

University of New Hampshire
University of New Hampshire Scholars' Repository

Master's Theses and Capstones

Student Scholarship

Spring 2016

BACTERIAL AND PHAGE INTERACTIONS INFLUENCING *Vibrio parahaemolyticus* ECOLOGY

Ashley L. Marcinkiewicz
University of New Hampshire, Durham

Follow this and additional works at: <https://scholars.unh.edu/thesis>

Recommended Citation

Marcinkiewicz, Ashley L., "BACTERIAL AND PHAGE INTERACTIONS INFLUENCING *Vibrio parahaemolyticus* ECOLOGY" (2016). *Master's Theses and Capstones*. 852.
<https://scholars.unh.edu/thesis/852>

This Thesis is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Master's Theses and Capstones by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.

BACTERIAL AND PHAGE INTERACTIONS INFLUENCING

Vibrio parahaemolyticus ECOLOGY

BY

ASHLEY MARCINKIEWICZ

Bachelor of Arts, Wells College, 2011

THESIS

Submitted to the University of New Hampshire

In Partial Fulfillment of

The Requirements for the Degree of

Master of Science

in

Microbiology

May, 2016

This thesis has been examined and approved in partial fulfillment of the requirements for the degree of Masters of Science in Microbiology by:

Thesis Director, Cheryl A. Whistler
Associate Professor of Molecular, Cellular, and Biomedical Sciences

Stephen H. Jones
Research Associate Professor of Natural Resources and the Environment

Jeffrey T. Foster
Assistant Professor of Molecular, Cellular, and Biomedical Sciences

On April 15th, 2016

Original approved signatures are on file with the University of New Hampshire Graduate School.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	vi
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
ABSTRACT.....	x

CHAPTER	PAGE
I. INTRODUCTION.....	1
II. CHAPTER 2: Bacterial Community Profiling of Individual Oysters: Taxonomic and Functional Associations with <i>Vibrio parahaemolyticus</i> Abundance.....	19
Abstract.....	20
Introduction.....	21
Results and Discussion.....	23
Conclusions.....	43
Methods.....	44
Acknowledgements.....	49
Supplementary Information.....	50
III. CHAPTER 3: Correlations between Abundance of <i>Vibrio parahaemolyticus</i> and the Microbiome of the Eastern oyster (<i>Crassostrea virginica</i>).....	51
Abstract.....	52
Introduction.....	53
Results and Discussion.....	54
Conclusions.....	70
Methods.....	71

Acknowledgements.....	75
Supplementary Information.....	76
IV. CHAPTER 4: Using Phylogenetic Relationships and Phage Content to Elucidate Invasion of <i>Vibrio parahaemolyticus</i> Sequence Type 36 into the United States North Atlantic Coast.....	78
Abstract.....	79
Introduction.....	80
Results and Discussion.....	81
Conclusions.....	106
Methods.....	107
Acknowledgements.....	110
Supplementary Information.....	111
V. REFERENCES.....	121

ACKNOWLEDGEMENTS

I give sincere thanks and gratitude to Dr. Cheryl Whistler, who has guided me through the ups and downs during my time at UNH. I also thank Drs. Steve Jones, Jeff Foster, and Vaughn Cooper, for helpful suggestions, thoughts, and discourse. I am indebted to my peers, especially Feng Xu and Jong Yu, for invaluable help, ideas, and support along the way, but most particularly for your willingness to teach. I couldn't have done this without my parents' support in getting me to where I am to get this degree, and Lindsey Marcinkiewicz, Tim Abrams, Sarah Bonawitz, and Nicole Peters for encouragement, support, and the occasional kick in the butt.

LIST OF TABLES

Table 1.1: <i>Vibrio parahaemolyticus</i> abundance in shellfish.....	10
Table 2.1. Results of select steps in the analysis pipeline with two different analysis programs.....	25
Table 2.2. OTUs at each level of taxonomic classification with two different analysis programs.....	25
Table 2.3. Nutrient levels at each collection site.....	37
Table 2.4. Distribution of <i>Vibrio parahaemolyticus</i> in oyster samples as determined by MPN and 16s sequencing.....	39
Table 2.5. Environmental conditions at each sampling site.....	40
Table 2S.1. Number of reads per sample.....	50
Table 2S.2. Functions positively correlated with <i>Vibrio parahaemolyticus</i> class....	50
Table 3.1. Distribution of <i>Vibrio parahaemolyticus</i> in oyster samples.....	55
Table 3S.1. Number of reads per sample.....	76
Table 4.1. Summary of screening environmental strains for phage.....	85
Table 4.2. All strains used in this study.....	87
Table 4.3. PCR primers used in this study.....	109
Table 4S.1. All results of screening environmental strains for phage.....	111

LIST OF FIGURES

Figure 1.1. Emergence of pathogenic <i>V. parahaemolyticus</i> in Pacific Ocean correlating with rises in ocean temperatures.....	7
Figure 2.1. Alpha diversity curves for all samples.....	27
Figure 2.2. Distribution of OTUs in oysters and overlaying water grouped by substrate and site.....	29
Figure 2.3. Dual hierarchical analysis of phyla-level classification for all samples.....	30
Figure 2.4. Unifrac phylogenetic distance analysis of all taxonomic levels for all samples.....	34
Figure 2.5. LEfSe analysis for oyster samples grouped by site.....	35
Figure 2.6. Annotated functions in significantly different numbers at each site....	37
Figure 2.7. Phylotypes significantly different based on <i>Vibrio parahaemolyticus</i> abundance class.....	42
Figure 3.1. Alpha diversity analysis.....	57
Figure 3.2. Distribution of phylotypes across sampling dates.....	60
Figure 3.3 Dual hierarchical analysis of phyla-level classification for all samples.....	62
Figure 3.4. Unifrac clustering of all taxonomic levels for all samples.....	65
Figure 3.5. LEfSe analysis for oyster samples grouped by date.....	67
Figure 3.6. LEfSe analysis for oyster samples grouped by <i>Vibrio parahaemolyticus</i> abundance.....	69
Figure 3S.1. LEfSe analysis for oyster samples grouped by date and <i>V. parahaemolyticus</i> abundance.....	77
Figure 4.1. Gene map of f237 and ST36-associated phage.....	83
Figure 4.2. Visualization of amplification by phage-detecting PCR primers.....	84
Figure 4.3. Maximum-likelihood tree of <i>V. parahaemolyticus</i> strains.....	99

Figure 4.4. Maximum-likelihood tree of <i>V. parahaemolyticus</i> ST36 strains.....	102
Figure 4.5. Phylogenetic tree of Vipa26.....	105
Figure 4S.1. Maximum-likelihood tree of phage-harboring <i>V. parahaemolyticus</i> ST36 strains.....	120

ABSTRACT

BACTERIAL AND PHAGE INTERACTIONS INFLUENCING

Vibrio parahaemolyticus ECOLOGY

By

Ashley Lynne Marcinkiewicz

University of New Hampshire, May, 2016

Vibrio parahaemolyticus, a human pathogenic bacterium, is a naturally occurring member of the microbiome of the Eastern oyster. As the nature of this symbiosis is unknown, the oyster presents the opportunity to investigate how microbial communities interact with a host as part of the ecology of an emergent pathogen of importance. To define how members of the oyster bacterial microbiome correlate with *V. parahaemolyticus*, I performed marker-based metagenetic sequencing analyses to identify and quantify the bacterial community in individual oysters after culturally-quantifying *V. parahaemolyticus* abundance. I concluded that despite shared environmental exposures, individual oysters from the same collection site varied both in microbiome community and *V. parahaemolyticus* abundance, and there may be an interaction with *V. parahaemolyticus* and *Bacillus* species. In addition, to elucidate the ecological origins of pathogenic New England ST36 populations, I performed whole genome sequencing and phylogenetic analyses. I concluded ST36 strains formed distinct subpopulations that correlated both with geographic region and unique phage content that can be used as a biomarker for more refined strain traceback. Furthermore, these subpopulations indicated there may have been multiple invasions of this non-native pathogen into the Atlantic coast.

CHAPTER 1

INTRODUCTION

1. Justification

Geologists categorize this age in time as the Cenozoic Era, or the Age of Mammals.

Fair enough, if we wish to honor multicellular creatures, but we are still not free of the parochialism of our scale. If we must characterize a whole by a representative part, we certainly should honor life's constant mode. We live now in the "Age of Bacteria." Our planet has always been in the "Age of Bacteria," ever since the first fossils - bacteria, of course - were entombed in rocks more than 3 billion years ago. On any possible, reasonable or fair criterion, bacteria are - and always have been - the dominant forms of life on Earth. Our failure to grasp this most evident of biological facts arises in part from the blindness of our arrogance but also, in large measure, as an effect of scale. We are so accustomed to viewing phenomena of our scale - sizes measured in feet and ages in decades - as typical of nature. [86]

The idea of the "Age of Bacteria" is especially relevant considering the co-evolution of a wide range of these higher-order organisms with their bacterial symbionts, which is frequently reflected in mirrored phylogenetic trees of host and symbionts [e.g., 42, 97, 151]. The interactions between host and microbial symbionts are quite intricate and susceptible to disruption in model and naturally-occurring systems. Both germ-free mice and mice treated with antibiotics are more vulnerable to infection from pathogens than conventionally-colonized mice, likely from both direct competition and up-regulation of host defenses by the gut microbiome

[20, 80, 92, 159]. Healthy and diseased states in humans are frequently correlated with changes in symbionts [46, 120, 134]. The function of such symbioses is not always known, particularly in lesser-studied environmental models. For example, the Eastern oyster (*Crassostrea virginica*) is host to *Vibrio*, a genus containing several human pathogenic lineages, and the nature of this symbiosis is unknown. The oyster presents the opportunity to investigate how microbial communities interact with a host as part of the ecology of an emergent pathogen of importance.

2. *Vibrio* spp.

An early description of the *Vibrio* genus as Gram-negative, rod-shaped bacteria with a single flagellum, preferential salinity (0.5-5.0%) and pH ranges (5.5-10), and which can tolerate temperatures between 15-41°C [45], although basically informative, hardly captures the essence of why these organisms are worthy subjects of study. These microorganisms are ubiquitous residents of estuarine and coastal marine habitats, where they live in water and animals inhabiting those waters, including corals [43], fish [28], shellfish [154], sponges [95], and shrimp [9]. There are well over 100 named species in this genus [http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=662], twelve of which cause illness in humans [38]. But as *Vibrio* spp. have been identified both as potential probiotics and antagonistic toward hosts [113], neither their successes as colonists nor their ecological function in these habitats are fully understood.

3. Epidemiology of *Vibrio* spp.

Illness from *Vibrio* spp. is collectively called vibriosis. Vibriosis reporting began in 1988 through the United States Center for Disease Control (CDC) with “Cholera and other *Vibrio*

illness surveillance system” (COVIS) including 4 states; now it is nationally reportable [36]. This ongoing epidemiological data collection allows for analyzing trends in illness, and vibriosis is increasing across the nation. The number of reported cases rose by 52% in 2014 compared to 2006-2008, an increase of ~120% from when analyses started in 1996 [35].

The regions where illnesses are prevalent are shifting as well. Historically, infections occurred in warmer regions. Illnesses traced to the Gulf coast states (Texas, Louisiana, Mississippi, Alabama, and Florida) made up 47.9% (449/937) of all reported infections in 1997-1998, while the Northeast Atlantic (NEA) states (New England and New York) were only 3% of total infections (28/937). But in 2013, the NEA states made up 23.0% of reported vibriosis infections (217/944) and the Gulf states were only 36.0% (340/944) [36]. Most cases are self-limiting so many patients do not seek medical treatment, and those do are not always tested for vibriosis. The estimations of rates of underreporting vary from 1:20 [225] to as high as 1:145 [178; Andy DePaola, personal communication]. *Vibrio cholerae*, *Vibrio vulnificus*, and *Vibrio parahaemolyticus* cause the most cases of vibriosis. However, most strains of these species are harmless and only a handful cause illness.

3.1. *Vibrio cholerae*

Perhaps the most well-known *Vibrio* spp., *Vibrio cholerae* is the causative agent of cholera, a severe gastroenteritis that if left untreated can lead to extreme dehydration and death [34]. *V. cholerae* was first isolated in 1854 by Filippo Pancini [160] but is historically credited to Robert Koch in 1884 after investigating outbreaks in Alexandria [98, 121]. While an estimated 3-5 million people are infected with *V. cholerae* worldwide with about 10% being severe enough to cause death if untreated [34], infections are relatively rare in the US. There

were only 73 reported cases in 2013 (6% of total reported vibriosis cases), killing six [36]. When considering underreporting, it is estimated 277 domestically-acquired cases occur per year [178].

Pathogenic *V. cholerae* strains contain two primary virulence factors. Cholera toxin is an enterotoxin encoded by a filamentous bacteriophage, CTXΦ, which integrates into the *ctx* site in the chromosome [218]. Interactions with KSF-1Φ, another filamentous bacteriophage, indirectly enhance the spread of cholera toxin genes into other strains of *V. cholerae* [66, 67]. The second main virulence factor is toxin-coregulated pili (TCP). These Type IV pili aggregate the bacterial cells to protect cells from host defenses and concentrate released cholera toxin [200], and are also the receptors for CTXΦ [218]. These phage (and other genetic elements with virulence-associated functions) were acquired through horizontal gene transfer [67] and are integral in the evolution and emergence of *V. cholerae* pathogenic lineages.

3.2. *Vibrio vulnificus*

Vibrio vulnificus was first isolated and characterized by the CDC in 1976 from blood cultures [96] but was not formally named until 1979 [65]. Pathogenic *V. vulnificus* strains cause gastroenteritis, but can also infect open wounds on the skin [37]. There were 137 reported cases in 2013, 12% of vibriosis cases [36], but it is estimated that 203 domestically-acquired cases occur yearly [178]. Infection by *V. vulnificus* is less frequent than *V. cholerae*, but this microbe is more deadly – sepsis from *V. vulnificus* infections has a 50% mortality rate [37].

V. vulnificus biotype I is particularly deadly, causing 95% of the shellfish-related deaths in the US [157]. There are two biotype I genotypes: C-type and E-type. Ninety percent of C-type strains are of clinical lineages, and 93% of E-types are environmental strains [174]. These

genotypes show about equal proportions in the water column, but 87% of strains isolated from oysters are E-type [221]. Oysters take up static C-type strains at greater rates than E-type [77] but E-type strains aggregated into marine snow are taken up at higher rates than C-type aggregates [75]. Hence, there is differential accumulation between clinical and environmental genotypes of *V. vulnificus* when interacting with organic and inorganic matter in the surrounding environment.

3.3. *Vibrio parahaemolyticus*

First identified from a shirasu (fried baby sardines) outbreak in Japan in 1950 that caused 272 illnesses and 20 deaths, infection from pathogenic *Vibrio parahaemolyticus* causes self-limiting gastroenteritis [79]. Albeit limited feeding studies, the infectious dose of *V. parahaemolyticus* ranges from 10^5 to 10^8 [94, 177] although pen shells containing 2.4×10^2 organisms per gram of oyster tissue (with an average of 50g oyster meat ingested) have been indicted in causing illness [89]. Illnesses from *V. parahaemolyticus* were relatively rare and sporadic in the US [15, 48] until a multi-state outbreak occurred in the summer of 1998, causing 23 infections [33]. The number of cases has only increased since [150], and it is estimated that with underreporting, 35,000 domestically-acquired cases occur each year in the US [178]. This organism is the leading cause of seafood borne illness [178], and in 2013 there were 594 reported cases in the US, comprising 51% of total vibriosis cases [36].

3.3.1. Pandemic *V. parahaemolyticus*

A few pathogenic *V. parahaemolyticus* strains have disseminated and one particularly virulent lineage, sequence type (ST) 3 (serotype O3:K6, among others [146]), has spread

globally to pandemic status. This began as an outbreak in February 1996 in Calcutta, India [155] that quickly spread around the world, with outbreaks being reported in every inhabited continent [146, 213]. The first outbreak outside of Asia was in Chile in 1997 [85], correlating with changing ocean temperatures of an El Niño event [139]. Warming ocean temperatures have influenced the spread of pathogenic strains, pandemic lineage and otherwise [10, 136, 138, 215; Fig. 1.1].

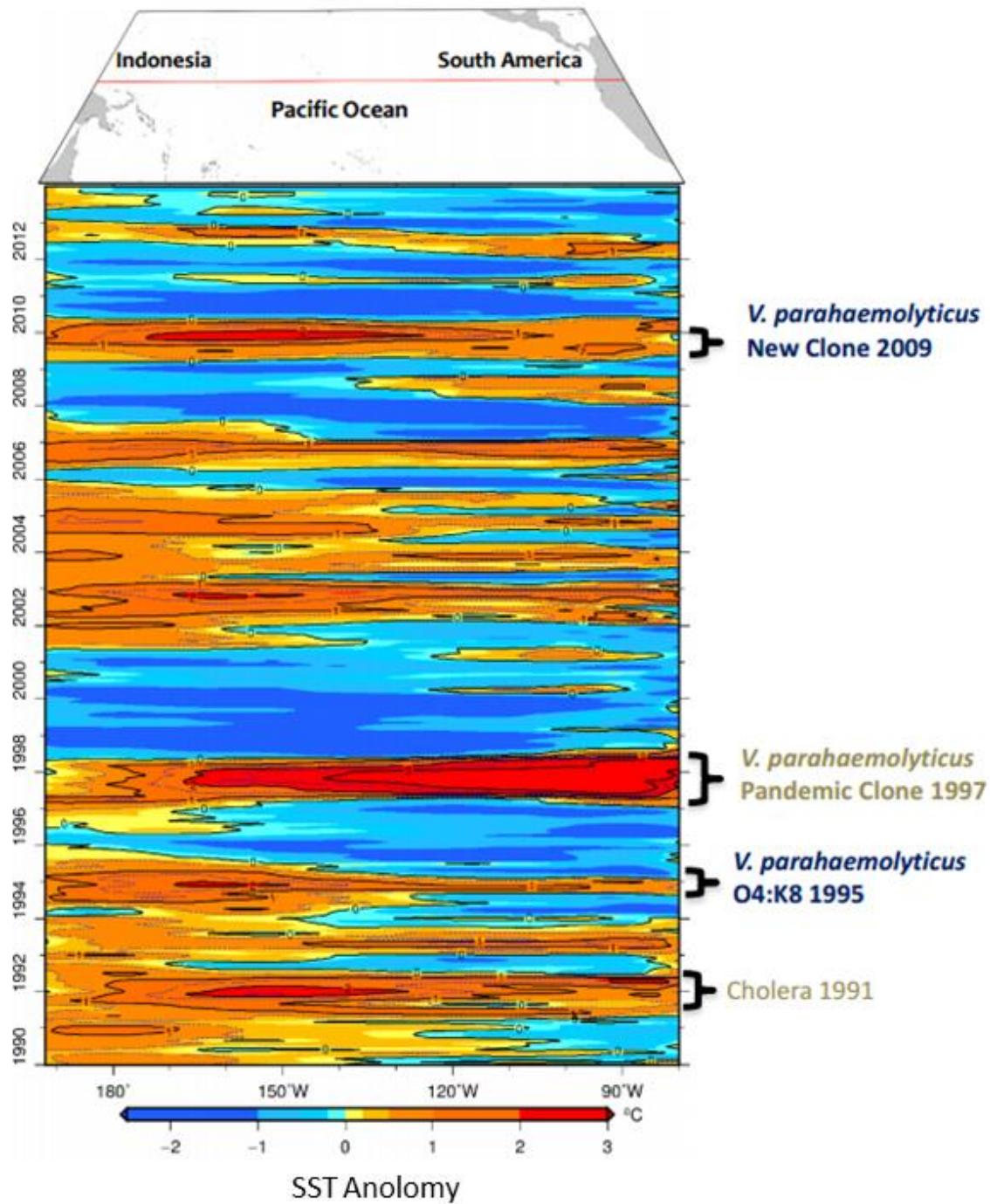


Figure 1.1. Emergence of pathogenic *V. parahaemolyticus* in Pacific Ocean correlating with rises in ocean temperatures.
From Martinez-Urtaza [136].

Researchers investigated the emergence of this pathogenic lineage with whole genome comparisons. Several unique pathogenicity islands exist that likely influence the highly virulent nature of this lineage [22]. f237, an Inovirus filamentous bacteriophage, is integrated into the chromosome at the *dif* site [100] of the pandemic strains and is unique to this lineage so is used for diagnostic identification [145, 147]. The specific diagnostic marker gene, ORF8, shows similar structure to the *plx* gene in *Drosophila*, encoding an adhesion molecule. It is proposed ORF8 may increase attachment to either the intestines during infection or to plankton which would increase dissemination [147]. ST3 does show increased cytotoxicity and adherence to HeLa cells compared to non-ST3 strains, and it is suggested ORF8 could be influencing these traits [232]. There are other phage integrated into the genome of pandemic *V. parahaemolyticus*. In particular, VP58.5, a Myovirus with high similarity to VHML in *Vibrio harveyi*, has been isolated in pandemic strains off the Chilean coast. Pandemic *V. parahaemolyticus* strains harboring this phage are up to 15 times more sensitive to UV radiation, presumably due to phage induction [234]. Phage act as biological mechanisms for controlling both the virulence [232] and the population structure of this lineage [234].

3.3.2. *V. parahaemolyticus* Sequence Type 36

ST36 strains (serotype O4:K12), which are native to the Pacific Northwest (PNW) and cause reoccurring illness in Canada [13], Washington [209], and Oregon [32], have recently been traced to outbreaks in other regions. Particularly, ST36 outbreaks have occurred in California, Spain [136, 141], and the North- to mid-Atlantic US coast [137], leading to illness of 104 people in the US in 2013 [149]. ST36 strains in the Northeast [224, 231], Maryland [88], and Spain [136, 141] likely derived from the PNW populations. While the pandemic lineage is highly

studied and well characterized, whole genome comparisons of ST36 strains to uncover specific population evolution and emergence are only just beginning.

