig, K. J. 2015. Phylogenetic diversity & biological activity of culturable Actinobacteria isolated from freshwater fish gut microbiota. *Microbiol Res* 175: 6–15.

- Jarman, S. N., Deagle, B. E., & Gales, N. J. (2004). Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. *Molecular Ecology*, 13:1313-1322.
- Jarman, S. N., Redd, K. S., & Gales, N. J. 2006. Group-specific primers for amplifying DNA sequences that identify Amphipoda, Cephalopoda, Echinodermata, Gastropoda, Isopoda, Ostracoda and Thoracica. *Molecular Ecology Notes* 6: 268–71.
- Jiang, B., Sun, J., Lv, A., Hu, X., Shi, H., Sung, Y., Wang, Q., & Wang, Y. 2019. Impact of DNA extraction methods on the observed microbial communities from the intestinal flora of the penaeid shrimp *Litopenaeus vannamei*. *FEMS Microbiol Lett* 366: 10.1093.
- Jo, H., Ventura, M., Vidal, N., & Gim, J. S. (2016). Discovering hidden biodiversity: the use of complementary monitoring of fish diet based on DNA barcoding in freshwater ecosystems. *Ecol & Evol* 6: 219-232.
- Jordan et al, 2005
- Judy, MH, & Lewis, R. 1983. Distribution of eggs and larvae of Atlantic menhaden, *Brevoortia tyrannus*, along the Atlantic coast of the United States. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service.
- June, FC, & Carlson, FT 1971. Food of young Atlantic menhaden, *Brevoortia tyrannus*, in relation to metamorphosis. *Fish Bull* 68: 493–512.
- Kendall, A. W., AND Reintjes, J. W. 1975. GEOGRAPHIC AND HYDROGRAPHIC DISTRIBUTION OF ATLANTIC MENHADEN EGGS AND LARVAE ALONG MIDDLE ATLANTIC COAST FROM RV DOLPHIN CRUISES, 1965-66. *Fishery Bulletin* 73: 317–35.
- Ketchum, R. N., Dieng, M. M., Vaughan, G. O., Burt, J. A., & Idaghdour, Y. 2016. Levels of genetic diversity and taxonomic status of *Epinephelus* species in United Arab Emirates fish markets. *Mar Pollut Bull* 105: 540–45.
- King, R. A., Read, D. S., Traugott, M., & Symondson, W. O. (2008). Molecular analysis of predation: a review of best practice for DNA-based approaches. *Mol Ecol*, 17(4), 947-963.
- King G. M., Judd, C., Kuske, C. R., & Smith, C. 2012. Analysis of stomach & gut microbiomes of the eastern oyster (*Crassostrea virginica*) from coastal Louisiana, USA. *PLoS One* 7: e51475.
- King, G. M., Smith, C., Tolar, B., & Hollibaugh, J. T. (2013). Analysis of composition and structure of coastal to mesopelagic bacterioplankton communities in the northern Gulf of Mexico. *Front Microbiol.* 3: 438.
- Klindworth, A., E. Pruesse, T. Schler, J. Peplies, C. Quast, M. Horn, & F. O. Glöckner. 2012. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research* 41: 1-11.

Koo *et al*, 2015

Kostka J. E., Prakash, O., Overholt, W. A., Green, S. J., Freyer, G., Canion, A., Delgardio, J., Norton, N., Hazen, T. C., & Huettel, M. 2011. Hydrocarbon-degrading bacteria & the bacterial community response in Gulf of Mexico beach impacted by the Deepwater Horizon oil spill. *Appl Environ Microbiol* **77**: 7962–74.

Kraus, RT, & Secor, D. H. 2005. Application of the nursery-role hypothesis to an estuarine fish. *Marine Ecology Progress Series* **291**: 301.

Kroger, R. L., AND Guthrie, J. F. 1973. Migrations of tagged juvenile Atlantic menhaden. *Transactions of the American Fisheries Society* **102**: 417–22.