4. Abundance of *V. parahaemolyticus* in shellfish

V. parahaemolyticus, the top threat for seafood-borne illnesses [178], naturally occurs in warm saline environments, and accumulates in animals living in these environments [35, 112]. As transmission into humans is typically through raw, undercooked, or mishandled shellfish, especially oysters, many studies all over the world evaluate the levels of *V. parahaemolyticus* in such animals (e.g., Table 1.1). Current methodology pools animals, usually in batches of 12, to determine average *V. parahaemolyticus* levels [71]. Very few experiments have considered levels in individual shellfish [e.g., 114] even though presumably, consuming only one contaminated animal could cause illness if it contained the infectious dose. The Food and Drug Administration (FDA) recommends shellfish safe for consumption contain < 10,000 MPN/g (4 log₁₀ MPN/g) total *V. parahaemolyticus* [70], but naturally occurring levels are sometimes higher than this (Table 1.1), and vary with abiotic and abiotic influences.

Table 1.1. *Vibrio parahaemolyticus* abundance in shellfish.

Quantification of *V. parahaemolyticus* in shellfish from around the world, across seasons. Studies are ordered by date. The means and ranges of *V. parahaemolyticus* were log-transformed.

^aBD = below detection

^bDoes not include BD

Site	Date	Animal	Mean	Range	Reference
Alabama, USA	March 1999 - Sept. 2000	Oysters	2.24 MPN/g	1.48-3.47 MPN/g	52
Mulki, India	Jan. 2002 - Dec. 2002	Oysters	3.73 CFU/mL	2-4.82 CFU/mL	50
Galicia, Spain	Jan. 2002 - Dec. 2004	Mussels	1.09 MPN/g	-0.92-1.51 MPN/g	140
Oregon, USA	Nov. 2002 - Oct. 2003	Oysters	1.34 MPN/g ^b	BD ^a -1.63 MPN/g	55
Sao Paulo State, Brazil	May 2004 - June 2005	Oysters	3.16 MPN/g	0.78-5.04 MPN/g	189

4.1. Environmental factors that correlate with *V. parahaemolyticus* abundance

Several environmental factors directly influence the abundance of *V. parahaemolyticus*.

Temperature is, to present knowledge, always positively correlated with naturally occurring *V. parahaemolyticus* when temperature change is captured in observations [50, 52, 55, 140, 189]. This species will grow in liquid media with a pH 4.8 and temperatures as low as 5°C [18], but 10°C is lowest for growth in natural environments [112]. During extended periods of temperatures below 10°C (e.g., winter), cells enter a viable but not culturable (VBNC) state. *V. parahaemolyticus* VBNC cells change morphology and no longer express epitopes [64], and show resistance to temperature extremes, low salinity, and high acid [227]. Correlating with temperature, *V. parahaemolyticus* levels have a seasonal growth pattern, increasing in warmer months and decreasing in cooler [52, 55, 140, 189]; tropical climates see the increase in the dry season and decrease in the wet season [50]. Correlations between the halophilic *V. parahaemolyticus* and salinity levels are mixed. Whereas both significant positive and negative

correlations have been reported [52, 140], others are not significant [55, 189]. This may be due to the extensive salinity range (3-35ppt) tolerated by *V. parahaemolyticus* [198].

4.2. Biotic factors that influence abundance of *V. parahaemolyticus*

In addition to environmental influences, biotic factors have also been reported to correlate with *Vibrio* spp. abundance. Dinoflagellates [60], diatoms [6, 196], copepods [59], and plankton [133, 208, 228] and more generally chlorophyll *a* [198, 210], are positively correlated with *Vibrio* levels, as these serve either as a food resource or a mechanism of dispersal. Protozoa [228] and nematodes (*Terschellinga*, *Molgolaimus*, and *Halolaimus* spp.) negatively correlate with *V. parahaemolyticus* in benthic sediment [216]; likely the nematodes graze on the bacterial species. Oyster hemocytes may control the abundance of certain *V. parahaemolyticus* strains, and therefore overall species abundance, as hemocytes differentially kill opaque and translucent strains [82] as well as environmental and clinical strains [217]. *In-vitro* bacterial-vibrio competitions illustrate the ability of several types of marine bacteria to influence *Vibrio* abundance [94, 177] but to present knowledge, *in-vivo* (e.g., in the oyster or water) experiments correlating *V. parahaemolyticus* abundance and other bacteria are lacking.

5. Post-harvest processing techniques reduce *V. parahaemolyticus* abundance in shellfish

In the US, the National Shellfish Sanitation Program “promote(s) and improve(s) the sanitation of shellfish (oysters, clams, mussels and scallops) moving in interstate commerce through federal/state cooperation and uniformity of State shellfish programs” [http://www.fda.gov/Food/GuidanceRegulation/FederalStateFoodPrograms/ucm2006754.htm]. Their model ordinance outlines requirements for states harvesting shellfish to promote safe

shellfish. It also includes additional strategies for reducing *Vibrio* abundance for states with a *Vibrio* control plan, like rapid cooling of shellfish in ice, slurries, or mechanical refrigeration down to 50°F(10°C) within a few hours after harvesting, all of which have been shown to significantly lower microbial load [131, 173].

Other strategies for reducing *Vibrio* abundance include depuration and relay. Depuration involves bathing oysters in sterilized (typically with UV light) water that is either recirculated or flow-through in design. Relaying oysters involves physically moving oysters to an environmental region or tanks other than the harvest site where *V. parahaemolyticus* is found in very low abundance, if at all, often due to high salinity levels. The effectiveness of depuration and relaying oysters to reduce *Vibrio* spp. abundance was tested by collecting oysters from the Great Bay Estuary (GBE), New Hampshire, and relaying in Spinney Creek (SC), Maine, a creek off the GBE with similar temperatures but increased salinity. Relaying oysters to SC was more effective than depuration in reducing abundance of *V. vulnificus* in oysters [106]. Later experiments illustrated this same result with levels of *V. parahaemolyticus*: relaying oysters to SC caused a more frequent decrease in *V. parahaemolyticus* abundance than depuration and tank-relay experiments, which used UV-sterilized SC water at temperatures set equal to current conditions in the creek [233]. This implies non-sterilized creek water is more effective at reducing *V. parahaemolyticus* abundance in oysters than sterilized water of the same temperature and salinity. The microbial communities in seawater and oysters may influence the abundance of *V. parahaemolyticus* in oysters.

6. Next-generation sequencing technology

Sequencing technologies have rapidly changed since Ray Wu and colleagues first published methods on elongating nucleotides [105], which Fredrick Sanger adopted and expanded to develop Sanger sequencing [176]. Currently, next-generation sequencing has expanded from allowing hundreds of thousands of short reads in a single run with 454 sequencing technologies to billions of reads with Illumina technologies. There are many applications of next-generation sequencing in the field of microbiology, including host-microbial interactions and bacterial population studies.

6.1. Bacterial community profiling

Next-generation sequencing technologies allow researchers to profile bacterial populations independent of culturing by amplifying and sequencing the hypervariable regions of the 16s rRNA gene. Historically, sequencing the entire 16s rRNA gene has been performed to determine the relationship between bacterial isolates/strains, distinguishing between species [54, 74, 142], and even uncovering previously unidentified or unculturable microbes [219]. The 16s rRNA gene contains alternating regions of highly conserved and hypervariable regions [226]. These hypervariable regions alone can be used for species identification [211], and even short regions (i.e. – 100bp) of certain hypervariable regions are capable of capturing species composition patterns seen from using the full 16s rRNA gene [129]. Next-generation sequencing technologies permit large-scale sequencing of short stretches of the 16s hypervariable regions to determine both relative composition and abundance of bacteria in a mixed community sample, and investigate differences in the microbiome between samples of different states or

environmental conditions. Analyzing the billions of reads produced to draw such comparisons requires the use of specific software programs designed for these purposes.

6.1.1. Generation and Analysis of 16s rRNA sequencing data

Generating 16s data requires extracting the bacterial DNA from samples and amplifying the variable region of choice. Primers have unique barcodes to allow sample multiplexing, as well as adapters necessary for the sequencing process [e.g., 31]. Data is returned as raw sequencing reads with per-base quality information. The two most commonly used software programs to analyze these short-read 16s rRNA data are QIIME [30] and mothur [181]. Both programs, albeit with core differences in algorithms, function in essentially the same way. Reads are quality-filtered, and overlapping paired-end reads can be used to extend the total read length and/or error correct the corresponding read. Quality-filtered reads are then clustered into groups based on similarity of the sequence, with 97% the most frequently used threshold as it is thought to best represent species-level classifications [192]. These clusters, which are known as operational taxonomic units (OTUs), can then be used to assess between and within-sample diversity, and outside software such as LEfSe [185] and PICRUSt [123] identify specific taxonomic or functional differences, respectively, between biological groupings. As useful as these techniques are in generating and analyzing sequencing data, there are several caveats that need to be considered when drawing conclusions from these data.

6.1.2. Caveats of 16s rRNA sequencing data

Generating 16s rRNA sequencing data involves more than simply loading raw sample into the sequencer, so each step in the preparation process introduce biases that may affect the

resulting OTUs. The first step is extracting and purifying bacterial DNA from the sample. Using different extraction methods on the same sample [73, 118], even laboratory-generated control samples [25, 195], and performing other steps in the same way alters OTU composition.

The purified bacterial DNA is then amplified using PCR to attach the sequencing adapters and sample identification barcodes. During the amplification process, errors could arise from the formation of chimerical and heteroduplex molecules, which if not removed in downstream processing, falsely inflate the total number of OTUs [2]. Differential amplification efficiency of DNA strands [2], starting concentration of the template [117, 168], and primer choice [61, 168, 195] all influence the relative abundance of OTUs.

Once the adapters and barcodes are attached to the DNA strand, the hypervariable region is sequenced. The two main sequencing platforms both have known error biases. 454 sequencing generates errors mostly associated with homopolymers, including insertions, deletions, and mismatches [99]. Illumina sequencing has been shown to cause insertions, deletions, and mismatches in certain motifs, but these do not occur randomly and are associated with library preparation and primer choice. In addition, Illumina characteristically drops in quality at the end of the read [122]. Failure to perform quality checking to remove such erroneous reads will falsely increase the number of OTUs.

Quality filtered reads are clustered into OTUs. QIIME currently offers the choice of twelve main programs for OTU clustering, each of which can be customized [30]. The default is uclust, which used a centroid-based algorithm. Each sequence in an OTU falls within the defined similarity threshold of the centroid sequence [56]. mothur implements only three clustering algorithms all based on distance matrices, the default being average linkage [181], where the average distance from all sequences within each OTU to all sequences within every

other OTU is greater than the defined similarity threshold [180]. Different programs, combined with the various options (e.g., referenced-based or *de novo*, pre-filtering/clustering steps) will influence the numbers of OTUs generated [14, 41, 180]. In addition to the program used to cluster, the standard 0.97 threshold is not consistent between lineages for distinguishing species [91, 180], and most organism have multiple copies of the 16s rRNA gene, which are not always within the 0.97 similarity threshold [104]. Collapsing OTUs into phylotypes, or OTUs with the same taxonomic classification, helps to alleviate these inconsistencies.

Once the OTUs are generated, taxonomy is assigned to each OTU by comparing a sequence in the OTU to a 16s rRNA taxonomy database. When using QIIME, the centroid sequence is used as the representative by default, but users may also choose the longest, most abundant, or a random sequence to assign taxonomy to the entire OTU [30]. In mothur, every sequence in an OTU is assigned taxonomy and the majority is used to assign the entire OTU [180]. SILVA [170], greengenes [53], and RDP [44] are the most frequently used taxonomy databases. As of January 2016, the SILVA database contains 4.9 million sequences (<http://www.arb-silva.de>), greengenes just over 1.2 million (<http://greengenes.lbl.gov>), and RDP, 3.2 million (<http://rdp.cme.msu.edu>). Older versions of QIIME used RDP by default; newer versions use greengenes [30]. mothur recommends SILVA [180]. The differences in these methodologies and databases will influence the generated taxonomic profile [223] and subsequent diversity or abundance pattern measurements. Comparisons between different methods can maximize the usefulness of the generated taxonomies.

Besides evaluating within-sample diversity, rarefied alpha diversity plots can be utilized to judge if samples were sequenced deeply enough to capture the full diversity. The samples are rarified stepwise and the chosen diversity measure is calculated at each step. This

data is plotted to visualize how diversity changes as the number of sequences considered increases. Rarefied alpha diversity plots that plateau demonstrate as more sequences are considered, the diversity stops increasing. In other words, deeper sequencing would not increase the diversity, so the plateau indicates that the true diversity of samples has been captured. Any study with evidence of under-sequencing [e.g., 115, 116, 204] needs to consider this when interpreting OTU-based patterns, as under-sequencing distorts community profiles [125]. One method to overcome under-sequencing is to disregard OTUs and/or phylotypes occurring below a defined threshold [e.g., 26, 165] but the rare OTUs/phylotypes may be of interest or important to the analyses [3].

Because of all these caveats of generating and interpreting short-read 16s rRNA sequencing data, it is vital the user understands exactly what each step of the analysis pipeline is doing to properly interpret their resulting data. This also permits informed deviation from program defaults to investigate how changes influence downstream steps. Heavily researching the currently used methods prior to any library preparations or analyses will help ensure the user picks the methods most appropriate to their sample types and research questions.

6.2 Whole genome sequencing

With the billions of reads produced from a single next-generation sequencing run, it is now possible to achieve publication-quality genomes using short reads. High-throughput pipelines using a 96-well format allow as many as 384 bacterial genomes to be sequenced at a time (e.g., Nextera XT DNA Library Preparation Kit, Illumina, San Diego, California, USA). This permits data generation for easy whole genome comparisons for investigating the origin and evolution of certain traits, genes, or the strain population itself.

7. Research objectives

7.1. Chapters 2 and 3: Is there a correlation between abundance of the members of the oyster microbiome and *V. parahaemolyticus*?

Given that non-sterilized water more effectively reduces *V. parahaemolyticus* abundance than sterilized water of the same temperature and salinity, it is possible an interaction with the seawater microbiome is facilitating this reduction. Antagonistic relationships or direct resource competition between *V. parahaemolyticus* and bacteria naturally in the seawater, coupled with less than favorable environmental conditions, could explain the greater reduction of *V. parahaemolyticus* in the oysters treated with non-sterilized seawater. I hypothesized there were correlations between abundance of members of the oyster microbiome and *V. parahaemolyticus*. Objective 1 investigated this hypothesis using MPN analyses to quantify *V. parahaemolyticus* and 16s rRNA sequencing to profile the oyster microbiome.

7.2 Chapter 4: Do whole genome comparisons of pathogenic *V. parahaemolyticus* strains reveal strain origins?

ST36 strains caused unprecedented outbreaks on the Atlantic Coast, a region where it was previously not detected. Comparing the whole genomes of ST36 isolates from the Atlantic to other regions may elucidate the origin and evolution of these strains. I hypothesized there were distinct sub-populations of ST36 strains in the Atlantic. In Objective 2, outbreak strains traced to Northeast Atlantic oysters were sequenced using next-generation technologies and their genomes compared to ST36 strains from other regions, and unique genomic content characterized.

CHAPTER 2

Bacterial Community Profiling of Individual Oysters: Taxonomic and Functional Associations with *Vibrio parahaemolyticus* Abundance

Ashley L. Marcinkiewicz^{1,2}, Brian M. Schuster^{2,3}, Stephen H. Jones^{1,4}, Vaughn S. Cooper^{1,2,5}, and Cheryl A. Whistler^{1,2}

¹ Northeast Center for Vibrio Disease and Ecology, University of New Hampshire, Durham, NH, USA

² Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH, USA

³ Seres Therapeutics, Cambridge, MA, USA

⁴ Department of Natural Resources and the Environment, University of New Hampshire, Durham, NH, USA

⁵ Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

ABSTRACT

Oysters naturally harbor the human pathogenic bacterium *Vibrio parahaemolyticus*, but the nature of this symbiosis is unknown. Precedent suggests the abundance of *V. parahaemolyticus* may be influenced by several biotic factors, including bacteria inhabiting the oyster and overlaying water. This study employed 16s rRNA sequencing to profile the microbiome of individual oysters and the overlaying water collected from two distinct sites in the Great Bay Estuary, New Hampshire, and Most Probable Number (MPN) analyses to quantify the levels of *V. parahaemolyticus* in those same oysters. Despite being filter-feeders, oysters are not a direct snapshot of the overlaying water, and likely selectively accumulate bacteria from the surrounding water. Several phylotypes exist in different proportions between the two collection sites, and may be a reflection of differences in site ecology. Individual oysters contain differential *V. parahaemolyticus* abundance, with culture-based methods agreeing with detection patterns from sequencing. Differences in the microbiome between oysters with differential abundance of *V. parahaemolyticus* were likely confounded by differences in collection sites, as all the highest abundance oysters were from a single site. This study was the first of its kind to correlate *V. parahaemolyticus* abundance with members of the oyster microbiome.

KEYWORDS

Vibrio parahaemolyticus, oysters, microbiome, 16s rRNA

INTRODUCTION

Shellfish, including the eastern oyster (*Crassostrea virginica*), are common vectors for the human pathogenic bacterium *Vibrio parahaemolyticus* which can cause self-limiting gastroenteritis [178] when rare pathogenic variants are consumed in large enough quantities [94, 177]. Since it was first identified in Japan following an outbreak in 1950 [79], *V. parahaemolyticus* has caused illness on every inhabited continent [146, 213]. Domestically-acquired cases of illness were relatively uncommon in the United States [15, 48] until a multi-state outbreak in the summer of 1998 caused 23 infections [33]. The number of cases has only increased since [150] and now an estimated at 35,000 cases occur annually [178]. Understanding factors that influence the accumulation of these potentially harmful organisms in shellfish could aid in proactive management to reduce infections.

The United States Food and Drug Administration (FDA) recommends shellfish safe for consumption contain < 10,000 Most Probable Number (MPN)/g *V. parahaemolyticus* [70], but this bacterium often naturally occurs at higher abundance than this, especially in the warm season when people are more inclined to consume raw oysters [e.g., 50, 52, 55, 140, 189]. *V. parahaemolyticus* abundance positively correlates with temperature, but on the other hand, reported correlations with salinity are mixed [50, 52, 55, 140, 189] likely due to the broad range of salinity tolerated by *V. parahaemolyticus* [198]. Strategies intended to reduce *Vibrio* in oysters include rapid cooling [131, 173] or post-harvest processing to decrease *Vibrio* abundance in live product. Some promising treatments include depuration in UV sterilized water and relay of oysters to a site where *V. parahaemolyticus* is of low abundance, often correlating with high salinity. Experiments indicate that, as with *Vibrio vulnificus*, relay of oysters into non-sterilized water is more effective than depuration in sterile water in reducing *V. parahaemolyticus* levels,

implying a potential interaction with the microbial communities in seawater may influence the abundance of *V. parahaemolyticus* in oysters [106, 233].

There is some precedent for biotic factors influencing *Vibrio* spp. abundance. Dinoflagellates [60], diatoms [6, 196], zooplankton [133, 208] and copepods [59] positively correlate with *Vibrio* levels, as these serve either as a nutrient resource or a mechanism of dispersal. *Vibrio* abundance also positively correlates with chlorophyll *a*, suggesting a general interaction with phytoplankton [198, 210]. Protozoa [228] and nematodes (*Terschellinga*, *Molgolaimus*, and *Halolaimus* spp.) negatively correlate with *V. parahaemolyticus* in benthic sediment [216], likely due to grazing. Oyster hemocytes may control the abundance of certain *V. parahaemolyticus* strains, and therefore overall species abundance, as hemocytes differentially kill opaque and translucent strains [82] as well as environmental and clinical strains [217]. *In-vitro* bacterial-*Vibrio* competitions illustrate several types of marine bacteria influence *Vibrio* abundance [94, 177] but *in-vivo* (i.e., in the oyster or water) experiments able to correlate *V. parahaemolyticus* abundance and other bacteria are lacking. Antagonistic relationships coupled with less than favorable salinity conditions could explain the greater reduction of *V. parahaemolyticus* in the oyster microbiome during relay.

Oysters could passively accumulate planktonic and particle-associated *Vibrio* spp. [143, 228] through filter-feeding, but several studies indicate potential preferential accumulation of microbiome communities that include *Vibrio* spp. Culture-based analysis reveal oysters contain over 100 times more *V. parahaemolyticus* per gram than overlaying water [47, 51, 83] and the overall oyster microbiome is more diverse than the overlying water microbiome [23, 156, 70], suggesting selective accumulation by the oyster. Although a few studies have employed next-generation sequencing to profile the oyster and overlying water microbiome [39, 40, 119, 202,

222, 235], none have correlated members of the oyster microbiome to the relative abundance of *V. parahaemolyticus*. Here we profiled the microbiome of 20 individual oysters and overlaying water, 10 each from two naturally occurring, ecologically-distinct oyster beds in the New Hampshire Great Bay Estuary, using 454 sequencing technology, and in parallel quantified *V. parahaemolyticus* abundance of those oysters using MPN analysis. With this approach that contrasts with the standard practice of pooling multiple oysters [71], we 1) examined variation in individual oyster microbiome communities, 2) compared core and variable microbiomes within and between sites, and between oysters and water, and 3) determined whether any phylotypes correlate with *Vibrio* abundance. There were differential *V. parahaemolyticus* levels, and microbiome and functional profiles, between individual oysters collected from each site.

RESULTS AND DISCUSSION

Sequencing the oyster microbiome

To identify the core and variable microbiome among individual oysters, assess variation associated with two distinct, naturally occurring oyster beds within the same estuarine system, and correlate *Vibrio parahaemolyticus* abundance with the associated microbiome, native oysters were collected from two ecologically distinct sites, less than five miles apart in the Great Bay Estuary of New Hampshire, and the bacterial communities sequenced. The Oyster River (OR) oyster bed is located within one of the seven tributaries of this estuary and the harvest area classification is prohibited due to its proximity to the outflow of a municipal waste water treatment plant, whereas the Nannie Island (NI) oyster bed is centrally located within the estuary and is classified as approved. Thus, these two sites may reveal how the different associated ecological factors pertaining to each site influence the microbial community composition.

To define the bacterial community, we generated and sequenced PCR amplicons of the V2-V3 region of the 16s rRNA gene (250bp) from total bacterial DNA isolated from the ten oysters and one overlaying water sample from each collection site. From the generated 1,487,480 reads, only the 512,220 reads with 100% identity to the forward primer and mid-tag were included in the analysis. Quality filtering with FlowClus removed an additional 6,995 reads. QIIME and mothur, which use different algorithms, were applied in parallel to determine the constituent microbial taxa and reveal which pipeline maximized the number of OTUs identified by species. Both programs removed comparable numbers of erroneous reads and generated very similar numbers of OTUs (Table 2.1). However, QIIME assigns taxonomy to an OTU based on a single representative sequence, and classified only 1.3% of OTUs at the species level. mothur, which assigns taxonomy to all sequences in an OTU and chooses the majority consensus taxonomy, resulted in the classification of 35.8% OTUs at the species level (Table 2.2). Because our investigations centered on determining correlations of phylotypes with sites and *V. parahaemolyticus* abundance, analysis continued with the mothur generated dataset. There were an average of 29,391 reads per sample from OR oysters (ranging from 9,670-47,231) and an average of 18,087 from NI oysters (10,338-31,788). The NI water sample had 6,493 reads, whereas OR water had 397, both considerably lower than the corresponding oysters (Table 2S.1).

Table 2.1. Results of select steps in the analysis pipeline with two different analysis programs.

Number of reads after each step in the QIIME and mothur pipelines. QIIME does not perform trimming or aligning steps.

Step in pipeline	QIIME	mothur
Trimmed, misaligned sequences removed	N/A	7,613
Chimeras removed	3,192	3,195
Singleton OTUs removed	9,418	9,189
Reads included in final OTUs	484,600	472,345
Total non-singleton OTUs	5,764	5,756

Table 2.2. OTUs at each level of taxonomic classification with two different analysis programs.

The number (percent) of OTUs at each classification level as determined with QIIME and mothur. Unclassified refers to OTUs that could not be assigned at any classification level.

	QIIME	mothur
Unclassified	1524 (26.4)	0 (0)
Kingdom	4240 (73.6)	5756 (100)
Phylum	4227 (73.3)	5446 (94.6)
Class	4173 (72.4)	5411 (94.0)
Order	3676 (63.8)	4987 (86.6)
Family	2666 (46.3)	4363 (75.8)
Genus	863 (15.0)	3155 (54.8)
Species	74 (1.3)	2059 (35.8)

Analysis of diversity and depth of sampling

Rarified alpha diversity metrics were applied to illustrate both within sample diversity and sufficiency in depth of sampling. The Shannon Index, which uses the proportion of any given OTU relative to the total number of OTUs, indicated that every oyster and water sample was sufficiently sampled and that additional sampling would not increase diversity (Fig. 2.1A).

However, most of the OTUs were rare (68.9% of the OTUs were doubletons) so the relative proportion of any given rare OTU included during rarefaction would likely not change diversity, suggesting this analysis was driven by the most abundant OTUs and not reflective of the true community diversity. The Chao 1 plot, which is commonly used in 16s rRNA studies and incorporates the proportions of singleton and doubleton OTUs, indicated most of the samples were sufficiently sampled (Fig. 2.1B). This, however, is likely an artifact of the removal of singleton OTUs as a means of reducing PCR generated errors. The rarefied PD plot demonstrated that for all samples, the total phylogenetic distances between all OTUs at each subsampling step increased with higher subsampling, indicating inadequate sequencing depth (Fig. 2.1C). As the 97% similarity threshold is not consistent between taxonomic lineages for distinguishing species [91, 180], we argue the PD measurement, incorporating actual sequence data into the diversity measurement, is the most appropriate for 16s rRNA microbiome studies [62, 63]. The overall higher index values in NI samples indicated higher alpha diversity than OR samples, but because the data did not capture total diversity, all interpretations and analyses were considerate of this limitation.

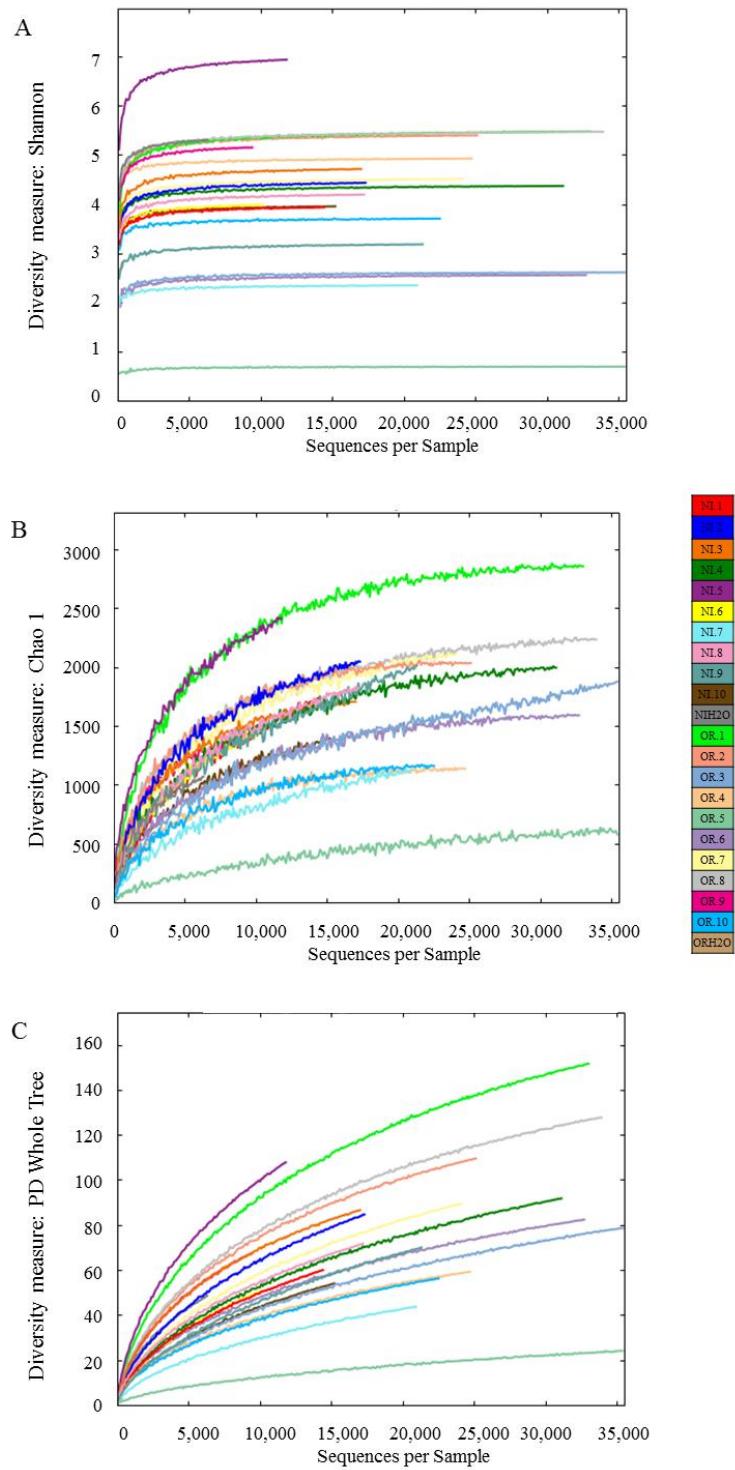


Figure 2.1. Alpha diversity curves for all samples.

Rarified within-sample diversity for individual oyster and overlying water samples generated with three indices: (A) Shannon index (B) Chao 1 and (C) Phylogenetic Distance (PD) whole tree index. The sample ID is indicated by unique color.

Comparison of the distribution of phylotypes by site and substrate

Comparisons of the distribution of OTUs in individual oysters and overlaying water can reveal the extent to which the microbiome of the filter-feeding oyster is reflective of the microbes in overlying water. Oysters from both sites harbor OTUs absent in the respective overlaying water (Fig. 2.2A). However, because there were relatively fewer sequences from the water samples, this lack of correlation could be an artifact of under-sequencing. OTUs present in water and not in oysters, even if not representative of the total diversity, could also suggest differential accumulation. Indeed, 1.35% of the OTUs from NI water and 0.19% from OR water were not detected in any oyster, indicating oysters are not an exact snapshot of the water. This is in agreement with other oyster microbiome studies [36, 195], as well as other filter feeders [124, 197]. Even though our data suggest differential accumulation of the oyster microbiome from water, 82% of the OTUs shared between every NI oyster were also present in overlaying water (Fig. 2.2B), indicating the core microbiome at this site is substantially present in, or possibly influenced by, the water column. Because the OR water sample yielded so few sequences for analysis (397 reads) meaningful comparisons were deemed unlikely.

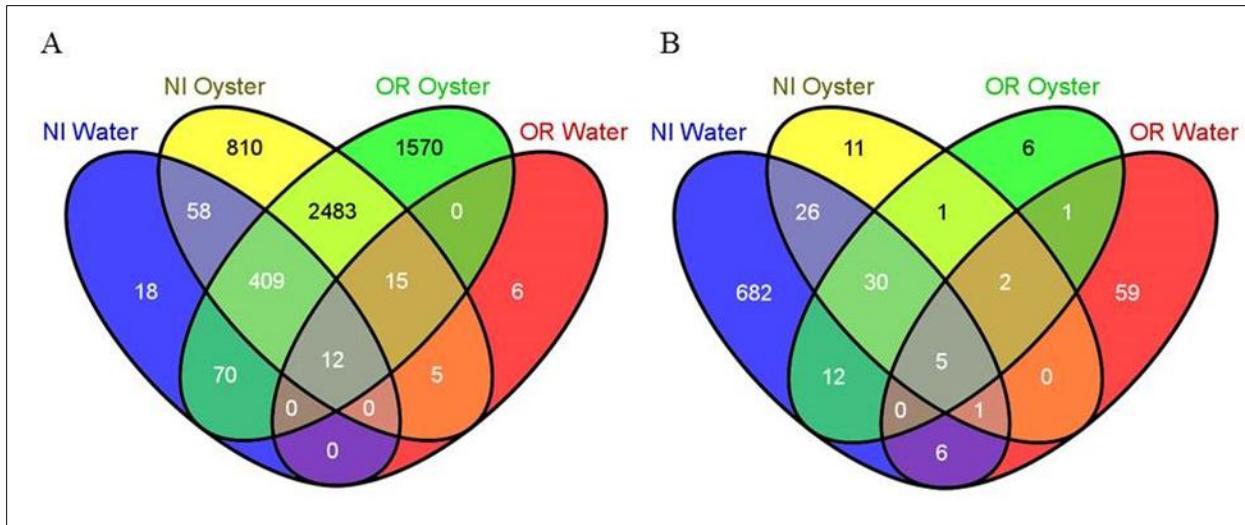
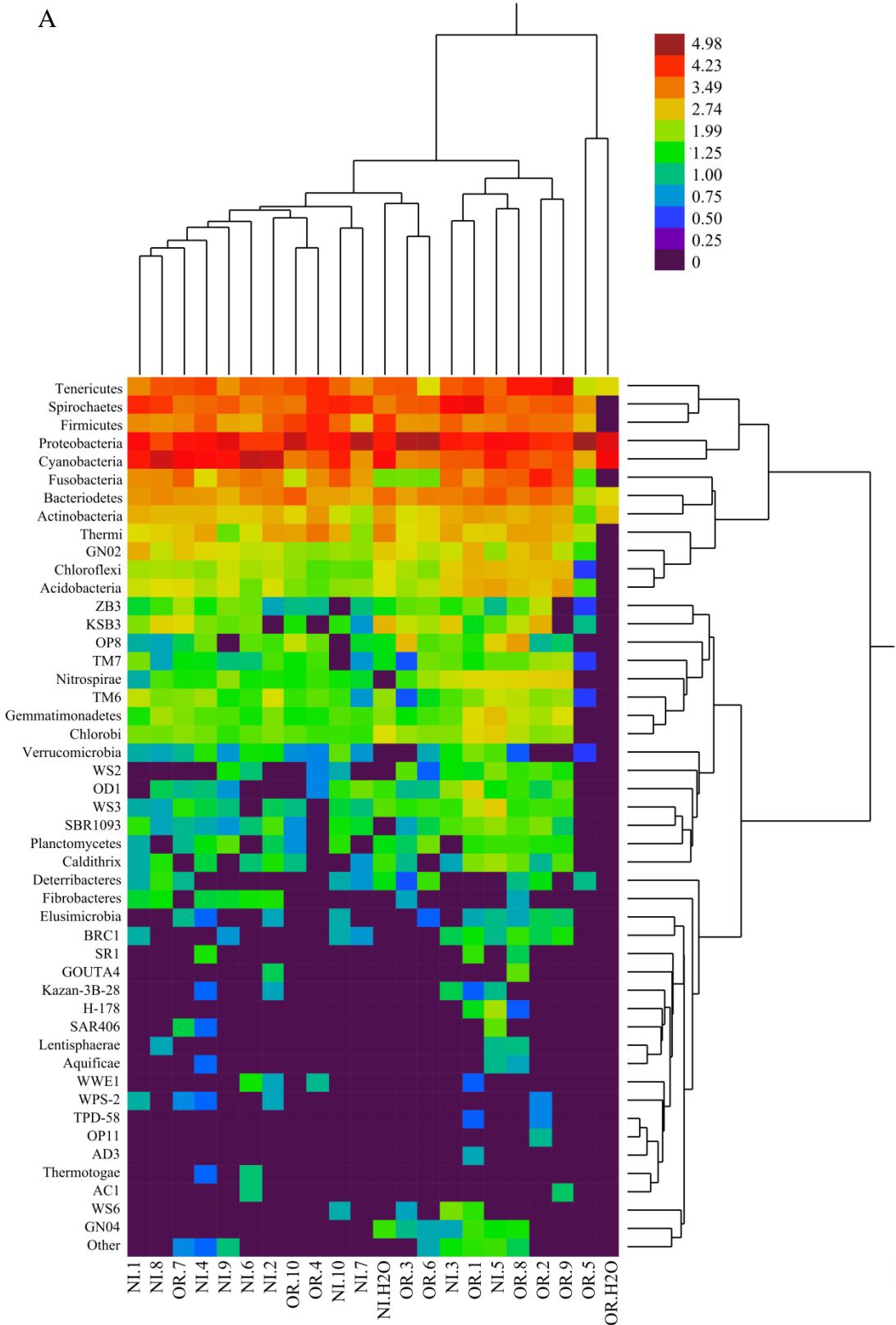


Figure 2.2. Distribution of OTUs in oysters and overlaying water grouped by substrate and site.

Distribution of (A) all OTUs in water and oyster samples and (B) all OTUs in water samples, and OTUs shared between every oyster by site, representing the site-specific and overall core microbiome.

Next we evaluated whether there were informative patterns in the abundance and distribution of phyla-level classifications by hierarchical cluster analysis. This analysis revealed delineation between the microbial communities of oysters by site, when considering both standardized and unstandardized clustering, with only a few exceptions (Fig. 2.3A, 2.3B). The most abundant phyla were consistent with other oyster microbiome studies. For example, Cyanobacteria is the most abundant phylum in Eastern oysters from Apalachicola, Florida [39], whereas the chief phyla in Eastern oysters from two sites in Louisiana include Chloroflexi, Firmicutes, Proteobacteria, and Planctomyces [119]. Proteobacteria and Bacteriodetes dominate the microbiome of *C. corteziensis*, *C. gigas* and *C. sikamea* oysters [202], and the digestive gland of Sydney rock oysters mainly contains Firmicutes, Proteobacteria, Cyanobacteria, and Spirochaetes [235].

A



B

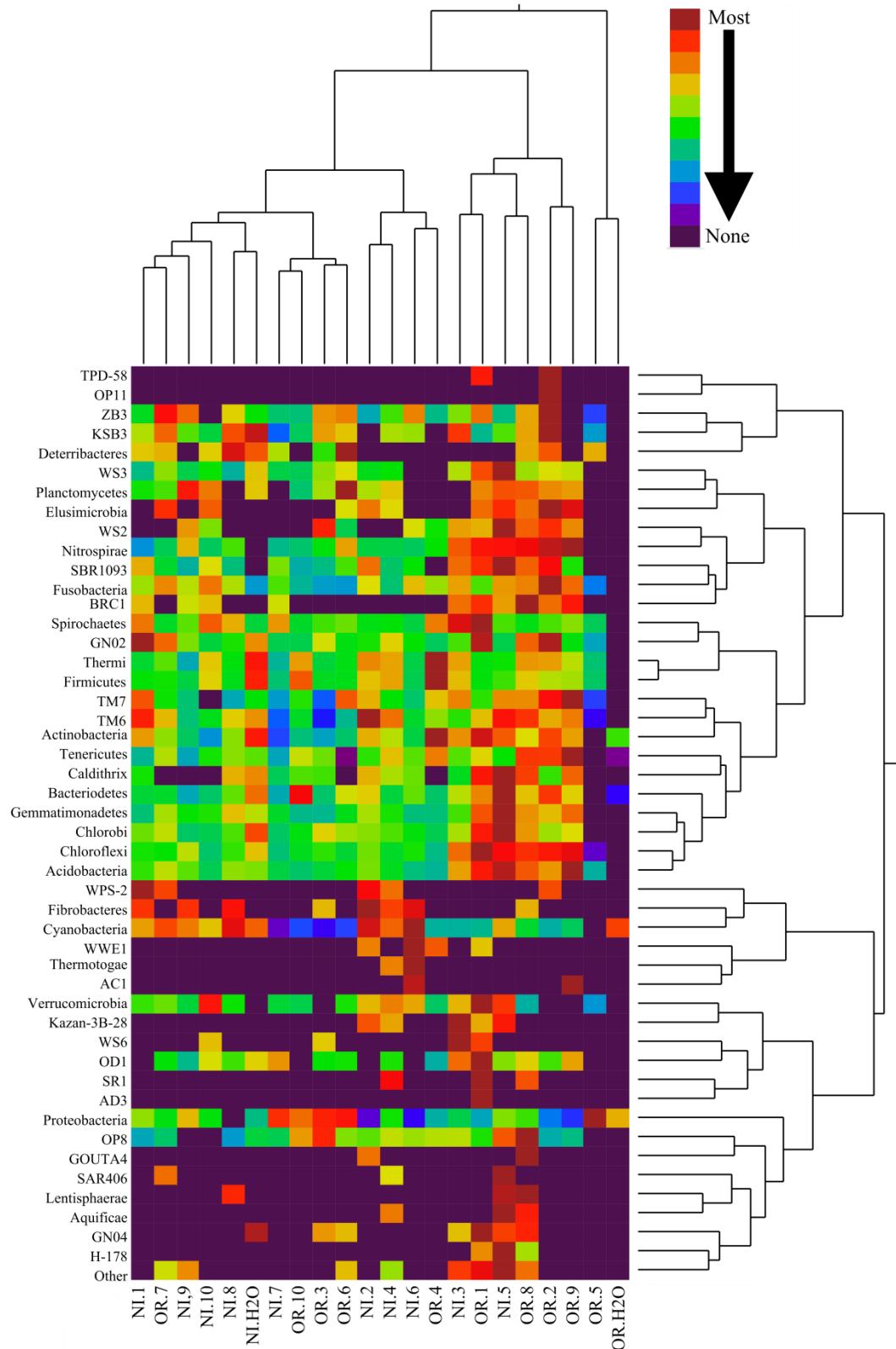


Figure 2.3. Dual hierarchical analysis of phyla-level classification for all samples.

The log-transformed percent abundance of each phylum is indicated by a color scale. Samples and phyla are clustered based on (A) unstandardized and (B) standardized average linkage. In unstandardized linkage, the abundance of each phylum in a given sample is colored based on relative abundance of all phyla, whereas in standardized the abundance of each phylum is colored based on the relative abundance of that phylum across all samples.

Differences between the overlying water and oysters were clearly demonstrated. Most of the variation between sample type, sites, and even individual oysters was explained by not the high abundance, but the low and mid-abundance phyla (Fig. 2.3A), and rarifying the sequences to remove the lowest-abundant OTUs, which is commonly performed [e.g., 119], could remove most of this variation and obscure these apparent site-associated differences.

One notable difference between the two sites is a higher abundance of Cyanobacteria at NI (33.8%, ranging from 1 to 69%) than OR (7.7%, ranging from 0.8 to 43.0%; Fig. 2.3B). Whereas some oyster microbiome studies have discarded cyanobacterial reads to eliminate sequenced chloroplasts from algal matter [119, 202], oysters will ingest Cyanobacteria as a food source [8] and accumulate Cyanobacteria in greater numbers than the surrounding water column [39] justifying retention of these reads as part of the microbiome. Cyanobacteria may even influence the abundance of other members of the oyster microbiome. For instance, Proteobacteria, Bacteriodetes, and Firmicutes have all been isolated from cyanobacterial blooms [16]. It is possible that the specific differences between NI and OR were influenced, at least in part, by the overall higher abundance of Cyanobacteria at NI.

Whereas differential abundances in broad phyla-level classifications reveal general patterns, considering all taxonomic levels with Unifrac uncovered more specific and relevant relationships between samples (Fig. 2.4). Unifrac delineated between sampling sites, with few exceptions, that were consistent with the subsampled jackknife analysis (data not shown). NI

oysters clustered in one branch, whereas OR oysters were distributed to several branches, indicating NI oyster microbiomes, which had overall fewer unique phylotypes than OT, are overall more similar to each other than are OR oyster microbiomes. Specifically, 2.29% of all OTUs in NI oysters are shared between all NI oysters, whereas only 1.25% OTUs in all OR oysters are shared between all OR oysters (Fig. 2.2). Unlike the dual-hierarchical clustering analysis, in the Unifrac analysis the microbiome for NI oyster #7 diverged from the other NI microbiomes, suggesting a classification level deeper than phylum caused this differentiation. NI oyster #7 contained over 15,000 reads (70.8%) assigned to *Nitratireductor pacificus*, whereas the other oysters contained on average 0.9% *N. pacificus* (ranging from 0 to 11.1%), and phyla-level clustering analysis included these as Proteobacteria, making NI oyster #7 appear more similar to NI oysters at that taxonomic level. In the OR water sample, which was quite distant from most samples in both clustering analyses, a proportionally high number of reads were assigned to the genus *Octadecabacter* (43.3%, compared to the average of 0.2% for all other samples, ranging from 0.05% to 0.5%) and the Mamiellaceae family (40.8%, compared to the average of 0.1% for all other samples, ranging from 0.002 to 0.5%). This could explain the distinctiveness of OR water sample and may also reflect low total reads.