Lagier, J.-C., Million, M., Hugon, P., Armougom, F., & Raoult, D. 2012. Human gut microbiota: repertoire and variations. *Front Cell Infect Microbiol* **2**: 136.

Langille, M. G. and others. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**: 814–21.

Lewis, V. P., & Peters, D. S. (1994). Diet of juvenile and adult Atlantic menhaden in estuarine and coastal habitats. *Transactions of the American Fisheries Society*, *123*(5), 803-810.

Lassuy, 1 D. R. 1983. Species profiles-life histories & environmental requirements (Gulf of Mexico): Gulf menhaden. repository.stdlorg.

Lejzerowicz, F, Esling, P, Pillet, L, Wilding, TA, Black & Pawlowski, J. 2015. High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. *Sci Report*, 5: 13932.

Leray, M, & Knowlton, N. 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *PNAS* **112**: 2076–81.

Leray, M, Yang, JY, Meyer, CP, Mills, SC, Agudelo, N, Ranlz, V, Boehm, JT & Machida, RJ. 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front Zool* *10*: 34.

Lessa, EP, & Applebaum, G. (1993). Screening techniques for detecting allelic variation in DNA sequences. *Molecular ecology*, *2*, 119-129.

Lewis, RM, & Roithmayr, CM. 1981. Spawning and sexual maturity of gulf menhaden, *Brevoortia patronus*. *Fishery Bulletin* *78*(4):947-951.

Lewis VP & Peters, DS 1994. Diet of juvenile & adult Atlantic menhaden in estuarine & coastal habitats. *Transactions of the American Fisheries Society* **123**: 803–10.

Li J, Lawson Handley LJ, Read DS, Hänfling B. 2018. The effect of filtration method on the efficiency of environmental DNA capture and quantification via metabarcoding. *Mol Ecol Resour*. doi: 10.1111/1755-0998.12899.

- Liljestrand, EM, Wilberg, MJ, AND Schueller, AM. 2019. Estimation of movement and mortality of Atlantic menhaden during 1966–1969 using a Bayesian multi-state mark-recovery model. *Fisheries Research* **210**: 204–13.
- Lindeque, PK, Parry, HE, Harmer, RA., Somerfield, PJ, & Atkinson, A. 2013. Next generation sequencing reveals the hidden diversity of zooplankton assemblage. *PLoS One* **8**: e81327.
- Lindenmeyer, D. B., AND Likens, G. E. 2011. Direct measurement versus surrogate indicator species for evaluating environmental change and biodiversity loss. *Ecosystems*,
- Livingston, RJ, X. Niu, GF Lewis III, & Woodsum, GC. 1997. Freshwater input to a Gulf estuary: Long-term control of trophic organization. *Ecological Applications* **7**:277-299.
- Livingston, RJ. 2007. Phytoplankton bloom effects on a gulf estuary: water quality changes and biological response. *Ecological Applications* **17**: S110–28
- Lochman, R, & Phillips, H. (1996). Stable isotopic evaluation of the relative assimilation of natural and artificial foods by golden shiners *Notemigonus crysoleucas* in ponds. *Journal of the World Aquaculture* **27**:168-177.
- Lozupone, C. A., M. Hamady, S. T. Kelley, and R. Knight. 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* **73**: 1576-1585.
- Lynch, PD, Brush, MJ, Condon, ED, & Latour, RJ. 2010. Net removal of nitrogen through ingestion of phytoplankton by Atlantic menhaden *Brevoortia tyrannus* in Chesapeake Bay. *Marine Ecology Progress Series* **401**: 195–209.
- Majaneva, M, Diserud, OH, Eagle, SHC, Boström, E, Hajibabaei, M, Ekrem T. 2018 Environmental DNA filtration techniques affect recovered biodiversity. *Sci Rep* **8**:4682.
- Mariani, S, Baillie, C, Colosimo, G, Riesgo. 2019 Sponges as natural environmental DNA samplers. *Current Biology* **29**: R395-R402.
- Maroneze, D. M., Tupinambas, T. H., Alves, C. B., & Vieira, F. (2011). Fish as ecological tools to complement biodiversity inventories of benthic macroinvertebrates. *Hydrobiologia*
- Maroneze, D. M., Tupinambas, T. H., Alves, C. B., & Vieira, F. (2011). Fish as ecological tools to complement biodiversity inventories of benthic macroinvertebrates. *Hydrobiologia* DOI 10.1007/s10750-011-0747-8.
- Martin, D. L., Ross, R. M., Quetin, L. B., & Murray, A. E. 2006. Molecular approach (PCR-DGGE) to diet analysis in young Antarctic krill *Euphausia superba*. *Marine Ecology Progress Series* **319**: 155–65.
- McMurdie, PJ, & Holmes, S. 2014. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol* **10**: e1003531.