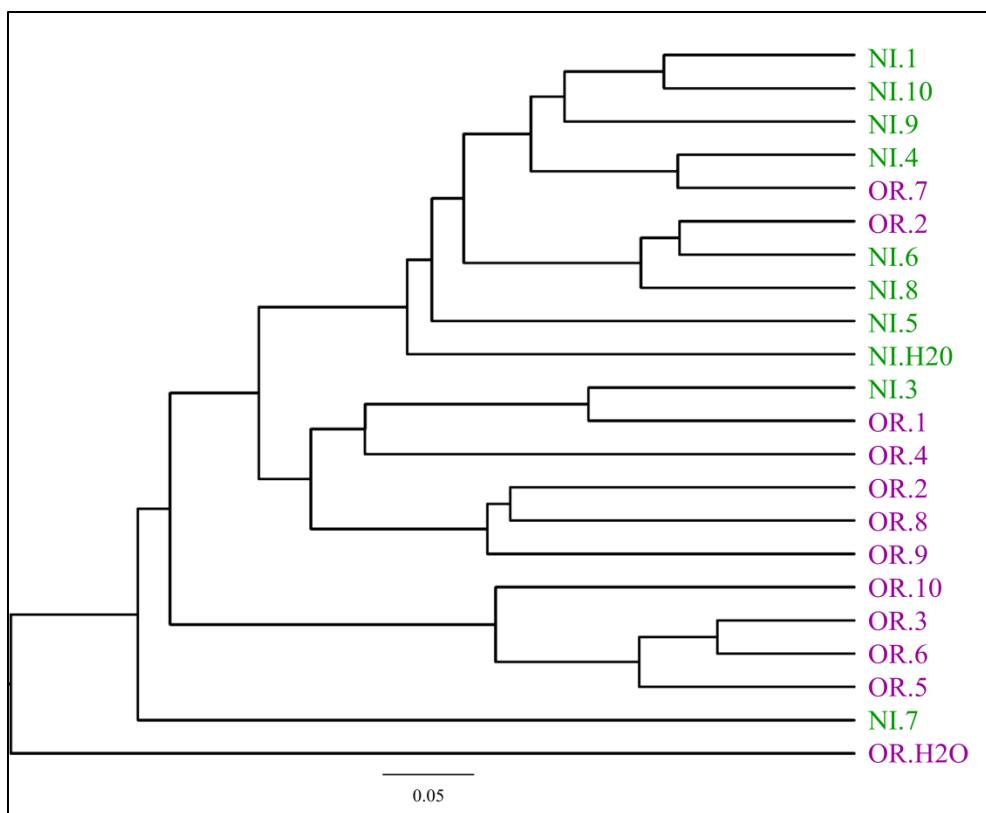


Figure 2.4. Unifrac phylogenetic distance analysis of all taxonomic levels for all samples. Samples are clustered based on the Unifrac distance metric for each pair of samples calculated by the total branch length of unique phylotypes over total branch length of all phylotypes.

The apparent differences in the oyster microbiome between the two sampling sites were further interrogated by employing LEfSe to identify which phylotypes significantly differ by site. The proportions of four phylotypes were significantly higher in OR oysters than NI oysters including: *Finegoldia*, *Bradyrhizobium*, *Roseateles depolymerans*, and *Brevundimonas intermedia* (Fig. 2.5). In contrast, the proportions of eight phylotypes were significantly higher in NI oysters compared to OR including: *Propionigenium*, M2PT2_76, *Reinekea*, *Pseudomonas viridiflava*, *Clostridium sticklandii*, *Vibrio fortis*, *Halobacillus yeomjeoni*, and *Endozooicimonaceae*. *Finegoldia* is typically found in the human gastrointestinal tract [126] and *Bradyrhizobium* is a soil-dwelling, root nodule organism [109], so these associations with OR

are consistent with site being a tidal tributary. Not much is known about the other organisms to draw conclusions on associations by site. However, these were all rare phylotypes, and as such, their higher abundance at one site could reflect sampling bias.

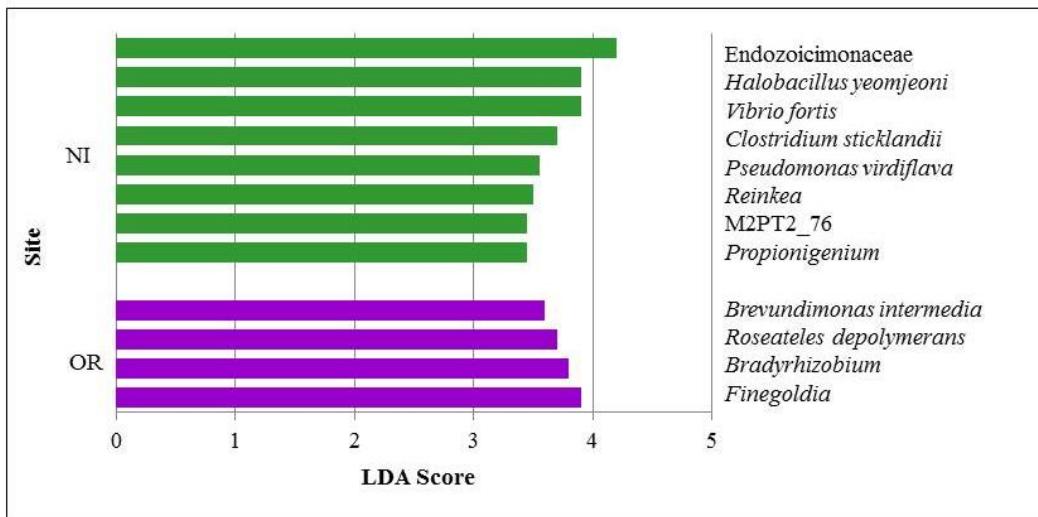


Figure 2.5. LEfSe analysis for oyster samples grouped by site.
Phylotypes in oysters at significantly different proportions at each collection site

Differences in predicted functional profiles between sites

In addition to defining the members of the oyster microbiome, we investigated potential functional differences inferred from phylotype composition between the sampling sites, which may be driven by their unique ecological and environmental associations. Although the bioinformatics tool PICRUSt predicts functions associated with 16s identity, it draws upon whole-genome sequenced species from NCBI and functional data from KEGG databases, so therefore can only predict functions associated with previously characterized species and pathways [123]. A total of 887 predicted gene functions significantly differed between NI and OR oysters ($p < 0.05$). Of these, 119 were $p < 0.005$, and 11 were $p < 0.0005$. Further

examination of the functional differences at $p < 0.0005$ reveals two distinct classes (Fig. 2.6).

OR oysters had a higher number of functions generally involved in cell growth, including nucleotide metabolism, tRNA synthesis and associated elongation factors, amino acid biosynthesis, and oxidative phosphorylation. NI oysters had a higher number of diverse metabolic functions (sugar, chlorophyll, carbon, and sulfur metabolism) as well as higher number of chaperone associated proteins. The higher chlorophyll metabolism logically related to the greater number of 16s sequences identified as Cyanobacteria at NI compared to OR. Overall, these variations could relate to nutrient concentrations at the two different sites. OR is impacted by an upstream tidal wastewater facility discharge, more directly influenced by rainfall/runoff events and nonpoint source pollution, and has slightly higher dissolved nutrients and suspended solids compared to NI (Table 2.3). NI is a much larger oyster bed with abundant oyster culch on coarser textured sediment compared to OR. The slightly more readily available nutrients at OR may support more rapid total bacterial growth, whereas the complex nutrient and energy sources available at NI may support a more diverse bacterial population.

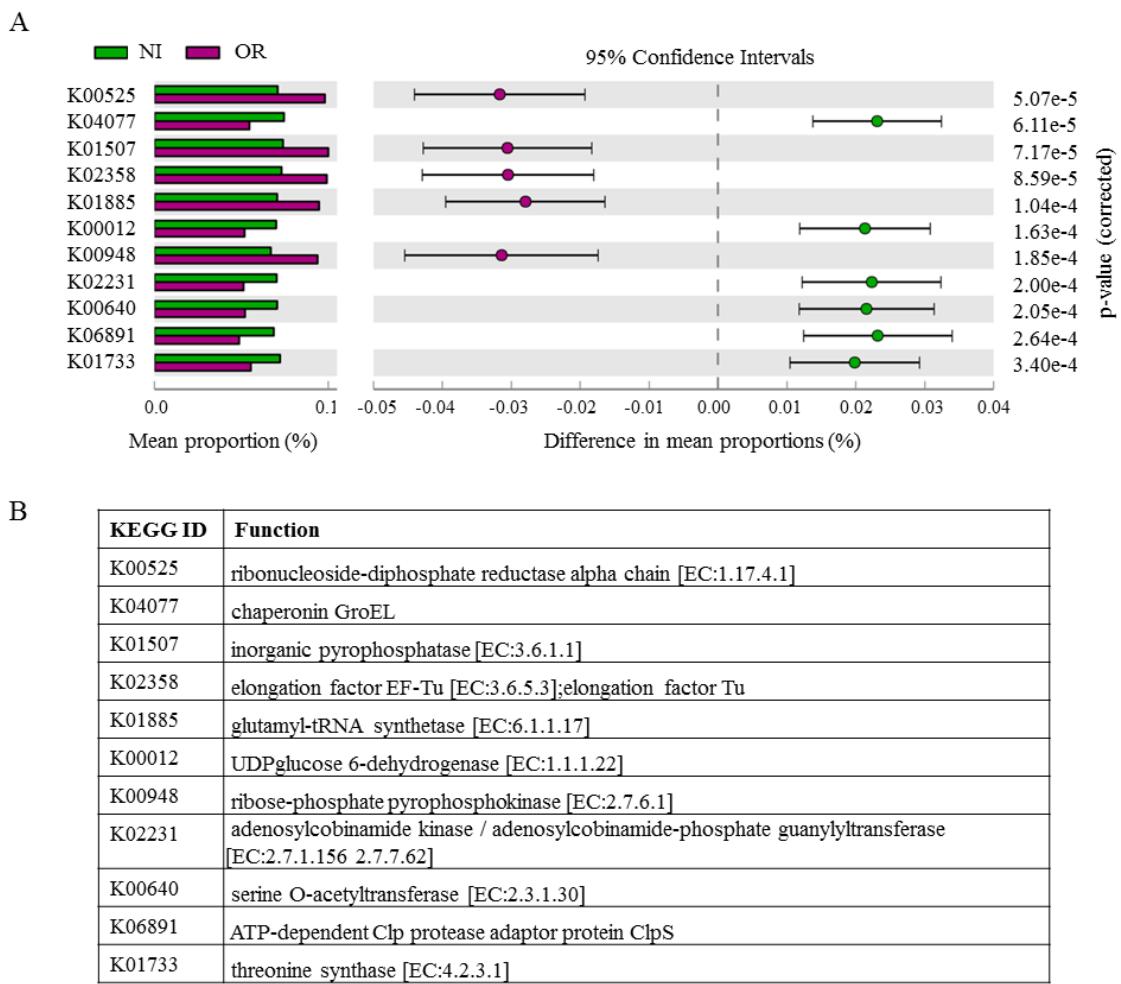


Figure 2.6. Annotated functions in significantly different numbers at each site.

(A) PICRUSt-derived KEGG orthology IDs at significant different ($p < 0.0005$) numbers at each site, and (B) the pathways associated with each ID.

Table 2.3. Nutrient levels at each collection site.

SONDE-collected nutrient levels from each collection site, representing the average of readings every 15 minutes from 2007-2013.

	Phosphate	Ammonium	Nitrogen	Turbidity	Chlorophyll <i>a</i>
NI	0.023	0.059	0.082	0.121	4.235
OR	0.051	0.065	0.123	0.177	4.614

Abundance of *V. parahaemolyticus* in individual oysters and correlations with microbiome

To evaluate potential correlations of microbiome with *V. parahaemolyticus*, we applied an enrichment and qPCR-based enumeration method of *V. parahaemolyticus* to individual oysters. Our approach was notably different from the standard practice of using a pooled oyster homogenate to capture the average number of bacteria by the MPN method, but allowed more critical comparison and evaluation of correlations between relative abundance of *V. parahaemolyticus* and phylotypes in individual oyster microbiomes. Individual oysters, even from the same site, differed dramatically in abundance of *V. parahaemolyticus* (Table 2.4). This has also been seen in individual oysters in Alabama [114]. To aid in additional comparisons, these oysters were grouped based on MPN/g abundance level, where the means of each group significantly differed from the other groups (Low: 0.48, Medium: 1.16, High: 2.51; $p < 0.0001$). *V. parahaemolyticus* was only captured via 16s sequencing in the medium and high abundance level oysters from NI (Table 2.4) but was a rare component of the oyster microbiome and as such could be influenced by under-sequencing. Whereas these independent methods did not align with regard to exact concentration in individual oysters, the general pattern of *V. parahaemolyticus* 16s rRNA detection aligned with culture-based methods, supporting the applied sequencing methods.

Table 2.4. Distribution of *Vibrio parahaemolyticus* in oyster samples as determined by MPN and 16s sequencing.

Oysters are separated by site, and ordered by increasing *V. parahaemolyticus* count as determined by MPN. 16s reads represents the number of *V. parahaemolyticus* sequences and relative percent abundance in parentheses.

Oyster	Log10 MPN/g	Classification	16s Reads
NI.10	0.52	Low	-
NI.5	0.72	Low	-
NI.7	1.01	Medium	-
NI.8	1.01	Medium	2 (0.011)
NI.1	1.34	Medium	3 (0.020)
NI.6	1.34	Medium	-
NI.2	2.38	High	-
NI.3	2.38	High	2 (0.011)
NI.4	2.38	High	1 (0.003)
NI.9	2.88	High	1 (0.009)
OR.10	0.13	Low	-
OR.4	0.28	Low	-
OR.6	0.28	Low	-
OR.2	0.49	Low	-
OR.8	0.52	Low	-
OR.5	0.93	Low	-
OR.1	1.01	Medium	-
OR.3	1.01	Medium	-
OR.7	1.20	Medium	-
OR.9	1.34	Medium	-

NI harbored the only high abundance level oysters, but NI also harbored low abundance level oysters. In contrast, OR contained only medium and low abundance level *V. parahaemolyticus* oysters. This suggested a high degree of oyster to oyster variation in *V. parahaemolyticus* abundance, perhaps driven in part by site differences. There are some differences in long-term nutrient loads between the two sites (Table 2.3), but with the exception

of chlorophyll *a*, there are no correlations between these nutrients and *V. parahaemolyticus* presence in the GBE [205]. Chlorophyll *a* positively correlates with *V. parahaemolyticus* presence [205, 198], yet was overall higher at OR. While temperature always positively correlates with *V. parahaemolyticus* abundance, and salinity sometimes correlates [50, 52, 55, 140, 189], the short-term environmental conditions averaged over the 12 hours prior to oyster harvest were essentially identical between the two sites (Table 2.5). It is therefore unlikely that these measured abiotic parameters drove higher levels of *V. parahaemolyticus* at NI in a subset of oysters, or comparatively lower levels of *V. parahaemolyticus* at OR.

Table 2.5. Environmental conditions at each sampling site.

SONDE-collected environmental conditions from each collection site, representing the average of readings every 15 minutes from the 12 hours prior to collection.

Site	Temperature	Salinity	DO %	DO mgL	pH	Turbidity
Nannie Island	20.8	20.8	82.8	6.6	7.5	12.1
Oyster River	20.8	20.7	81.8	6.5	7.4	12.4

To investigate any biotic correlation with *V. parahaemolyticus* abundance, microbiome data for individual oysters were analyzed with Unifrac distance trees to determine similarity of the microbiome of oysters in the same MPN abundance level group, separated by site. Branching patterns did not correspond with *V. parahaemolyticus* abundance level (data not shown), indicating there is no overall similarity in the microbial community correlating with *V. parahaemolyticus* abundance. Despite a lack of clustering of samples by *V. parahaemolyticus* abundance level, there were 24 phylotypes significantly higher in number in high abundance level oysters, one phylotype in medium abundance level oysters, and three in low abundance level oysters (Fig. 2.7). However, a caveat to this data and its interpretation is that these were all

rare phylotypes of which the proportion could be influenced by under-sequencing, in addition to the relationships potentially being confounded by site-specific differences. Specifically, the 19 phylotypes that were exclusive to high abundance level oysters (and by default present in significantly higher proportions) could be an artifact of differences between collection sites, as there were no high abundance level OR oysters for comparison. PICRUSt was again used to evaluate differences in functional capabilities of microbiomes between oysters containing different abundance levels of *V. parahaemolyticus*. In total, there were 320 predicted functions that differed significantly by relative *V. parahaemolyticus* abundance level ($p < 0.05$), ten of which also correlated with *V. parahaemolyticus* class (Table 2S.2). However, because these ten functions were also significantly different between sites, they are compounded by associated differences between the two sites. The differences in the microbiome between the sampling sites, rather than *V. parahaemolyticus* abundance, were a driving force in this study.

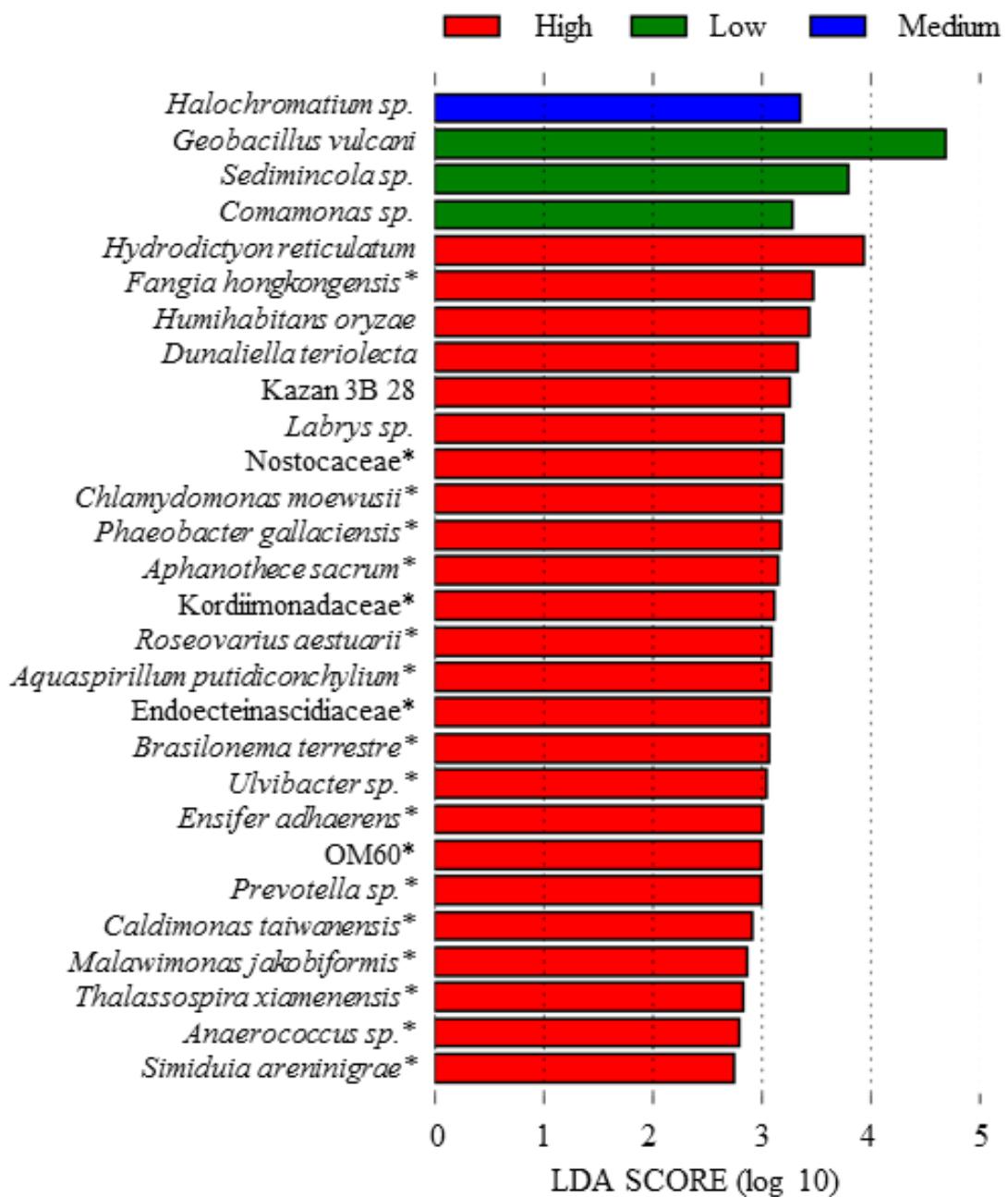


Figure 2.7. Phylotypes significantly different based on *Vibrio parahaemolyticus* abundance class.

PICRUSt-derived phylotypes in oysters at different proportions by *V. parahaemolyticus* abundance class. Phylotypes followed by an * were only found in high oysters.

Even though there were insufficient individual oysters at each *V. parahaemolyticus* abundance level and under-sequencing for a robust analysis of differences in the taxonomic profile or functional capabilities of microbiomes of oysters associated with *V. parahaemolyticus* abundance, this study does indicate the general approach may reveal phylotypes and functional differences associated with *V. parahaemolyticus* abundance by applying deeper sampling to more oysters.

CONCLUSIONS

In this study, we profiled the microbiome and quantified *Vibrio parahaemolyticus* abundance of ten oysters and overlaying water from each of two oyster beds from the Great Bay Estuary in New Hampshire. We determined individual oysters and collection sites have differences in the taxonomic and functional microbiome profiles. Several studies conclude the microbiomes of marine animals are highly specific based on individuals' surrounding habitat [201, 235] and diet [84] so it is unsurprising there is a different microbiome between oysters at Nannie Island and Oyster River, located at ecologically distinct, albeit geographically proximal areas within the same estuary.

Both culture and non-culture based methods revealed *V. parahaemolyticus* did not equally accumulate in individual oysters, despite the oysters exposure to the same environmental conditions at each site. It is therefore unlikely these measured environmental conditions contributed to the differences in *V. parahaemolyticus* levels. Bivalves actively filter water based upon particle size [220] bacterial species [19, 27], strains within the same species with known or introduced (i.e. mutations) genetic variation [138, 144, 162, 191], and even viral particles [190]. The *V. parahaemolyticus* strains themselves may contain genetic factors or phenotypic traits

influencing uptake and/or depuration and it is possible different strains are accumulated at different rates, much like *Vibrio vulnificus* [77, 221].

The higher abundance of Cyanobacteria at NI may influence the abundance of other phyla at this site [16]. It may also explain the higher abundance of *V. parahaemolyticus* at NI. Cyanobacteria and *V. cholerae* will associate [102, 103], and *Vibrio* spp. make up as much as 6% of all cultivable heterotrophic bacteria isolated from cyanobacterial blooms [16]. In addition, cyanobacterial-derived organic matter increases *Vibrio* abundance [59]. With this study, we were unable to identify specific species that always correlate with *Vibrio* abundance. However, there are several potential species that, with increased sample size, more targeted sampling, and deeper sequencing, may be able to be significantly tied to *V. parahaemolyticus* abundance.

METHODS

Oyster Collection and Processing

One water and ten oyster samples were collected at low tide on September 1st, 2009 from two distinct naturally-occurring oyster beds in the New Hampshire Great Bay Estuary (GBE), one (Nannie Island) within an area approved for recreational shellfish harvesting, and the other (Oyster River) falling within an area prohibited for recreational harvesting due to its proximity to a wastewater treatment facility effluent outfall. Oysters were collected using oyster tongs whereas water samples were collected by submerging capped sterile bottles ~0.5m below the water surface and uncapping to fill. Samples were immediately stored on ice packs in coolers until laboratory processing. Environmental and nutrient conditions per each site were assessed from the NOAA National Estuarine Research Reserve System (<http://nerrs.noaa.gov/>) which measures conditions every 15 minutes by YSI datasondes. Short-term environmental conditions,

including temperature, salinity, dissolved oxygen, pH, and turbidity were averaged for the 12 hours prior to sampling. Long-term nutrient patterns were assessed by averaging all readings from 2007-2013.

Individual oysters were cleaned, aseptically shucked, and thoroughly homogenized with a surface disinfected (using 90% ethanol and filter sterilized water) Tissue Tearor (Biospec Products, Bartlesville, OK). Most Probable Number (MPN) analyses were performed on individual oyster homogenate and water samples as described in Schuster *et al.* [184]. In brief, samples were serially diluted tenfold into Alkaline Peptone Water (APW) and incubated at 37°C for 16 hours, and the tubes scored by turbidity. To positively identify the presence of *V. parahaemolyticus*, 1.0mL of each turbid dilution was pelleted, and the DNA obtained by a CTAB-NaCl precipitation followed by phenol-chloroform extraction [7], which was used to score MPN results with qPCR as described below. Remaining water samples were spun in a 5810R centrifuge (Eppendorf, Hamburg, Germany) at 4,000 rpm; the supernatant was discarded. The water bacterioplankton pellet and unenriched oyster homogenate not immediately used for MPN analysis were frozen at -80°C.

MPN/g enumeration

MPN tubes were scored as positive for *V. parahaemolyticus* by detection of the thermolabile hemolysin gene (*tlh*) with qPCR [153]. The reaction contained 1x iQ Supermix SYBR Green I (Bio-Rad, Hercules, CA) and 2µL of the DNA template in a final volume of 25µL. An iCycler with the MyiQ Single Color Real-Time PCR Detection system with included software (Bio-Rad, Hercules, CA, USA) was used with the published cycling parameters [153]. A melting curve was performed to ensure positive detection of the correct amplicon compared to

a control DNA sample (*V. parahaemolyticus* F11-3A). MPN tubes were scored as positive or negative based on whether qPCR starting quality values were below (negative) or above (positive) the threshold value determined by the standard curve using purified F11-3A and water blank with iCycler software. The *V. parahaemolyticus* MPN/g was calculated for each oyster according the FDA BAM [71] and grouped by high, medium, or low abundance level based on 10-fold differences in MPN/g.

Metagenetic preparation

Metagenetic DNA was isolated from archived oyster homogenates. The homogenates were thawed on ice for 10 minutes, the top ~1cm was aseptically removed and discarded, and 1.0g of each oyster homogenate was aseptically collected. The entire bacterioplankton pellet was used for the water samples. The total bacterial DNA was extracted using the E.Z.N.A. Soil DNA Kit (Omega Bio-Tek, Norcross, GA, USA) following standard protocols for Gram-negative and -positive bacterial isolation.

The V2 to V3 region of 16s rRNA gene was amplified from each individual sample in triplicate using PCR with standard 16s F8 (5' – AGTTTGATCCTGGCTCAG – 3') with GS FLX Titanium Primer A (5' – CGTATCGCCTCCCTCGCGCCATCAG – 3') and R357 (5' – CTGCTGCCTYCCGTA – 3') with Primer B (5' – CTATGCGCCTGCCAGCCCCGCTCAG – 3'), with each pair of corresponding forward and reverse primer sets having a unique 6bp MID tag [125]. The PCR reaction containing 45µL Platinum PCR Supermix (Invitrogen, Carlsbad, CA, USA), 3µL of sample DNA, and 2µL molecular grade water, was ran in an iCycler thermocycler (Bio-Rad, Hercules, CA, USA) at the following conditions: 94°C for 90 seconds; 30 cycles of 94°C for 30 seconds, 50.7°C for 45 seconds, 72°C for 30 seconds; and 72°C for 3

minutes. The triplicate samples were combined and then purified using the MinElute PCR Purification Kit (Qiagen, Valencia, CA, USA) following standard protocols. Each purified sample was visualized on a 1.2% agarose gel to ensure purity and quality including expected amplicon size.

A 10ng/mL multiplexed sample was prepared for the Roche Genome Sequencer FLX System using Titanium Chemistry (454 Life Sciences, Branford, CT, USA). The DNA concentration for each sample was quantified using a NanoDrop 2000c (Thermo Scientific, Wilmington, DE, USA) and pooled with equal proportions of the twenty oyster and two water samples. The pooled mixture was purified using the AMPure XP Purification Kit (Beckman Coulter Genomics, Danvers, MA, USA) by manufacturers protocols, with the final samples suspended in 20uL elution buffer EB from the MinElute PCR Purification Kit (Qiagen, Valencia, CA, USA). The pooled tagged single-stranded pyrosequencing library underwent fusion PCR and pyrosequencing using a Roche 454 FLX Pyrosequencer (454 Life Sciences, Branford, CT, USA) according to the manufacturer instructions at the University of Illinois W.M. Keck Center High-Throughout DNA Sequencing Center.

Metagenetic analysis

The forward 454 pyrosequencing reads were quality filtered and denoised to reduce erroneous PCR and sequencing errors using FlowClus, setting zero primer and barcode mismatches, a minimum sequence length of 200, zero ambiguous bases and seven homopolymers allowed before truncation, a minimum average quality score of 25, and k=5 for the flow value multiple [81]. These sequences were then further filtered and clustered with both QIIME 1.8 [30] and mothur 1.22.0 [181] to select the pipeline that would produce the most

OTUs classified to the species level. In QIIME, *de novo* chimeric sequences were identified with usearch 6.1 [56] split by sample. *De novo* operational taxonomic units (OTUs) were picked from non-chimeric sequences using usearch 6.1 [56] with maximum rejects set to exceed the number of sequences, and maximum accepts set to 20. Taxonomy was assigned to the representative sequence (most abundant) of each OTU using greengenes 13.8 [53]. Singleton OTUs were deleted to further reduce PCR artifacts. The mothur workflow followed the 454 SOP accessed September 2014 [181] with some modifications. The pre-clustering step was performed permitting one difference. Chloroplasts were retained, as cyanobacteria have previously been identified as part of the oyster microbiome [39, 235]. greengenes 13.8 [53] was used to assign taxonomy to OTUs. After removing singleton OTUs, mothur 1.33.0 [181] was used to generate a distance matrix, pick representative OTUs, and create a phylogenetic tree using clearcut 1.0.9 [186] for determining alpha diversity.

Rarified alpha diversity measurements were calculated with QIIME 1.8 [30] to determine both the within-sample diversity and sequencing depth using three index measurements. The Shannon Index, a traditional ecological diversity measurement, incorporates relative proportions of both species abundance and evenness. Chao 1 considers the proportion of singleton and doubleton OTUs relative to the total number to estimate diversity. Whole-tree phylogenetic diversity (PD) uses a phylogenetic tree of all OTUs to assess species richness where index values reflect relatedness of OTUs, or more specifically inferred phylogenetic distance of OTUs, within a sample [62, 63]. All indices were calculated with ten iterations of 100 reads added at each rarefaction step, up to 75% of the sample with the highest number of reads. The distribution of OTUs between sampling sites and substrates was determined with Venny 1.0 [158]. Patterns in abundance in phyla-level classifications in all samples were revealed with a dual-hierarchical

clustering performed with JMP 12 (SAS Institute Inc., Cary, North Carolina, USA) for log-transformed percent abundance using both standardized and unstandardized average linkage. Weighted and normalized Fast Unifrac [130], which uses all levels of taxonomic assignment to create a distance matrix and groups samples based on similarity, was used to perform beta diversity clustering and jackknife analyses for samples, jackknifing at 1000 permutations at 75% of the sample with the lowest number of reads. LEfSe [185], PICRUSt [123] and STAMP [164] were all used at default settings, to determine taxonomic and profile similarities between sample groups, and calculate statistical significance, respectively, pre-normalizing samples to 1M in LEfSe.

To compare the sequenced-based abundance of *V. parahaemolyticus* to abundance quantified with the culture-based MPN method, all quality-filtered, denoised reads were aligned to the region of *V. parahaemolyticus* strain RIMD 2210633 (GCA_000196095.1) that would be amplified by the F8-R357 primer pair at 99.0% with PyNast [29] through QIIME 1.8 [30]. The identity of matching sequences was confirmed with BLAST [4].

ACKNOWLEDGEMENTS

We thank Crystal Ellis and Anna Tyzik for assistance with MPNs, Jessica Jarett and John Gaspar for assistance with learning bioinformatics software, and W. Kelley Thomas for philosophical debates.

SUPPLEMENTARY INFORMATION

Table 2S.1. Number of reads per sample.

The number of reads in each oyster and water sample derived using the mothur pipeline.

Oyster	NI	OR
1	14786	34420
2	17900	25737
3	17478	36390
4	31802	25115
5	12500	47231
6	10390	33229
7	21227	24765
8	17670	34639
9	21719	9672
10	15468	22802
Water	6495	397

Table 2S.2. Functions positively correlated with *Vibrio parahaemolyticus* class.

PICRUSt-identified KEGG IDs significantly different and positively correlated ($p < 0.05$) between oysters of different *V. parahaemolyticus* classes.

KEGG ID	% abundance			p-value
	High	Medium	Low	
K00525	0.089	0.091	0.117	0.040
K02231	0.081	0.075	0.058	0.044
K02628	0.00036	0.00017	0.00007	0.040
K02629	0.00036	0.00017	0.00007	0.040
K02630	0.00036	0.00017	0.00007	0.040
K02631	0.00036	0.00017	0.00007	0.040
K02632	0.00036	0.00017	0.00007	0.040
K05587	0.0004	0.0002	0.00012	0.029
K05886	0.00022	0.00013	0.00006	0.035
K07358	0.00016	0.00005	0.00001	7.50e-3

CHAPTER 3

Correlations between Abundance of *Vibrio parahaemolyticus* and the Microbiome of the Eastern oyster (*Crassostrea virginica*)

Ashley L. Marcinkiewicz^{1,2} and Cheryl A. Whistler^{1,2}

¹ Northeast Center for Vibrio Disease and Ecology, University of New Hampshire, Durham, NH, USA

² Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH, USA

ABSTRACT

Shellfish, including the eastern oyster (*Crassostrea virginica*), are common vectors for the human pathogenic bacterium *Vibrio parahaemolyticus*. Individual oysters, even from the same collection site, do not contain the same levels of *V. parahaemolyticus* but it is unclear what factors influence this variability. Based on previous analyses, we hypothesize that bacteria composing the oyster microbiome may influence *V. parahaemolyticus* abundance. To investigate potential correlation of abundance with the microbiome, we used Most Probable Number analyses to quantify the levels of *V. parahaemolyticus* in oysters collected on three dates from the Great Bay Estuary in New Hampshire, and from these same individual oysters sequenced the V4 region of the 16s rRNA gene to profile the bacterial community of those oysters. From this analysis we evaluated whether phylotypes correlated with *V. parahaemolyticus* abundance. Oysters contained *V. parahaemolyticus* in low, medium, and high abundance levels, providing a basis for these comparisons. Our analysis indicated the microbiome of oysters correlated far more strongly with collection date than with *V. parahaemolyticus* abundance, but there were significant differences in the concentration of several phylotypes between high and low *V. parahaemolyticus* abundance levels, particularly *Bacillus* species. With future studies to confirm the relationship, these phylotypes may represent potential alternative indicator and/or antagonistic organisms against *V. parahaemolyticus*.

KEYWORDS

Vibrio parahaemolyticus, oysters, microbiome, 16s rRNA

INTRODUCTION

The human pathogenic bacterium *Vibrio parahaemolyticus* is a constituent of marine waters and forms symbiosis with animals inhabiting those waters, including the eastern oyster, *Crassostrea virginica* [154]. The nature of this symbiosis is unknown, and provides an opportunity to evaluate the ecology of a pathogen in its natural environment as it interacts with other bacteria. Previous studies [114, 135] illustrate the levels of *V. parahaemolyticus* in oysters varies, even from the same collection site with exposure to the same environmental conditions. The microbiome of the surrounding seawater, and the microbiome of the oyster itself, may influence the levels of *V. parahaemolyticus* in individual oysters [106, 233].

Microorganisms in the oyster microbiome that positively correlate with *V. parahaemolyticus* abundance could represent potential alternative indicator organisms, as *V. parahaemolyticus* comprises a small proportion of the oyster microbiome [135] and detection methods require enrichment [71]. Microbes that inversely correlate with *V. parahaemolyticus* abundance could potentially antagonize *V. parahaemolyticus*, especially when paired with less than favorable salinity conditions during relaying oysters to reduce *Vibrio* levels [106, 233]. Indeed, even though oysters establish a resilient microbiome in the larval stage that is resistant to depuration [MP Doyle and JD Oliver, unpublished, as cited by Froelich *et al.* [76]), high salinity allows bacteria more tolerant to these high levels to outcompete *V. vulnificus* in oyster colonization [78, 193].

Here we determined the abundance of *V. parahaemolyticus* in 25 individual oysters harvested during three collection dates, from early July through late August 2014 from a single site. From these oysters, we identified a subset harboring high, medium, and low *V. parahaemolyticus* abundance levels, for targeted profiling of their constituent microbiome to

elucidate correlations between *V. parahaemolyticus* abundance and community members. The microbiome of these oysters correlated more strongly with collection date than with *V. parahaemolyticus* abundance. Different members of the *Bacillus* spp. both negatively and positively correlated with *V. parahaemolyticus* abundance, suggesting species-specific interactions that, with further investigation into the exact relationship, may be good alternate indicator organisms and/or mechanisms of reducing *V. parahaemolyticus* abundance.

RESULTS AND DISCUSSION

Individual oysters harbor differential abundance of Vibrio parahaemolyticus

To facilitate analysis of correlations of microbiome with pathogen abundance, the most probable number (MPN) of *V. parahaemolyticus* was determined for individual oysters, collected from a single site collected on three dates in the summer of 2014 (July 7, July 30, and August 20). These dates were selected as they captured the warmest season in the Great Bay Estuary, New Hampshire, when *V. parahaemolyticus* levels are the highest and the chances for collecting high *V. parahaemolyticus* abundance oysters for targeted comparisons was the greatest. Using individual oysters rather than pooled as is standard [71] allows fine-scale resolution of both the *Vibrio* abundance and the bacterial communities composing oysters that would be masked in pooling samples, especially when evaluating differences between sampling dates only weeks apart.

As expected [114, 135], individual oysters contained different levels of *V. parahaemolyticus* abundance (Table 3.1). To aid in statistical comparisons, samples were grouped into three abundance levels, where means of each group significantly differed from the other groups (Low: 1.30, Medium: 2.38, High: 3.27, p < 0.0001). Our goal was to target only

high and low *V. parahaemolyticus* abundance oysters based on initial MPN tube turbidity, but this method was ineffective and resulted in almost half of all oysters with medium *V. parahaemolyticus* abundance.

Table 3.1. Distribution of *Vibrio parahaemolyticus* in oyster samples.

Oysters are separated by date, identified by sample number, and ordered by increasing *V. parahaemolyticus* abundance as determined by MPN. Length is the shell from the umbo to growth edge rounded to the nearest 0.5cm. A subset of oysters was used for microbial analysis.

Date	Oyster	Length	Log10 MPN/g	Abundance level	Sequenced for microbial analysis
7-Jul-14	11	7.5	0.477	Low	N
	20	9	1.362	Low	Y
	2	6.5	1.964	Low	N
	13	6	2.380	Medium	N
	16	5.5	2.380	Medium	Y
	5	8	2.869	High	N
	15	6.5	2.964	High	N
	19	7.5	2.964	High	Y
30-Jul-14	19	7.5	0.964	Low	Y
	3	8	1.041	Low	Y
	15	7	2.176	Medium	Y
	2	9	2.322	Medium	Y
	9	9	2.380	Medium	Y
	18	9.5	2.380	Medium	Y
	22	5.5	2.663	Medium	Y
20-Aug-14	5	6	1.176	Low	Y
	9	10.5	1.633	Low	Y
	21	10	1.633	Low	Y
	22	7.5	2.322	Medium	Y
	3	7	2.380	Medium	Y
	4	6.5	2.380	Medium	Y
	7	5	2.380	Medium	N
	24	7	2.380	Medium	Y
	11	7	2.968	High	Y
	23	8.5	3.869	High	Y

Assembling the oyster microbiome

After quantifying the abundance of *V. parahaemolyticus* in individual oysters, we correlated *Vibrio* abundance with the oyster microbiome, identified the core and variable microbiome among individual oysters, and assessed microbiome variation over multiple sampling dates. The total bacterial DNA was extracted from select individual oysters (due to limited space in the sequencing run, but we included samples from each *V. parahaemolyticus* abundance level from each sampling date), and the V4 of the 16s rRNA gene sequenced to profile the oyster microbiome. As the reverse reads had overall higher quality scores than the forward, the 7,046,472 reverse reads were used for analysis. Quality trimming, aligning, and chimera removal eliminated 613,340 reads. The remaining reads were clustered into OTUs, and after removing 7,181 singleton OTUs (37.5%), 11,950 OTUs remained, representing 1,176 genus-level phylotypes. There was an average of 336,373 reads per sample, ranging from 68,080-484,007 (Table 3S.1).

A rarified PD whole tree alpha diversity plot allowed assessment of both within-sample diversity and sampling depth. The rarefaction plot of every sample approached a plateau characteristic of adequate sampling depth (Fig. 4.1). The diversity values were generally higher than the diversity values in previously sequenced GBE oysters [135] indicating the higher number of reads achieved with Illumina sequencing in this study were better at capturing true sample diversity.

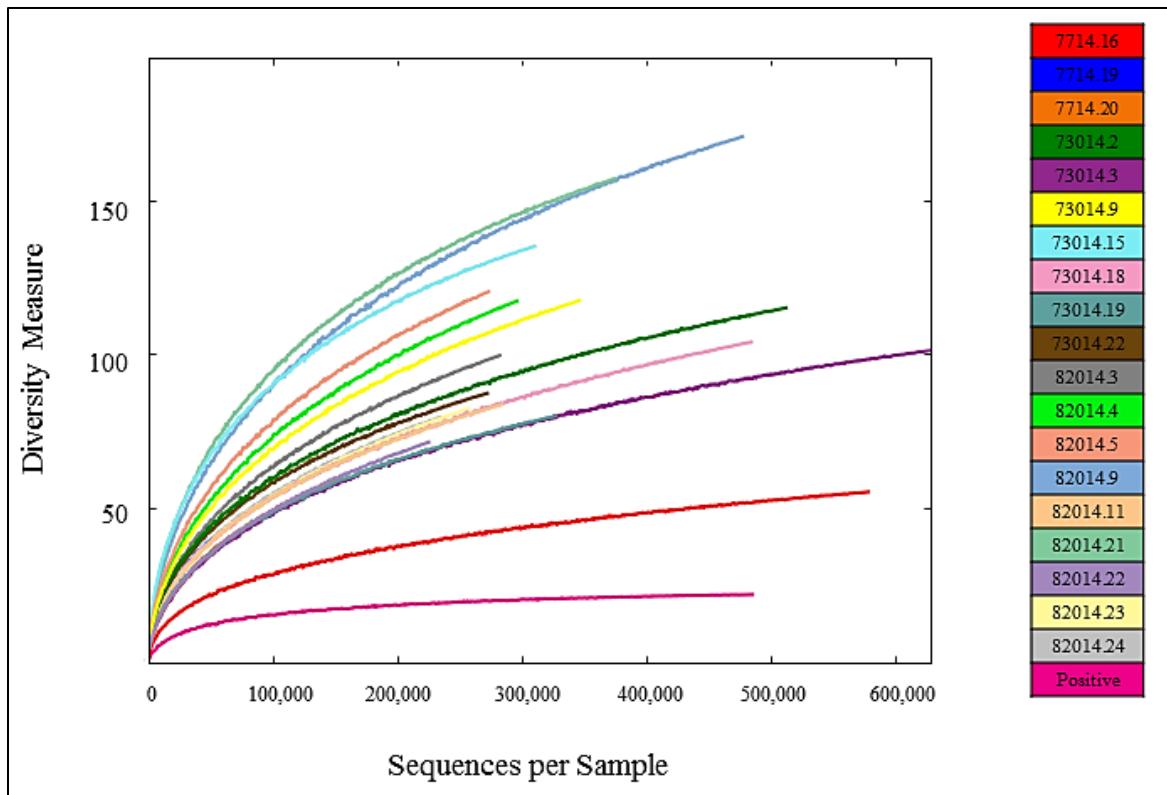


Figure 3.1. Alpha diversity analysis.

Rarified within-sample diversity of individual oyster samples generated with PD whole tree. The sample ID is identified by unique color.