- Meeter D. A., Livingston, R. J., & Woodsum, G. C. 1979. Ecological processes in coastal & marine systems. 315–38.
- Menhaden Advisory Commission. 2015. The Menhaden Fishery of the Gulf of Mexico: a regional management plan. Gulf States Marine Fisheries Commission.
- Meusnier, I., Singer, G. A. C., Landry, J.-F., Hickey, D. A., Hebert, P. D. N., & Hajibabaei, M. 2008. A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* **9**: 214.
- Minello, T.J., & Webb, J.W. 1997. Use of natural and created *Spartina alterniflora* salt marshes by fishery species and other aquatic fauna in Galveston Bay, Texas, USA. *Mar Ecol Progr Ser* **151**: 165–79.
- Moortazavi, B, Iverson, RL, Landing, WM, & Huang, W. 2000. Phosphorus budget of Apalachicola Bay: a river-dominated estuary in the northeastern Gulf of Mexico. *Mar Ecol Progr Ser* **198**: 33–42.
- Moortazavi, B, Iverson, RL, Landing, WM, Lewis, FG, & Huang, W. 2000. Control of phytoplankton production and biomass in a river-dominated estuary: Apalachicola Bay, Florida, USA. *Mar Ecol Progr Ser* **198**: 19-31.
- Moortazavi, B, Iverson, RL, & Huang, W. 2000. 2001. Dissolved organic nitrogen and nitrate in Apalachicola Bay, Florida: spatial distributions and monthly budgets. *Mar Ecol Progr Ser* **214**: 79-91.
- Moreby, S. J. 1988. An aid to the identification of arthropod fragments in the faeces of gamebird chicks (Galliformes). *Ibis* **130**: 519–26.
- Mühling, M., J. Woolven-Allen, J. C. Murrell, & I. Joint. 2008. Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities. *ISME J* **2**: 379-392.
- Murray, D. C., Bunce, M., Cannell, B. L., Oliver, R., Houston, J., White, N. E. *et al.* (2011). DNA-based faecal dietary analysis: a comparison of qPCR and high throughput sequencing approaches. *PLoS One*, *6*(10), e25776.
- Muyzer, G., De Waal, E. C., & Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.*, *59*(3), 695-700.
- Nelson, W. R., Ingham, M. C., AND Schaaf, W. E. 1977. LARVAL TRANSPORT AND YEAR-CLASS STRENGTH OF ATLANTIC MENHADEN, BREVOORTIA-TYRANNUS. *Fishery Bulletin* **75**: 23–41.
- Nicholson, WR. 1978. Gulf menhaden, *Brevoortia patronus*, purse seine fishery: Catch, fishing activity, and age and size composition, 1964-73, (ed.), US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service.
- Nicholson, N, 1978. Gulf menhaden, *Brevoortia patronus*, purse seine fishery: Catch, fishing activity, and age and size composition, 1964-73. US Department of

- Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Technical Report SSRF-722. 8 pp.
- NOAA-NMFS. 2015. Forecast for the 2015 Gulf and Atlantic Menhaden Purse-Seine Fisheries and Review of the 2014 Fishing Season. Gulf and Atlantic Forecast 2015.
- O'Farrell, H., Gruss, A., Sagarese, S. R., Babcock, E. A., & Rose, K. A. 2017. Ecosystem modeling in the Gulf of Mexico: current status and future needs to address ecosystem-based fisheries management and restoration activities. *Rev Fish Biol Fisheries* **27**: 587–614.
- Olsen, Z., Fulford, R., Dillon, K., & Graham, W. 2014. Trophic role of gulf menhaden *Brevoortia patronus* examined with carbon and nitrogen stable isotope analysis. *Marine Ecology Progress Series* **497**: 215–27.
- PACHECO, A. L., AND G. C. GRANT. 1965. Studies of the Early Life History of Atlantic Menhaden in Estuarine Nurseries: Part I, Seasonal Occurrence of Juvenile Menhaden and Other Small Fishes in a Tributary Creek of Indian River, Delaware, 1957-58. US Fish and Wildlife Service.
- Peck, J. I. 1893. On the food of menhaden. *Bull US Fish Comm.* p113.
- Pendleton, B. A. G. 1989. Proceedings of the Northeast Raptor Management Symposium and Workshop. Institute for Wildlife Research, National Wildlife Federation.
- Peters D. S., & Schaaf, W. E. 1981. Food requirements & sources for juvenile Atlantic menhaden. *Transactions of the American Fisheries Society* **110**: 317–24.
- Peterson, B. J., & Fry, B. (1987). Stable isotopes in ecosystem studies. *AnnRev Ecol Systemat.* **18**: 293-320.
- Pikitch, E., Boersma, P. D., Boyd, I. L., Conover, D. O., Cury, P., Essington, T., Heppell, S. S., Houde, E. D., Mangel, M., & Pauly, D. 2012. Little fish, big impact: managing a crucial link in ocean food webs. Lenfest Ocean Program, Washington, DC **108**: csiro:EP12435.
- Pikitch, E. K., Rountos, K. J., Essington, T. E., Santora, C., Pauly, D., Watson, R., Sumaila, U. R., Boersma, P. D., Boyd, I. L., AND Conover, D. O. 2014. The global contribution of forage fish to marine fisheries and ecosystems. *Fish and Fisheries* **15**: 43–64.
- Plagányi, É. E., AND Essington, T. E. 2014. When the SURFs up, forage fish are key. *Fisheries Research* **159**: 68–74.
- Pollock, J., Glendinning, L., Wisedchanit, T., & Watson, M. 2018. The Madness of Microbiome: Attempting to find consensus “Best Practice” for 16S microbiome studies. *Appl Environ Microbiol* **84**: e02627-17.
- Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., & Taberlet, P. 2012. Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology* **21**: 1931–50.

- Potter, IC, Tweedley, JR, Elliott, M, AND Whitfield, AK. 2015. The ways in which fish use estuaries: a refinement and expansion of the guild approach. *Fish and Fisheries* **16**: 230–39.
- Post, DM. 2002. Using stable isotopes to estimate trophic position: models, methods, and assumptions. *Ecology* **83**: 703–18.
- Powell (1993). M. (2002). Using stable isotopes to estimate trophic position: models, methods, and assumptions. *Ecology*, *83*, 703-718.
- Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, & F. O. Glöckner. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188-7196.
- Putland, J. N. (2005). *Ecology of phytoplankton, Acartia tonsa, and microzooplankton in Apalachicola Bay, Florida*. Florida State University.
- Putland, J. N., & Iverson, R. L. (2007). Ecology of *Acartia tonsa* in Apalachicola Bay, Florida, and implications of river water diversion. *Mar Ecol Progr Series* **340**, 173-187.
- Putland, J. N., & Iverson, R. L. (2007). Microzooplankton: major herbivores in an estuarine planktonic food web. *Mar Ecol Progr Series* **345**, 63-73.
- Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Sirdlow, H. P., & Gu, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**: 341.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schler, T., Yarza, P., Peplies, J., & Glöckner, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–6.
- Rajilić-Stojanović, M., Smidt, H., & de Vos, W. M. 2007. Diversity of the human gastrointestinal tract microbiota revisited. *Environ Microbiol* **9**: 2125–36.
- Reid, G. K. 1955. A summer study of the biology and ecology of East Bay, Texas. Part II. The fish fauna of East Bay, the Gulf beach, and summary. [tamug-irtdl.org](http://tamug-irtdl.org)
- Reintjes, JW. 1962 Development of eggs and yolk-sac larvae of yellowfin menhaden. U.S. Dep. Inter., Fish Wildl. Serv., Fish Bull. **62**: 93-102
- Reintjes, JW. 1970. The Gulf menhaden and our changing estuaries. *Proc Gulf Carrib Fish Inst* **13**: 87–90.
- Riemann, LH, Alfredsson, MM, Hansen, TD, Als, TG, Nielsen, P, Munk, K, Aarestrup, GE, Maes, H, Sparholt, MI, Petersen, M, Bachler & Castonguay, M. 2010. Qualitative assessment of the diet of European eel larvae in the Sargasso Sea resolved by DNA barcoding. *Biol Lett* **6**: 819–22.