Assessment of accuracy and efficiency of microbiome analysis pipeline

There are no standard procedures for assembling and analyzing 16s microbial reads, and the use of different pipelines on the same dataset produce different results [14, 41, 135, 180], illustrating the need for an internal control. In addition, current methodologies rely on PCR amplification of the 16s rRNA to attach both unique identifying barcodes and sequencing adapters. To sequence all bacterial DNA in a sample, the primers anneal to the highly conserved regions of the 16s gene. DNA from bacteria introduced at any stage prior to amplification, or remaining in the sequencer lane from a previous run with the same barcodes, could contaminate the samples and interfere with accurate identification of the sample microbiota.

To assess the efficiency of the analysis pipeline and test for contamination, we included a positive control containing both Gram-positive and Gram-negative bacterial species of equal proportions: *V. parahaemolyticus*, *Staphylococcus aureus*, *Shigella flexneri*, *Serratia marcescens*, *Pseudomonas aeruginosa*, *Bacillus subtilis*, and *Escherichia coli*. With the exception of the extraction method (see Methods), this control underwent the same amplification and assembly pipeline as the experimental samples. Whereas we were expecting seven phylotypes in similar proportions in the positive control, there were 132 genus-level phylotypes detected, 91.5% of which were assigned to the *Lactococcus* genus. This likely represented incorrectly-identified *Staphylococcus aureus*, as both are in the Streptococcaceae family. There were 7% of reads assigned as *Bacillus*, 0.028% as *Pseudomonas*, 0.005% as *Shigella*, 0.005% as *Vibrio*, and no *Serratia* or *Escherichia*. It is likely OD600 was not an accurate for normalizing concentrations across species. Rather, each species should have been extracted independently and the gDNA quantified then pooled.

Whereas very few microbiome studies include a positive control, the Microbiome Quality Control Project shows similar numbers of OTU inflation in controls. Specifically, the positive controls (bacterial communities of known composition) containing 20 expected OTUs consistently resulted in 50 to 150 OTUs, and the negative controls (no bacteria) resulted in several hundred OTUs [188]. Erroneous OTUs in a negative control could only be from contamination, either from any step prior to PCR or in the sequencer. The erroneous OTUs in the positive control could be either contamination, or PCR or sequencing errors not removed during microbiome assembly [2] or even improper taxonomic classification [91, 104, 180]. One way to eliminate this possible contamination is to use the relative percent abundance of the most abundant contaminant OTU as a threshold, and remove all OTUs occurring at or below that

abundance [212]. This would not have been appropriate to apply to our study, as *Serratia* and *Escherichia* in the positive control were not detected at all so a cutoff threshold may have eliminated part of the true microbiome. Rather, we retained all OTUs but considered any patterns in rare members of the microbiome with the caveat that it may be erroneous data.

Distribution of and relative abundance of phylotypes across oysters

Comparing the distribution of phylotypes (OTUs collapsed by taxonomic similarity) revealed patterns in the microbiome of oysters across sampling dates. Out of 1,176 genus-level phylotypes, 602 (51.2%) were in oysters from July 7, 1061 (90.2%) from July 30, and 1099 (93.5%) from August 20 (Fig. 3.2A). This distribution supported the overall higher alpha diversity in oysters from the later sampling dates (Fig. 3.1). Almost half of all phylotypes were present across all three sampling dates. The oysters from the two later sampling dates shared more phylotypes than either did with the oysters from the earlier sampling date.

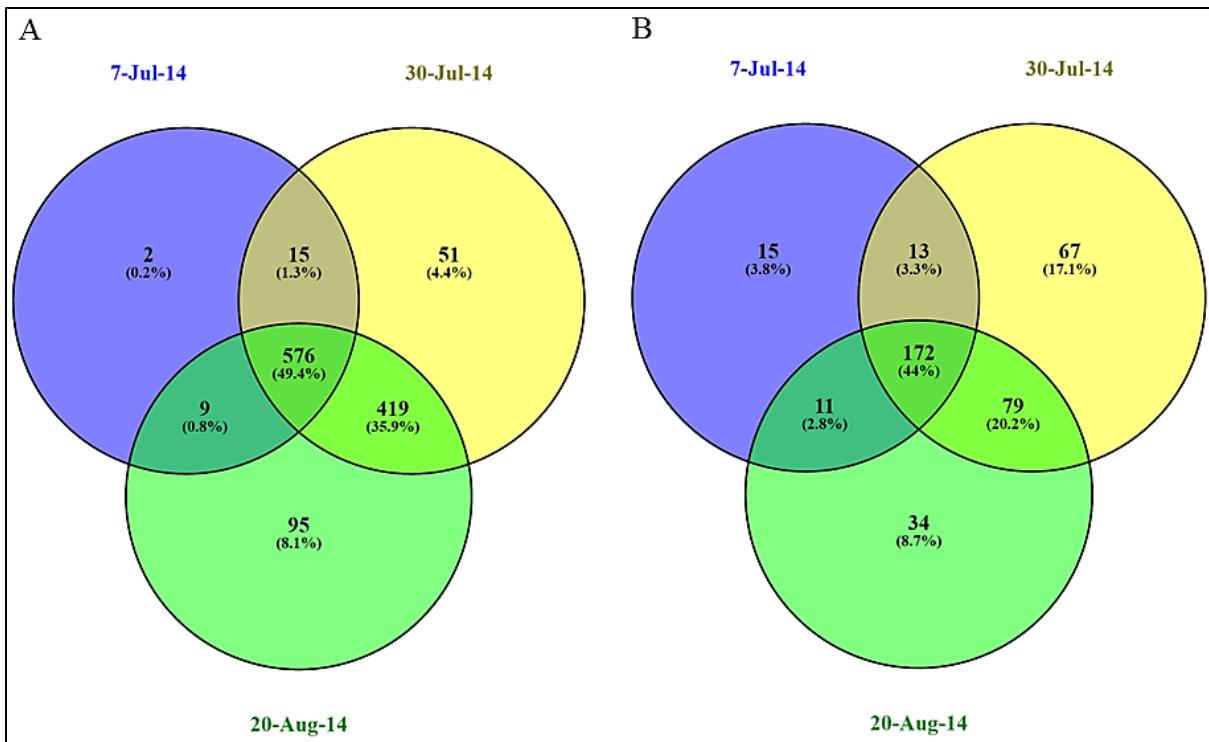


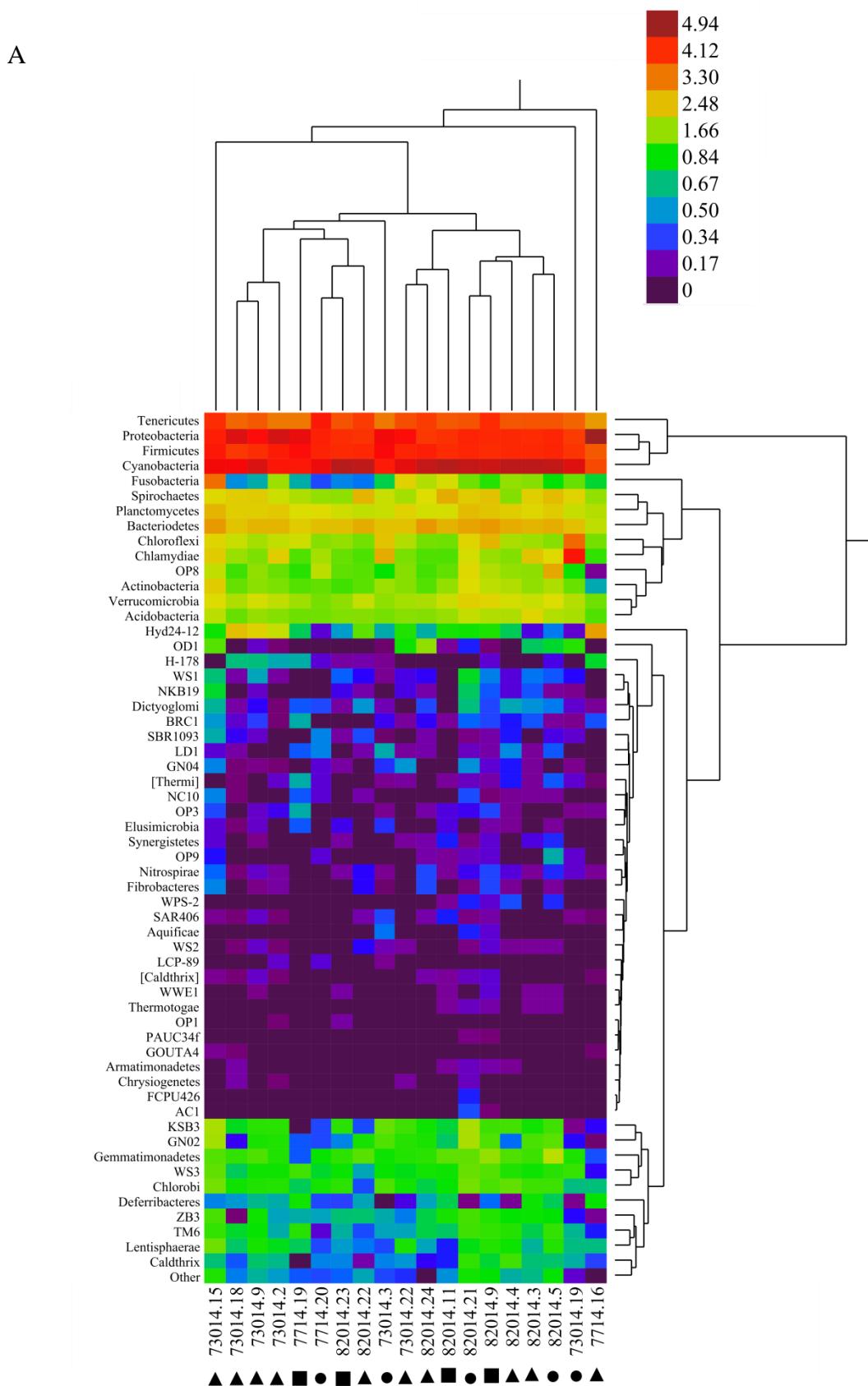
Figure 3.2. Distribution of phylotypes across sampling dates.

Distribution of phylotypes in (A) all oysters from each sampling date (B) every oyster from each sampling date, representing the core microbiome.

When considering only phylotypes present in all oysters from each sampling date, 221 (18.8%) were in all oysters from July 7, 331 (28.1%) from July 30, and 296 (25.2%) from August 20. 44% of phylotypes were shared between every single oyster across the three collection dates (Fig. 3.2B), representing the NI warm season core microbiome. This was a much larger proportion of phylotypes than when comparing across multiple GBE collection sites sequenced with 454 technology [135]. Other studies using 454 sequencing also identify far fewer core OTUs from animals from the same site [84, 194]. The use of Illumina sequencing in this study permitted deeper sequencing and captured more of the core microbiome between samples. Despite the core microbiome, the microbiome of oysters from the later sampling dates

was more similar to each other than the earlier date. While these oysters were collected with the purpose of representing warm season, it is apparent that even within a span of a few weeks there was a shift in the NI oyster microbiome community.

We next analyzed the distribution and relative abundance of phyla-level phylotypes in all samples using dual-hierarchical clustering analysis. When considering the abundance of a given phylum relative to the abundance of all other phyla (unstandardized average-linkage clustering), oysters clustered by collection month (Fig. 3.3A). Specifically, all August oysters contained higher levels of Cyanobacteria than both July collection dates. Most August oysters contained higher levels of Fusobacteria, with the exception of two oysters that grouped closer to the July oysters. Cyanobacteria and Fusobacteria likely shaped this delineation by date. Cyanobacteria represented the most abundant phylum in this study, which was consistent with other oyster microbiome studies [39, 235]. There was no clear delineation between the microbiome of oysters of different *V. parahaemolyticus* abundance levels, implying any differences in the microbiome associated with *V. parahaemolyticus* abundance level was not captured at the phylum level.



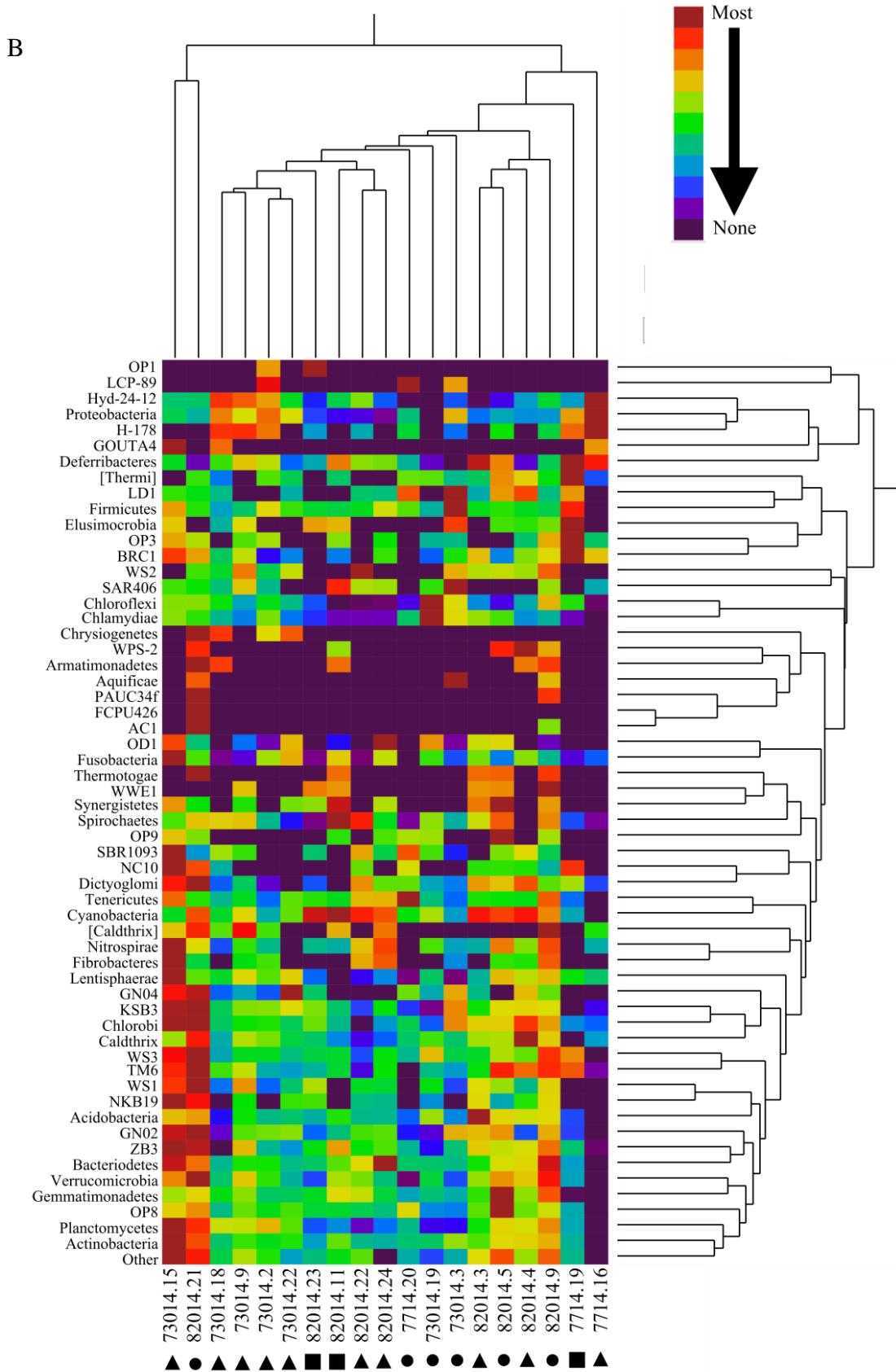


Figure 3.3. Dual hierarchical analysis of phyla-level classification for all samples.

The log-transformed percent abundance of each phylum is indicated by a color scale. Samples and phyla are clustered based on (A) unstandardized and (B) standardized average linkage. In unstandardized linkage, the abundance of each phylum in a given sample is colored based on relative abundance of all phyla, whereas in standardized the abundance of each phylum is colored based on the relative abundance of that phylum across all samples. Oysters are labelled with a shape to represent *V. parahaemolyticus* abundance level: the circle is low, triangle is medium, and square is high.

The standardized average-linkage clustering analysis, which calculates relative abundance of a phylum in a given sample relative to all other samples, does not show the same strict separation by date (Fig. 3.3B), illustrating the variation in any given phylum between individuals collected on the same day outweighs the total relative phylum abundance on the same collection day. In other words, variation between individuals is greater than variation between dates. For example, whereas the relatively higher abundance of Cyanobacteria in the microbiome of August oysters was captured in the standardized clustering, the higher abundance of Fusobacteria was not due to very high proportions of this phyla in oyster #703014.15 Once again, there was no clear delineation between oysters in different *V. parahaemolyticus* classes.

Whereas phyla-level associations reveal general patterns in the microbiome profile, comparing the relative distribution of all phylotypes between samples revealed more specific patterns. The Unifrac distance matrix for all sample pairs was visualized by a PCoA plot (beta diversity) and demonstrated strong clustering for the August oysters and weak clustering between the July oysters, illustrating between-sample similarity (Fig. 3.4A). The relative abundance of all phylotypes elucidated patterns more clearly than merely the presence of phylotypes. When coloring this PCoA plot by *V. parahaemolyticus* class, there is weak clustering between low abundant oysters, but nowhere near as strong as the date relationships

(Fig 3.4B.) The microbiome of oysters is likely highly influenced by day to day fluctuations that had little correlation with *V. parahaemolyticus* abundance.

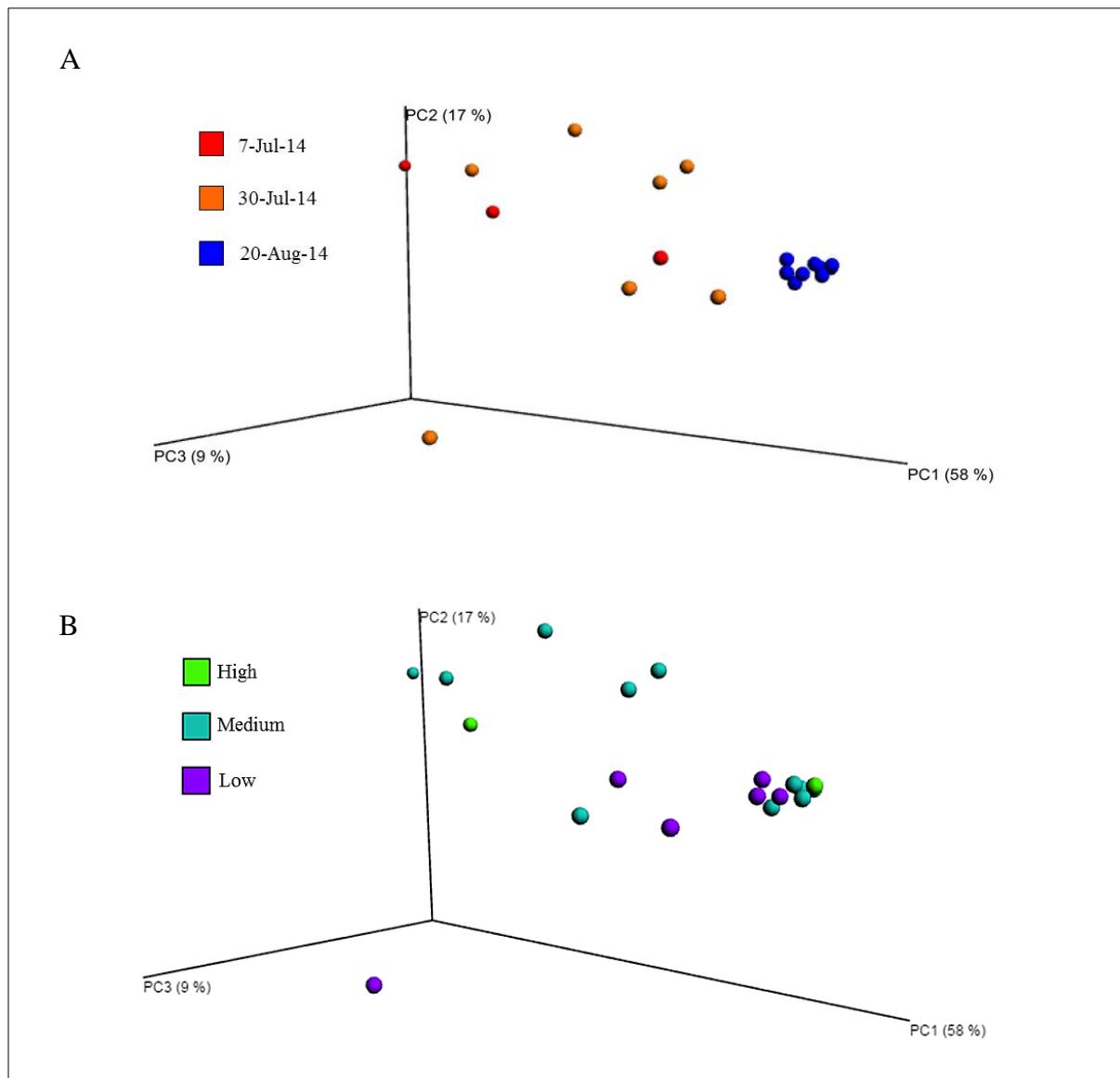


Figure 3.4. Unifrac clustering of all taxonomic levels for all samples.

Sample clustering was based on the Unifrac distance metric for each pair of samples calculated by the total branch length of unique phylotypes over total branch length of all phylotypes, and is visualized with EMPor. Panel (A) is colored by sampling date, and panel (B) is colored by *Vibrio parahaemolyticus* abundance class.

Knowing there are differences between the microbiome of oysters, we employed LEfSe to determine what specific phylotypes differed between both collection date and *V. parahaemolyticus* abundance class. Seventy phylotypes significantly differed by collection date. Four were present at higher levels on July 30, and 66 higher on August 20 (Fig. 3.5). Of these 66, only four were exclusive to the microbiome of oysters collected in August. The remaining were present in the oyster microbiome from July 30, but in significantly lower numbers, and absent from July 7. The distribution of phylotypes across sampling dates suggested the microbiome of oysters from the later sampling dates shared more phylotypes than either did with the earlier collection date, but the PCoA plot suggested the microbiome of July oysters were more similar to each other than the August oysters. There may have been a minor shift in the population of the microbiome from early to late July, and 62 of those phylotypes increased to high enough proportions in August to drive the strong clustering between August oysters.