- Riccioni, G., Stagioni, M., Piccinetti, C., AND Libralato, S. 2018. A metabarcoding approach for the feeding habits of European hake in the Adriatic Sea. *Ecol Evol* **8**: 10435–47.
- Robinson, K., Ruzicka, J., Hernandez, F., Graham, W., Decker, M., Brodeur, R., Sutor, M., 2015. Evaluating energy flows through jellyfish and gulf menhaden (*Brevoortia patronus*) and the effects of fishing on the northern Gulf of Mexico ecosystems. *ICES Journal of Marine Science*. 72 10.1093/icesjms/fsv088.
- Roeselers G., Mittge, E. K., Stephens, W. Z., Parichy, D. M., Cavanaugh, C. M., Guillemin, K., & Rawls, J. F. 2011. Evidence for a core gut microbiota in the zebrafish. *ISME J* **5**: 1595–608.
- Roithmayr, C. M., & Waller, R. A. 1963. Seasonal occurrence of *Brevoortia patronus* in the northern Gulf of Mexico. *Transactions of the American Fisheries Society* **92**: 301–2.
- Rollo, F., Ubaldi, M., Ermini, L., & Marota, I. (2002). Otzi's last meals: DNA analysis of the intestinal content of the Neolithic glacier mummy from the Alps. *Proc Natl Acad Sci U S A*, 99(20), 12594-12599.
- Sagarese, SR, Nuttall, MA, Geers, TM, Lauretta, MV, Walter, JF, Serafy, JE. 2016. Quantifying the trophic importance of Gulf menhaden within the Northern Gulf of Mexico ecosystem. *Mar & Coastal Fish*, 8:23-45.
- Schiebelhut, LM, Abboud, SS, Gómez Daglio, LE, Swift, HF, & Dawson, MN. 2017. A comparison of DNA extraction methods for high-throughput DNA analyses. *Mol Ecol Resour* **17**: 721–29.
- Schloss, PD, Istcote, SL, Ryabin, T, Hall, JR, Hartmann, M, Hollister, EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Iber, CF 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environ Microbiol*, 75, 7537–7541.
- Schrader, C, Schielke, A, Ellerbroek, L, & Johne, R. 2012. PCR inhibitors - occurrence, properties and removal. *J Appl Microbiol* **113**: 1014–26.
- Schnell, IB., Bohmann, K, & Gilbert, MTP. 2015. Tag jumps illuminated—reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular ecology resources* **15**: 1289–303.
- Schnell, IB, Thomsen, PF, & Gilbert, MTP. 2012. Screening mammal biodiversity using DNA from leeches. *Curr Biol*, 22: R262-R263.
- SEDAR, 2103. *SEDAR 34A – Gulf of Mexico Menhaden Stock Assessment Report*. SEDAR, North Charleston, SC 422 p.
- SEDAR63, 2018. *SEDAR 63 – Gulf of Mexico Menhaden Stock Assessment Report*. SEDAR, North Charleston, SC