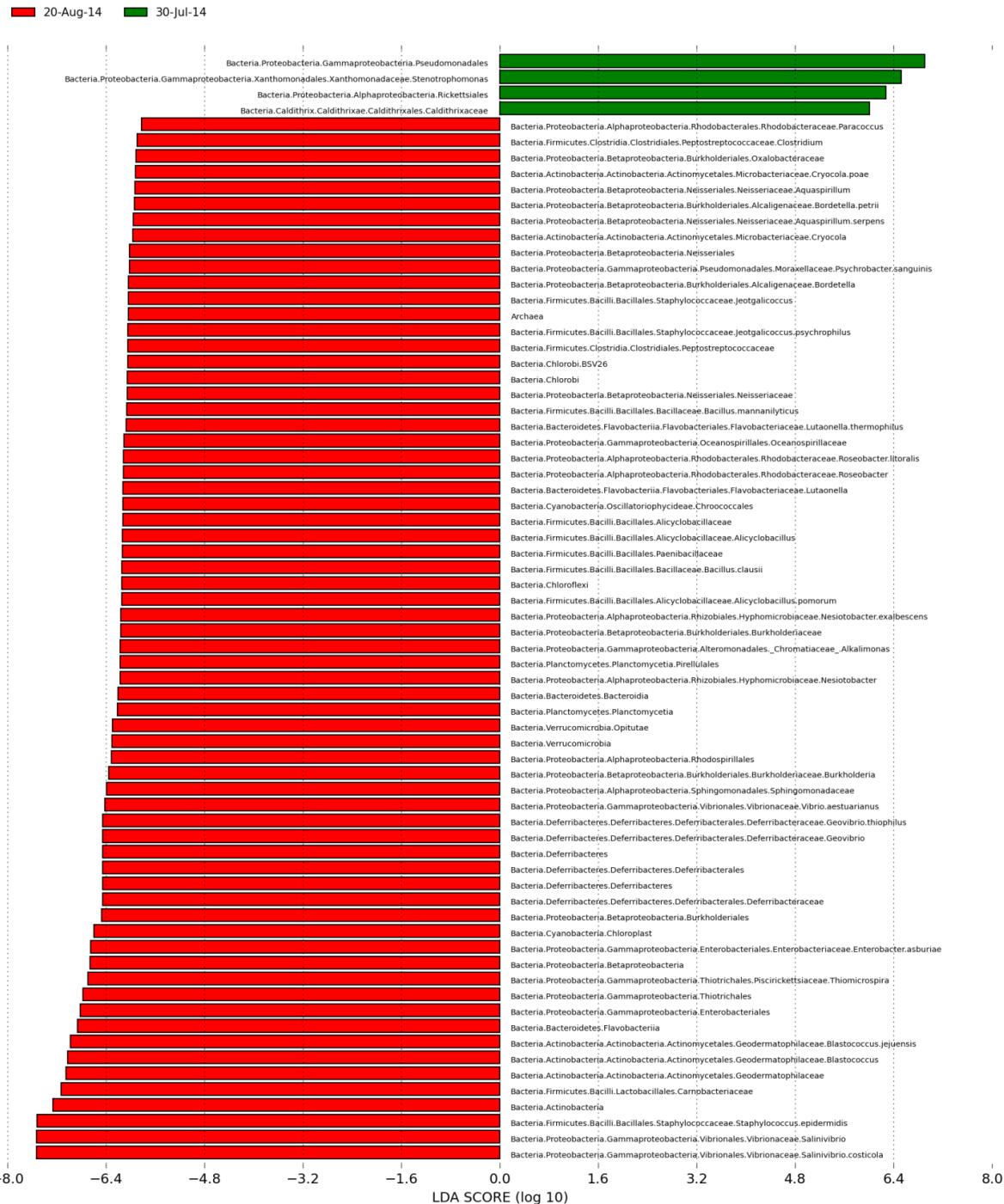


Figure 3.5. LEfSe analysis for oyster samples grouped by date.
Phylotypes in oysters at significantly different proportions from each collection date.

Of the phylotypes which were present at significantly differing proportions between collection dates, three phylotypes inversely correlated with *V. parahaemolyticus* abundance class as a secondary LEfSe grouping: Actinobacteria (Fig 3S.1A), Geodermatophilaceae (Fig 3S.1B), and *Bacillus clausii* (Fig 3S.1C). This inverse correlation is suggestive of a possible antagonistic relationship of these organisms against *V. parahaemolyticus*. Whereas there is no known precedent in literature for an inverse correlation between Actinobacteria and Geodermatophilacece with *V. parahaemolyticus*, *B. clausii* has strong antibacterial activity against *V. parahaemolyticus* [5]. *B. clausii* was a relatively rare component of the oyster microbiome, averaging 0.0013%, and while not significantly different between *V. parahaemolyticus* abundance levels ($p = 0.80$), the average proportions of this organism in the total microbiome inversely correlated with *V. parahaemolyticus* abundance class (Low: 0.0018%, Medium: 0.0012%, High: 0.008%). The inverse correlation and literature precedent supports potential antagonism between *B. clausii* and *V. parahaemolyticus*. Whereas we cannot confidently draw conclusions due to the ambiguity in the positive control coupled with the rarity of this species in the oyster microbiome, future studies can confirm the relationship between *B. clausii* and *V. parahaemolyticus*.

When classifying abundance class as the primary grouping in LEfSe, the levels of four phylotypes significantly differed between high and low *V. parahaemolyticus* abundant oysters (Fig. 3.6). The phylotype that inversely correlated with *V. parahaemolyticus* abundance, which could represent antagonistic bacteria, was *Phaeospirillum fulvum*. These purple photosynthetic bacteria [197] were a very rare component of the microbiome (averaging 0.001%) and because the level ambiguity in the positive control, was too low of a proportion to draw any conclusions about antagonistic potential.

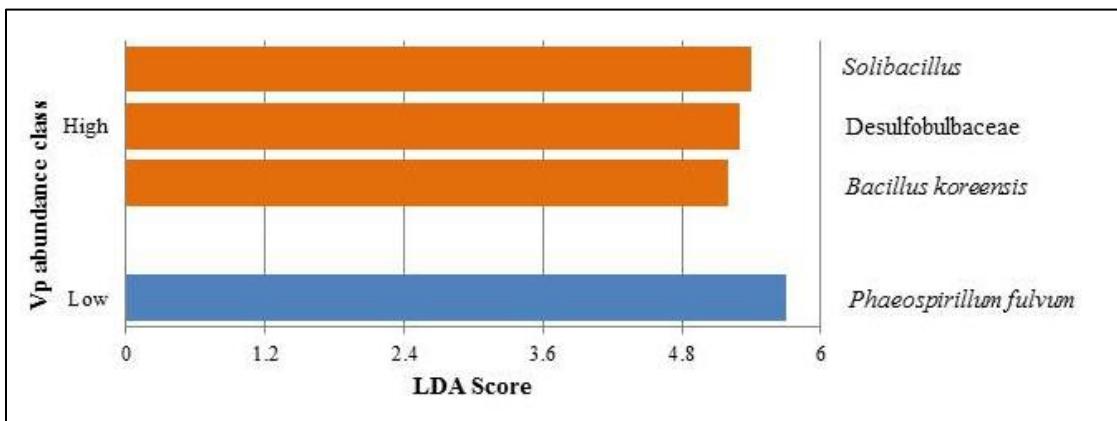


Figure 3.6. LEfSe analysis for oyster samples grouped by *Vibrio parahaemolyticus* abundance.

Phylotypes in oysters at significantly different proportions from each *V. parahaemolyticus* abundance class.

The phylotypes that positively correlated with *V. parahaemolyticus* abundance, which could act as alternative indicator organisms with further investigations, were Desulfobulbaceae, *Solibacillus*, and *Bacillus koreensis*. Desulfobulbaceae is a filamentous family is well known for producing electric currents along the ocean floor [166]. *V. parahaemolyticus* and Desulfobulbaceae have been detected in the microbiome in other marine organisms [152], but there hasn't been any direct research between this family and *V. parahaemolyticus*. It was only present in the oyster microbiome at proportions averaging 0.005%. Because the level ambiguity in the positive control, this is too low of a proportion to be conclusive about potential use as an indicator organism. Not much is known about *Solibacillus*, but at one point it was classified in the *Bacillus* genus [122]. This genus was also only a very small proportion of the oyster microbiome (averaging 0.003%). *B. koreensis* was an even smaller proportion of the

microbiome (averaging 0.00005%) and was only present in three oysters, so it is possible this was influenced by errors and/or contamination. This positive correlation may appear to contradict the inverse correlation with *V. parahaemolyticus* and *B. clausii*, but there is substantial *Bacillus* intra-species diversity [169] so the relationship between *V. parahaemolyticus* and *Bacillus* is likely species specific.

Bacillus was not a rare component of the microbiome (averaging 9.2%, ranging from 3.7 to 29.1%) and was present in every oyster. It was the second most abundant phylotype in the positive control (7%), so it is unlikely the detection was due to errors or contamination. Several *Bacillus* spp. antagonize *V. parahaemolyticus*, including *B. subtilis* [11, 12, 225, 229], *B. pumilus* [128, 229], *B. mojavensis* [128], and *B. cereus* [229]. *B. cereus* in particular was detected in every oyster, and while not significantly different between *V. parahaemolyticus* abundance level ($p = 0.49$), the average proportions of this organism in the total microbiome inversely correlated with *V. parahaemolyticus* abundance class (Low: 4.40%, Medium: 3.95%, High: 2.81%). Therefore, *B. cereus* may have potential as a mechanism of bio-control for *V. parahaemolyticus* abundance.

CONCLUSIONS

In this study, we quantified *Vibrio parahaemolyticus* abundance and profiled the microbiome of oysters collected from Nannie Island in the New Hampshire Great Bay Estuary (GBE). The microbiome of oysters harvested in mid-August were more similar to each other than the oysters harvested earlier in July, but all oysters shared a considerably larger core microbiome than our previous study of GBE oysters from two different collection sites on the same day [135]. Hence, the oyster core microbiome is likely shaped by site-specific conditions.

The microbiomes of marine animals are highly specific based on individuals' surrounding habitat [201, 235] and diet [84], and because oysters are filter feeders, day to day fluctuations in the overlying water likely influence the bacterial community composing oysters.

Phylogenotypes that inversely or positively correlated with *V. parahaemolyticus* levels represent potential antagonistic or indirect indicator microorganisms, respectively. In particular, there may be an interaction with multiple members of the *Bacillus* genus. Precedent in literature suggests *Bacillus* species antagonize *V. parahaemolyticus*, and whereas both *B. clausii* and *B. cereus* inversely correlated with *V. parahaemolyticus* abundance, *B. koreensis* positively correlated. There may be species-specific interactions between *V. parahaemolyticus* and *Bacillus*. With further investigation into the intricacies of the interactions, including culture-based quantification methods to confer the relationships observed in this study, *Bacillus* species could be used as an alternative indicator or probiotic for reducing *V. parahaemolyticus* abundance.

METHODS

Oyster collection and processing

Twenty four oysters were collected at low tide on July 7, 2014, July 30, 2014, and August 20, 2014, from Nannie Island, a naturally-occurring oyster bed in the Great Bay Estuary in New Hampshire. Oysters were collected with oyster tongs, and immediately stored on ice packs in a cooler until laboratory processing. Each individual oyster was processed and homogenized as described in Marcinkiewicz *et al.* [135]. The length (cm) of each oyster was measured from the umbo to growth edge. A three-tube Most Probable Number (MPN) analysis was performed on individual samples by diluting the oyster homogenate 10-fold into APW tubes following the

FDA Bacteriological Analytical Manual (BAM) [71] to quantify the abundance of *Vibrio parahaemolyticus*. Multiple 1mL samples of oyster homogenate not used for MPN analysis were frozen at -80°C until needed for bacterial DNA extraction (see below).

MPN determination of *V. parahaemolyticus* abundance

After incubating overnight at 37°C, MPN tubes were scored for turbidity. Only oysters with turbidity representing high or low *V. parahaemolyticus* abundance were used in subsequent steps (i.e., either very few or many tubes were turbid). To identify the presence of *V. parahaemolyticus* in these turbid tubes, the DNA was extracted from 1.0mL culture from the first 1cm of each turbid tube via crude lysate. Specifically, the culture was pelleted by centrifugation, resuspended in 100µL nuclease-free water, heated to 100°C for 10 minutes, and centrifuged once again. The supernatant was collected and used as template for PCR.

MPN tubes were scored as positive for *V. parahaemolyticus* by detecting the thermolabile hemolysin gene (*tlh*) with PCR [161]. Each 10µL reaction contained 1x Accustart II Supermix, 1µL DNA template, 0.2µM each forward and reverse primer, and nuclease-free water to volume. The thermocycler parameters were as follows: 94°C for 3 minutes; 30 cycles of 94°C for 1 minute, 55°C for 1 minute, and 72°C for 1 minute; and 72°C for 5 minutes. Reactions were then visualized on a 0.7% agarose gel and samples with a band at the expected size (450bp) were scored as positive for *V. parahaemolyticus*. Any oysters demonstrating evidence of PCR inhibition (i.e., presence of *tlh* in a more diluted tube and not less diluted tubes) would have been disregarded from the analysis. The *V. parahaemolyticus* MPN/g was calculated for each oyster according the FDA BAM [71] and grouped by high, medium, or low abundance level based on 10-fold differences in MPN/g.

Metagenetic sequencing preparation

Bacterial DNA was isolated from archived oyster homogenate. The homogenates were thawed, and bacterial DNA extracted from 19 oysters using E.Z.N.A. soil DNA Kit (Omega Bio-Tek, Norcross, GA, USA) following standard protocols for both Gram-negative and Gram-positive bacteria, including incubating the preparations at -20°C for one hour. Genomic DNA was quantified on Nanodrop 2000c (Thermo-Scientific, Wilmington, Delaware, USA) to ensure successful extractions with minimal RNA contamination. To judge the effectiveness of the pipeline at removing PCR and sequencing errors, a positive control was generated, containing seven known bacterial species: *V. parahaemolyticus*, *Staphylococcus aureus*, *Shigella flexneri*, *Serratia marcescens*, *Pseudomonas aeruginosa*, *Bacillus subtilis*, and *Escherichia coli*. These species were grown at 37°C for 3 hours and pooled in equal proportions based on OD600. Total bacterial DNA was extracted using CTAB-NaCl precipitation followed by phenol-chloroform extraction [7], as extraction with the E.Z.N.A. Soil DNA Kit proved to be too harsh for pure culture.

The V4 region of the 16s rRNA gene from each sample and the positive control was amplified and Illumina adapters attached with PCR in triplicate using standard 515F (5' – AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGTGTGCCAGCMGCCGCG GTAA – 3') and 806R (5' – CAAGCAGAAGACGGCATACGAGAT[X]AGTCAGTCAGCCGGACTACHVGGGTWTCT AAT – '3) primers, with each reverse primer having a unique 10-12bp barcode indicated by the brackets [31]. Each 15µL reaction contained 1x HotMasterMix (5 Prime, Inc., Gaithersburg, Maryland, USA), 0.6µL template, 0.2µM forward and reverse primer, and nuclease-free water to volume. The cycling parameters were as follows: 94°C for 2 minutes; 30 cycles of 94°C for 45

seconds, 50°C for 1 minute, 72°C for 1.5 minutes; and 72°C for 10 minutes [31]. Amplicons were run on a 1.2% agarose gel to ensure purity and amplicon size (400bp). Triplicates were pooled into a single sample, and remaining PCR reagents removed with UltraClean 96-well PCR clean up kit (MO BIO Laboratories, Inc., Carlsbad, California, USA) following manufacturer's instructions. Samples were quantified in microtiter plates using Quant-iT PicoGreen dsDNA (Invitrogen, Thermo Fisher Scientific, Waltham, Massachusetts, USA) following manufacturer's instructions, in the Infinite M200 Microplate Reader (Tecan, Männedorf, Switzerland). All samples were pooled into a single tube in equal ratios and submitted to the University of New Hampshire Hubbard Center for Genomic Studies (Durham, New Hampshire, USA) for sequencing on Illumina HiSeq 2500 (Illumina, Inc., San Diego, California, USA).

Metagenetic sequencing analysis

The reverse reads were analyzed with mothur 1.35 [181], which produces more taxonomic classifications to the species level than other pipelines [135]. Reads were quality trimmed to reduce PCR and sequencing errors by setting a qaverage of 30 and removing sequences with ambiguous bases and homopolymers 7 or greater. Sequences were aligned to the silva v119 database [170] and outlier sequences removed. After pre-clustering with one difference, chimeric sequences were detected with uchime [57] and removed to reduce PCR errors. A phylip distance matrix was generated and used to cluster sequences into operational taxonomic units (OTUs) using the furthest method at a cutoff of 0.03. greengenes 13_8 [53] was used to taxonomically classify OTUs, and singletons were removed to further reduce PCR and sequencing errors. OTUs were collapsed into phylotypes using QIIME 1.8 [30].

Representative sequences of each OTU were chosen and used to make a distance matrix to generate a phylogenetic tree with clearcut [58] in mothur [181]. This phylogenetic tree was used to calculate rarified alpha diversity using PD whole tree in QIIME 1.8 [30], as this index was previously determined to be most appropriate for metagenetic studies [135], with ten iterations of 1000 reads added at each rarefaction step to determine both the within-sample diversity and sequencing depth. Weighted and normalized Fast Unifrac [130], which uses all levels of taxonomic assignment to create a distance matrix and groups samples based on similarity, was used to perform beta (between sample) diversity clustering in QIIME [30] and was visualized with EMPeror [214]. The distribution of genus-level phylotypes between sampling dates was determined with Venny 2.0 [158], whereas patterns in abundance in phyla-level classifications in all samples were revealed with a dual-hierarchical clustering performed with JMP 12 (SAS Institute Inc., Cary, North Carolina, USA) for log-transformed percent abundance using both standardized and unstandardized average linkage. LEfSe [185] was used at default settings, pre-normalizing samples to 1M, to determine statistical significance of taxonomic similarities between oysters of the same sampling date and *V. parahaemolyticus* abundance.

ACKNOWLEDGEMENTS

We thank Michael Taylor for assisting with the preparation of the metagenetic sequencing protocols, Tiffany DeGroot for assistance with processing oysters and inoculating MPN tubes, and members of the Jones lab for collecting oysters.

SUPPLEMENTAL INFORMATION

Table 3S.1. Number of reads per sample.

Number of 16s reads per oyster after all quality filtering and clustering into OTUs using mothur.

Date	Oyster	Number of Reads
7-Jul-14	2	155776
	16	578453
	19	68088
30-Jul-14	2	512043
	3	627811
	9	346162
	15	310924
	18	484007
	19	327775
	22	272191
	3	282449
20-Aug-14	4	296413
	5	273588
	9	477717
	11	284927
	21	375021
	22	225795
	23	257029
	24	234929

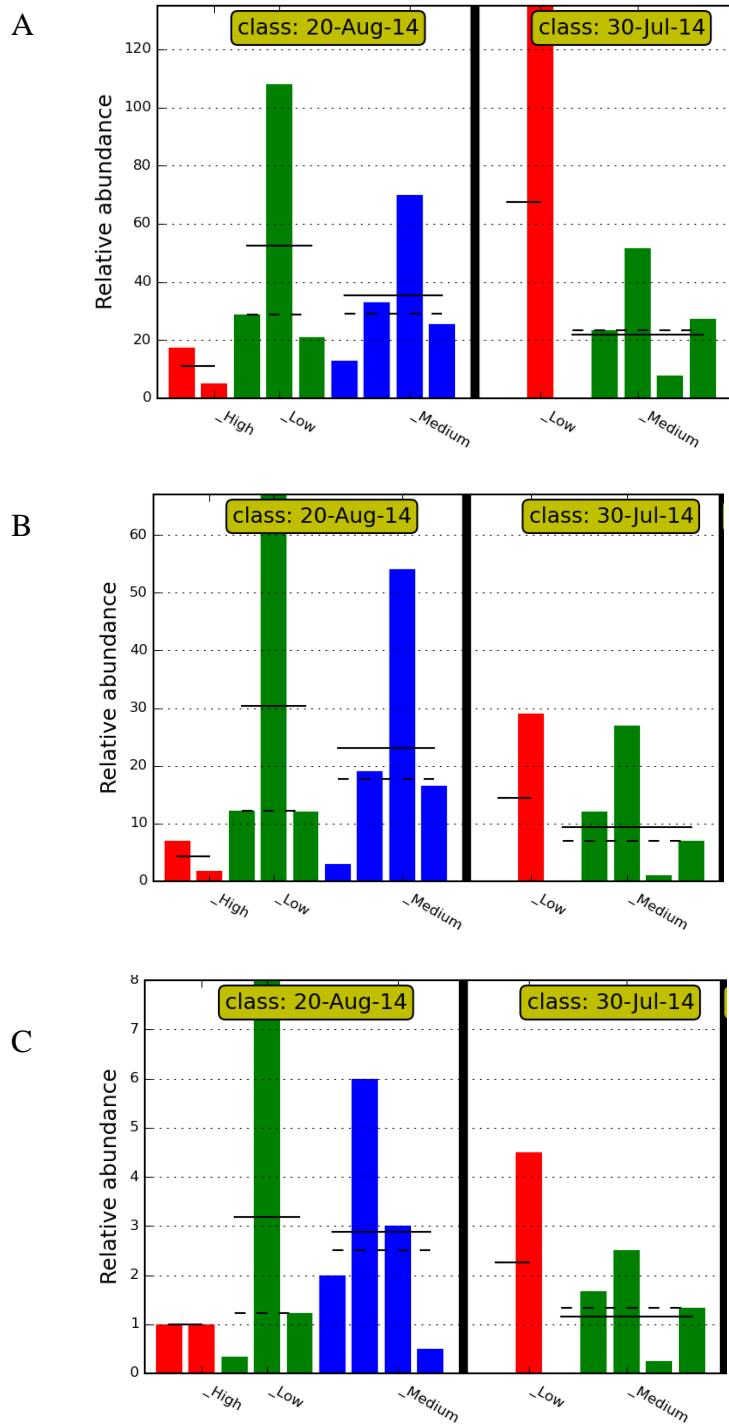


Figure 3S.1. LEfSe analysis for oyster samples grouped by date and *V. parahaemolyticus* abundance.