- Shaw, RF, Cowan, JH, & Tillman, TL. 1985. Distribution and density of *Brevoortia patronus* (Gulf menhaden) eggs and larvae in the continental shelf waters of western Louisiana. *Bulletin of Mar Sci* **36**: 96–103.
- Siegenthaler, A, Wangensteen, OS, Soto, AZ, Benvenuto, C, Corrigan, L, & Mariani, S. (2018). Metabarcoding of shrimp stomach content: Harnessing a natural sampler for fish biodiversity monitoring. *Mol. Ecol. Resour.* 19, 206–220.
- Simon M, Scheuner, C, Meier-Kolthoff, JP, Brinkhoff, T, Wagner-Döbler, I, Ulbrich, M, Klenk, HP, Schomburg, D, Petersen, J, & Göker, M. 2017. Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine & non-marine habitats. *ISME J* **11**: 1483–99.
- Simmons, EG. 1967. The Texas menhaden fishery. Texas Parks and Wildlife Department.
- Simmons, EG, & Breuer, JP. 1964. edition. The Texas menhaden fishery. Texas Parks and Wildlife Department Bulletin.
- Simpson, C. A., Bi, H., Liang, D., Wilberg, M. J., Schueller, A. M., Nesslage, G. M., & Walsh, H. J. 2017. Spawning locations and larval dispersal of Atlantic Menhaden during 1977–2013. *ICES Journal of Marine Science* **74**: 1574–86.
- Simpson et al, 2016
- Soininen, EM, Valentini, A, Coissac, E, Miquel, C, Gielly, L, Brochmann, C. et al. (2009). Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Front Zool*, 6, 16.
- Song W, Li, L, Huang, H, Jiang, ., Zhang, F, Chen, X, Zhao, M, & Ma, L. 2016. The gut microbial community of Antarctic fish detected by 16S rRNA gene sequence analysis. *Biomed Res Int* **2016**: 3241529.
- Sullam KE, Essinger, SD, Lozupone, C., O'Connor, MP, Rosen, GL, Knight, R, Kilham, SS, & Russell, JA 2012. Environmental & ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Mol Ecol* **21**: 3363–78.
- Suttkus, RD 1956. Early life history of the Gulf menhaden, *Brevoortia patronus*. *Louisiana Transactions of the North American Wildlife Conference* **21**: 390–406.
- Symondson, W. O. C. 2002. Molecular identification of prey in predator diets. *Mol Ecol* **11**: 627–41
- Taberlet, P, Coissac, E, Pompanon, F, Brochmann, C, Willerslev, E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol*, 21: 2045-2050.
- Tarnecki A., Burgos, FA, Ray, CL, & Arias, CR. 2017. Fish intestinal microbiome: diversity and symbiosis unravelled by metagenomics. *J Appl Microbiol* **123**: 2–17.
- Telechea, 2009

- Thomsen, PF, Kielgast, J, Iversen, LL, Møller, PR, Rasmussen, M, & Willerslev, E. 2012. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE*, 7, e41732.
- Tsilimigras, MC and Fodor, AA. 2016. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Ann Epidemiol.*, 26(5): 330-5.
- Tuoinambas, T. H., Callisto, M., & Santos, G. B. (2007). Benthic macroinvertebrate assemblages structure in two headwater streams, south-eastern Brazil. *Rev Brasileira de Zool*, 24.
- Turner, WR. 1969. Life history of menhadens in the eastern Gulf of Mexico. *Transactions of the American Fisheries Society* **98**: 216–24.
- Ushio, M. 2019. Use of a filter cartridge combined with intra-cartridge bead-beating improves detection of microbial DNA from water samples. *Methods in Ecology & Evolution* **10**: 1142–56.
- Valentini, A, Miquel, C, Nawaz, MA, Bellemain, E, Coissac, E, Pompanon, F, Gielly, L, Cruaud, C, Nascetti, G, Wincker, P, Swenson, JE, Taberlet, P, 2009. New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Mol. Ecol. Resour.* 9: 51e60.
- Van Valkenburg, SD, Jones, JK, & Heinle, DR. 1978. A comparison by size class and volume of detritus versus phytoplankton in Chesapeake Bay. *Estuarine and Coastal Marine Science* **6**: 569–82.
- Vander Zanden, MJ, & Rasmussen, JB. 1999. Primary consumer  $\delta^{15}\text{N}$  and  $\delta^{13}\text{C}$  and the trophic position of aquatic consumers. *Ecology* 80: 1395–1404.
- Vanysacker, L, Declerck, SA, Hellemans, B, De Meester, L, Vankelecom, I, & Declerck, P. 2010. Bacterial community analysis of activated sludge: an evaluation of four commonly used DNA extraction methods. *Applied Microbiology and Biotechnology* **88**: 299–307.
- Vaughan, D. S., Shertzer, K. W., AND Smith, J. W. 2007. Gulf menhaden (*Brevoortia patronus*) in the US Gulf of Mexico: fishery characteristics and biological reference points for management. *Fisheries Research* **83**: 263–75.
- Walter J. M., Bagi, A., & Pampanin, D. M. 2019. Insights into the potential of the Atlantic cod gut microbiome as biomarker of oil contamination in the marine environment. *Microorganisms* **7**: 209.
- Wang, Y, Tian, RM, Gao, ZM, Bougouffa, S, & Qian, P-Y. 2014. Optimal eukaryotic 18S and universal 16S/18S ribosomal RNA primers and their application in a study of symbiosis. *PLoS One* **9**: e90053.
- Waraniak, JM, Marsh, TL, & Scribner, KT. 2019. 18S rRNA metabarcoding diet analysis of a predatory fish community across seasonal changes in prey availability. *Ecol Evol* **9**: 1410–30.
- Ward, DM., Weller, R, & Bateson, MM. 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **345**: 63–65.

- Weigand, H, A.J. Beermann, F. Čiampor, F. O. Costa, Z. Csabai, S. Duarte, M. F. Geiger, M. Grabowski, F. Rimet, B. Rulik, M. Strand, N. Szucsich, A. M. ligand, E. Willassen, S. A. Wyler, A. Bouchez, A. Borja, Z. Čiamporová-Zaťovičová, S. Ferreira, K. B. Dijkstra, U. Eisendle, J. Freyhof, P. Gadawski, W. Graf, A. Haegerbaeumer, B. B. van der Hoorn, B. Japoshvili, L. Keresztes, E. Keskin, F. Leese, J. N. Macher, T. Mamos, G. Paz, V. Pešić, D. M. Pfannkuchen, M. A. Pfannkuchen, B. W. Price, B. Rinkevich, M. A. L. Teixeira, G. Várбірó, & T. Ekrem. 2019. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Sci Total Environ* 678: 499-524.
- Weisburg, WG, Barns, SM, Pelletier, DA, & Lane, DJ. 1991. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173**: 697–703.
- Weis, S., Xu, ZZ, Peddada, S., et al. 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5, 27.
- White, JR, Nagarajan, N, & Pop, M. 2009. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* **5**: e1000352.
- Wustman, BA., Lind, J., Wetherbee, R., Gretz, MR., 1998. Extracellular Matrix Assembly in Diatoms (Bacillariophyceae). *Plant Physiology* Apr. 116(4) 1431-1441.
- Yarza, P, Yilmaz, P, Pruesse, E, Glöckner, FO, Ludwig, W, Schleifer, KH. et al. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol*, 12(9), 635-645.
- Zarrinpar, A, Chaix, A, Yooseph, S, & Panda, S. (2014). Diet and feeding pattern affect the diurnal dynamics of the gut microbiome. *Cell Metab*, 20(6), 1006-1017.
- Zeale, MR, Butlin, RK, Barker, GL, Lees, DC, & Jones, G. (2011). Taxon-specific PCR for DNA barcoding arthropod prey in bat feces. *Mol Ecol Resour*, 11(2), 236-244.