Phylotypes in oysters at significantly different proportions from each collection date that inversely correlated with *V. parahaemolyticus* abundance class, (A) Actinobacteria, (B) Geodermatophilaceae and (C) *Bacillus clausii*. None of these phylotypes were present in the microbiome of oysters collected on July 7. Relative abundance was calculated through LEfSe.

CHAPTER 4

Using Phylogenetic Relationships and Phage Content to the Elucidate Invasion of *Vibrio parahaemolyticus* Sequence Type 36 into the United States North Atlantic Coast

Ashley L. Marcinkiewicz^{1,2}, Feng Xu^{1,2}, Jeffery A. Hall^{2,3}, Kevin Drees^{1,2}, Vaughn S. Cooper⁴, and Cheryl A. Whistler^{1,2}

¹ Northeast Center for Vibrio Disease and Ecology, University of New Hampshire, Durham, NH, USA

² Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH, USA

³ Hubbard Center for Genome Studies, University of New Hampshire, Durham, NH, USA

⁴ Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

ABSTRACT

Vibrio parahaemolyticus sequence type (ST) 36 strains caused 104 infections in a multi-state outbreak in 2013. These illnesses were unprecedented, as the ST36 lineage is native to the Pacific Northwest yet caused infections on the Atlantic coast. To investigate the origin and evolution of this lineage on the Atlantic coast, we compared the genomes of publicly available ST36 strains to our collection of New England ST36 clinical isolates. The biggest difference in strains from the Pacific and Massachusetts was bacteriophage. We identified three unique phage that share similar architecture to the f237 phage: Vipa10290, Vipa26, and Vipa36. These phage were likely horizontally acquired after the evolution of the ST36 lineage, unlike other pathogenic lineages which appear to have acquired phage in their evolutionary past. Vipa10290, Vipa36, and Vipa26 correlated with the phylogeography of ST36 subpopulations. Specifically, ST36 strains from the Gulf of Maine and Long Island Sound with phage all have Vipa26 and Vipa36, respectively, whereas Vipa10290 is characteristic of an older Pacific lineage. These phage have promise as a biomarker with our PCR primers to assist with identifying the geographic region of strain origin. In addition, strain phylogeny suggests there were multiple ST36 invasions into the Atlantic coast.

KEYWORDS

Vibrio parahaemolyticus, sequence type 36, bacteriophage, New England

INTRODUCTION

While illnesses from *Vibrio parahaemolyticus* were once rare and sporadic in the United States [15, 48], this human pathogenic bacterium now causes an estimated 35,000 illnesses each year [178]. Historically, infections occurred in warmer areas such as the Gulf of Mexico, but this has changed in recent years. Specifically, illnesses traced to the Gulf coast states (Texas, Louisiana, Mississippi, Alabama, and Florida) made up 48% of all reported infections in 1997-1998, while the United States Northeast Atlantic (NEA) states (New England and New York) were only 3% of total infections. But in 2013, the NEA states made up 23.0% of reported vibriosis infections [36].

Recent vibriosis in NEA states is caused by the emergence of both resident and non-resident pathogenic lineages, including sequence type (ST) 631, 34, 674, and 36 [231]. ST36 strains (serotype O4:K12), which are native to the United States Pacific Northwest (PNW), have recently been traced to outbreaks in California and Spain, as well as the North- to mid-Atlantic coast [137, 141, 149]. Whole genome comparisons reveal ST36 strains both in the Northeast [224, 231] and Maryland [88] derived from the PNW populations.

This lineage was first detected in the Northeast in the Long Island Sound (LIS) in 2012 [149] and subsequently spread throughout New England [231]. However, an LIS ST36 strain shares more genomic content with Maryland strains than with PNW and Gulf of Maine ST36 strains [231], implying the ancestry of these New England populations may not be shared. This study performed whole genome comparisons of all available PNW and NEA ST36 strains to gain further insight into the origin and evolution of the ST36 population in New England. We found phage that was uniquely associated with geographically-distinct ST36 subpopulations, which can be used for more accurate strain traceback and inferred strain ancestry. In addition, the

phylogeographically distinct subpopulations implied multiple invasions of ST36 strains into the Atlantic coast.

RESULTS AND DISCUSSION

Comparisons between *Vibrio parahaemolyticus* Sequence Type 36 strains from different regions reveal unique features of populations

To elucidate the similarity of *Vibrio parahaemolyticus* sequence type (ST) 36 strains from different geographic regions, *breseq* was employed to perform whole genome comparisons between strain 10290 from the United States Pacific Northwest (PNW), and strains MAVP26 and MAVP36 from Massachusetts. There were very few differences between the strains, with the exception of a region annotated as bacteriophage. Strains 10290, MAVP36, and MAVP26 each harbored a unique phage, which we named Vipa10290, Vipa36, and Vipa26, respectively. These three phage have similar structure to the filamentous Inovirus f237 (Fig. 4.1), the phage uniquely associated with the ST3 pandemic *V. parahaemolyticus* lineage [147]. In addition to shared gene architecture, all four are integrated into the *dif* site of the harboring genome [100].

Besides the absence of ORF8 in the ST36-associated phage, the main differences between these four phage were hypothetical proteins (Fig. 4.1). Hypothetical protein A is unique to the two Massachusetts phage, and Vipa10290, Vipa36, and Vipa26 contain unique hypothetical proteins after seven shared ORFs. Despite these differences, all four phage were flanked by ORFs 10 and 9, and contained ORFs 1-7, which were almost identical in sequence. These seven core genes likely underwent purifying selection (codon-based Z test; dS – dN =9.425, p < 0.001) and may be essential to phage function. Two of the seven core f237-like ORFs were the accessory cholera toxin and the zona occludens toxin. Both are secondary toxins

in CTXΦ, the filamentous bacteriophage encoding cholera toxin, which also integrates into the *dif* site of *Vibrio cholerae* [218]. There have been eight *V. cholerae* pandemics, of which the 6-8th were caused by unique lineages that independently acquired virulence-associated traits from horizontal gene transfer, including CTXΦ [67]. It is likely the ST36 lineage also horizontally acquired these phage. Vipa26 and Vipa36, which were harbored in Massachusetts ST36 strains, may have been acquired from New England waters from local *V. parahaemolyticus* strain populations.

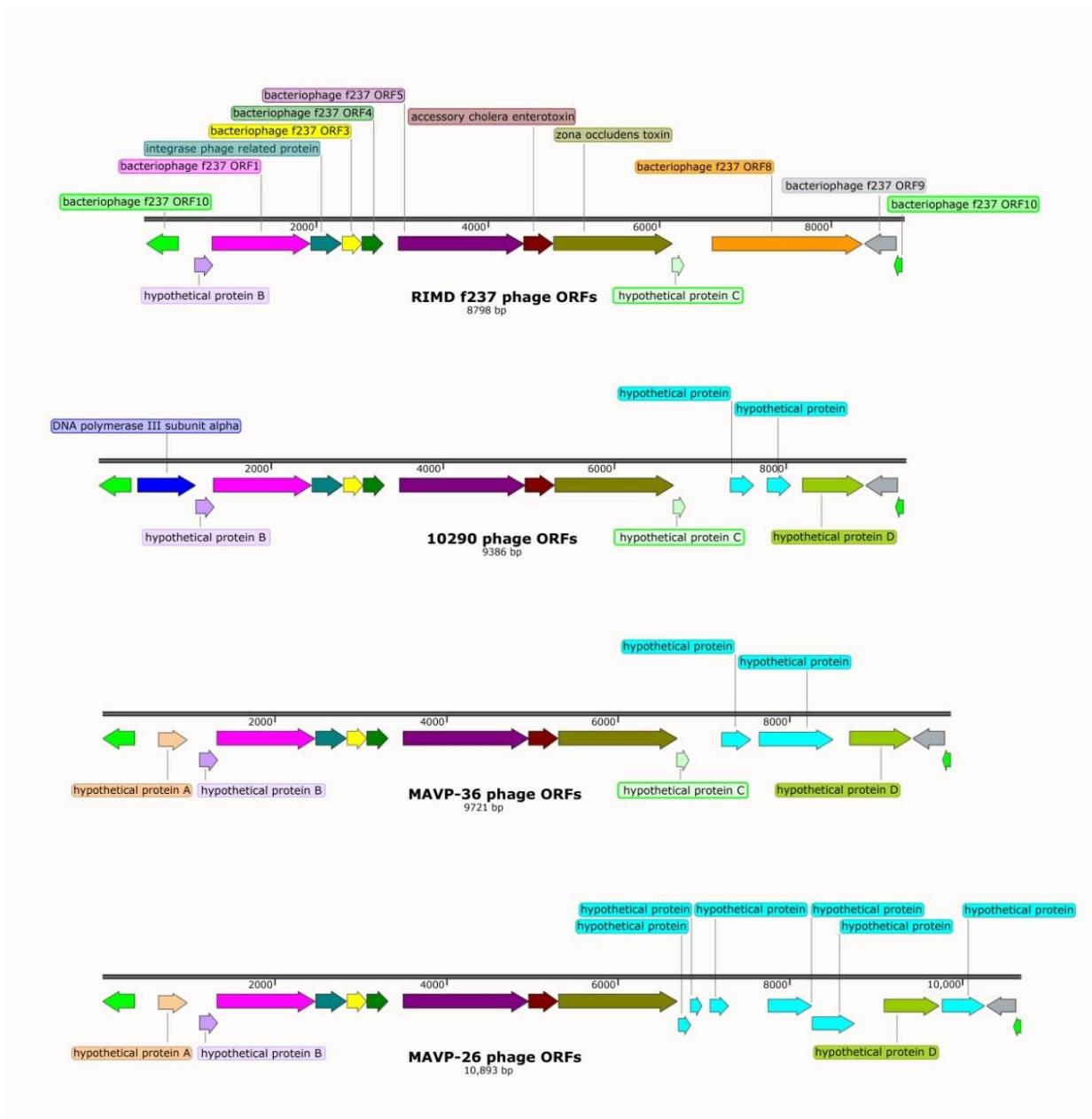


Figure 4.1. Gene map of f237 and ST36-associated phage.

Gene content of f237, Vipa10290 (from 10290), Vipa36 (from MAVP36), and Vipa26 (from MAVP26). Gene regions of the same color are the same genes, with the exception of hypothetical proteins, which are unique unless specified with a letter designation.

Screening environmental isolates for Vipa26 and Vipa36 reveals diverse phage population

To determine if ST36 strains horizontally acquired Vipa26 and Vipa36 from New England (NE), we designed PCR primers to detect f237-like phage, NE-like phage, and Vipa26 and Vipa36, by targeting both conserved and unique hypothetical genes producing amplicons of different sizes (Fig. 4.2). Screening environmental isolates from New Hampshire, Connecticut, and Massachusetts revealed what appeared to be a highly diverse population of phage. Several environmental strains tested positive for the primer sets as expected for Vipa26 (f237-like, NE-like, and Vipa26) and Vipa36 (f237-like, NE-like, and Vipa36) (Table 4.1). But there were also strains with ambiguous banding patterns that, while they were the expected size, were an unexpected combination of bands (Table 4.1, Table 4S.1).

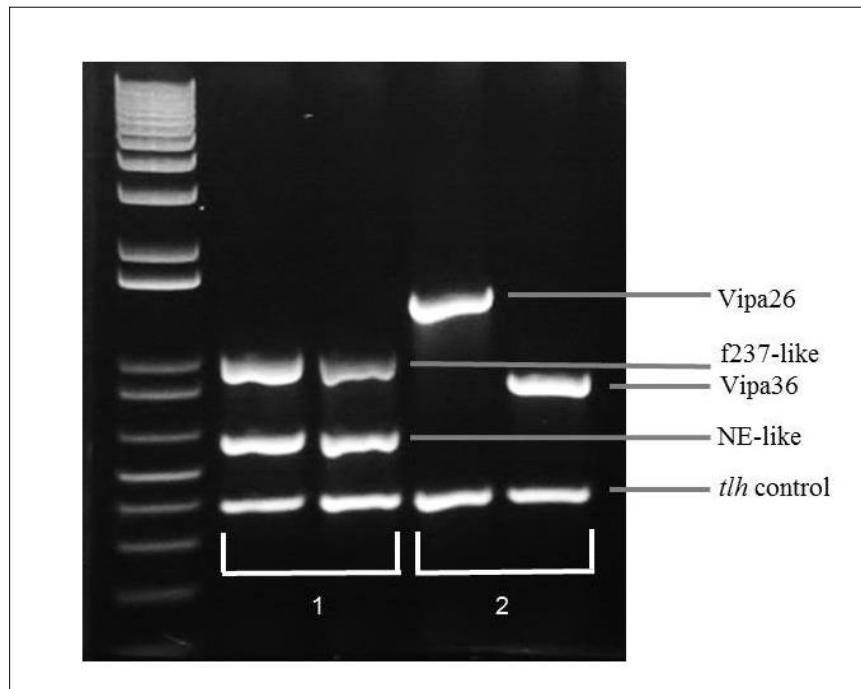


Figure 4.2. Visualization of amplification by phage-detection PCR primers.

The PCR to detect f237-like phage, NE-like phage, and Vipa26 and Vipa36 was developed as a two-step multiplex using *tlh* as an internal control. Reactions were visualized on a 0.7% agarose gel at 60V, using 1Kb+ ladder for size reference.

Table 4.1. Summary of screening environmental strains for phage.

Total numbers of Vipa26, Vipa36, and other phage harbored in environmental isolates from the Long Island Sound and Gulf of Maine. Ambiguous refers to results not matching Vipa26 or Vipa36.

Phage	LIS		GOM	
	CT	MA	NH	MA
Vipa26	1	1	3	7
Vipa36	1	0	2	7
Ambiguous	5	12	22	39
Total screened	50	27	136	153

The phage population was more diverse than first anticipated, and there appeared to be other f237-like and NE-like phage in environmental strains that were neither Vipa26 nor Vipa36. We sequenced the genomes of several of these phage-harboring environmental strains to confirm the accuracy of the primer sets. Whole genome sequencing confirmed there were f237-like phage other than Vipa26 and Vipa36 present in New England *V. parahaemolyticus* strains. Of the strains that were sequenced, none of the environmental strains identified as harboring Vipa36 from PCR actually harbored this phage. Only two sequenced environmental strains carried Vipa26, G1445 and G1449, but G1449 did not show the expected PCR bands. Further investigations into the gene amplified by the NE-like primers, hypothetical protein A, revealed this gene was present in other *V. parahaemolyticus* strains from other regions, both with and without f237-like phage content, and therefore was not effective in detecting phage unique to NE. The ambiguity and inconsistencies indicated the primers in this study were not able to uniquely detect Vipa26 and Vipa36 from environmental strains, and they need to be modified for future use.

Relationship of all *V. parahaemolyticus* strains with distribution of f237-like phage

Whereas the *V. parahaemolyticus* ST3 lineage is historically characterized by f237 [147], we identified three unique ST36 strains that harbored three unique f237-like phage. This suggested f237-like phage were basally acquired for some lineages but were a more recent acquisition for others. To analyze strain relatedness and the presence of f237-like phage, publicly available and in-house *V. parahaemolyticus* genomes were sequence-typed and included in a maximum-likelihood phylogenetic tree. The sequence types and phage content of these strains are in Table 4.2. The phylogenetic tree, which included 24.2% of the reference genomes, revealed clonal clades of the same known ST, including STs 3 and 36, as well as several previously undefined STs (Fig. 4.3). Whereas the close grouping of three sets of sequence types (STs 3 and 36; 111, 653, and 114; and 8 and ST-NF) indicated shared ancestry, other pathogenic lineages (e.g., 34, 43, 417, 647, and 636) were elsewhere in the tree. The lack of a single lineage for all pathogens indicates pathogenicity evolved multiple times, similar to *V. cholerae* [67]. A maximum-likelihood phylogenetic tree of just the phage-harboring strains, which used 58.1% of the reference genomes, had a very similar topography (Fig. 4S.1) to the whole genome tree with all *V. parahaemolyticus* strains, implying ST was a good representation of overall genome similarity.

Table 4.2. All strains used in this study.

List of all strains used in this study, including NCBI assembly ID (or none if sequenced in-house), year collected, geographic region and country, clinical (C) or environmentally isolated (E), sequence type (ST) and phage type, if present. N/A for any category indicates the information was not available. Phage types with an * indicate phage content was present over multiple scaffolds with complete coverage, or nearly complete coverage over one scaffold. NF for sequence type indicates there is not yet a defined sequence type matching that strain.

Strain	Assembly ID	Year	Location	Country	C/E	ST	Phage
029-1b	GCA_000707045.1	1997	Oregon	USA	E	36	Vipa10290*
04-1290	GCA_000878815.1	2004	Alberta	Canada	C	36	None
04-2549	GCA_000951795.1	2004	Saskatchewan	Canada	C	3	f237
04-2551	GCA_000975195.1	2004	Ontario	Canada	C	3	f237
07-1339	GCA_000972045.1	2007	British Columbia	Canada	C	3	f237
07-2965	GCA_000960565.1	2007	Saskatchewan	Canada	C	326	None
08-0278	GCA_000960645.1	2008	Alberta	Canada	C	216	None
08-7626	GCA_000960665.1	2008	Alberta	Canada	C	417	None
09-3216	GCA_000878785.1	2009	British Columbia	Canada	C	36	None
09-3217	GCA_000960655.1	2009	British Columbia	Canada	C	43	Other
09-3218	GCA_000974905.1	2009	British Columbia	Canada	C	417	None
09-4434	GCA_000972125.1	2009	Alberta	Canada	C	417	None
09-4435	GCA_000960685.1	2009	British Columbia	Canada	C	3	f237
09-4660	GCA_000972055.1	2009	British Columbia	Canada	C	417	None
09-4661	GCA_001559885.1	2009	British Columbia	Canada	C	417	None
09-4663	GCA_001006115.1	2009	British Columbia	Canada	C	417	None
09-4664	GCA_000972035.1	2009	British Columbia	Canada	C	417	None
09-4681	GCA_000972025.1	2009	New Brunswick	Canada	C	632	None
090-96	GCA_000701045.1	1996	N/A	Peru	C	265	None
10-4241	GCA_000878805.1	2006	British Columbia	Canada	C	36	None
10-4242	GCA_000878755.1	2006	British Columbia	Canada	C	36	None
10-4243	GCA_000972105.1	2006	British Columbia	Canada	C	141	None
10-4244	GCA_001006125.1	2006	British Columbia	Canada	C	141	None
10-4245	GCA_000878725.1	2006	British Columbia	Canada	C	36	None
10-4246	GCA_000878705.1	2006	British Columbia	Canada	C	36	None
10-4247	GCA_000878665.1	2006	British Columbia	Canada	C	36	None

10-4248	GCA_000878675.1	2006	British Columbia	Canada	C	36	None
10-4251	GCA_001006105.1	2006	British Columbia	Canada	C	3	f237
10-4255	GCA_001006195.1	2006	British Columbia	Canada	C	43	Other
10-4274	GCA_000878735.1	2005	British Columbia	Canada	C	36	None
10-4287	GCA_001006185.1	2003	British Columbia	Canada	C	50	None
10-4288	GCA_000878645.1	2003	British Columbia	Canada	C	36	None
10-4293	GCA_000878595.1	2002	British Columbia	Canada	C	36	Vipa10290
10-4298	GCA_000878565.1	2001	British Columbia	Canada	C	36	None
10-4303	GCA_000878575.1	2000	British Columbia	Canada	C	36	Vipa10290
10-7197	GCA_000878585.1	2008	British Columbia	Canada	C	36	None
10-7205	GCA_001006205.1	2008	British Columbia	Canada	C	417	None
10290	GCA_000454205.1	1997	Washington	USA	C	36	Vipa10290
10292	GCA_000707245.1	1997	Washington	USA	C	50	None
10296	GCA_000500105.1	1997	Washington	USA	C	36	Vipa10290
10329	GCA_001188185.1	1998	Washington	USA	C	36	Vipa10290
12310	GCA_000500755.1	2006	Washington	USA	C	36	None
12315	GCA_000877535.1	2006	Washington	USA	C	36	None
13-028-A3	GCA_000737635.1	2013	N/A	Vietnam	E	1166	None
22702	GCA_000958645.1	1998	Georgia	USA	E	NF	Other
3256	GCA_000519405.1	2007	N/A	USA	C	36	None
3259	GCA_000454245.1	N/A	N/A	N/A	N/A	479	None
3324	GCA_000877495.1	2007	Washington	USA	C	36	None
3355	GCA_000877615.1	2007	N/A	USA	C	65	None
3631	GCA_000877595.1	2007	Washington	USA	C	417	None
3644	GCA_000877755.1	2007	Washington	USA	C	43	Other*
3646	GCA_000877765.1	2007	Washington	USA	C	417	None
48057	GCA_000706825.1	1990	Washington	USA	C	36	Vipa10290
48291	GCA_000707525.	1990	Washington	USA	C	36	Vipa10290
49	GCA_000877625.1	2007	Washington	USA	E	137	Other
50	GCA_000519385.1	2006	N/A	USA	C	34	None
605	GCA_000519365.1	2006	Washington	USA	E	3	f237
846	GCA_000877405.1	2007	Washington	USA	E	36	None
861	GCA_000524535.1	2007	Washington	USA	E	3	f237
863	GCA_000877485.1	2007	Washington	USA	E	3	f237
901128	GCA_000877675.1	1997	N/A	USA	C	135	None
930	GCA_000877475.1	2007	Washington	USA	E	3	f237

949	GCA_000454455.1	2006	Washington	USA	C	3	f237
97-10290	GCA_000877425.1	1997	Washington	USA	C	36	Vipa10290
98-513-F52	GCA_000707605.1	1998	Louisiana	USA	E	34	None
A1EZ919	GCA_001559895.1	2001	British Columbia	Canada	C	N/A	Vipa10290
A4EZ700	GCA_001559805.1	2004	British Columbia	Canada	C	N/A	Other*
AN-5034	GCA_000182385.1	1998	N/A	Bangladesh	C	3	f237
AQ3810	GCA_000154045.1	1983	N/A	Japan	C	87	None
AQ4037	GCA_000182365.1	1985	N/A	Japan	C	96	None
ATC210	GCA_001270885.1	1998	N/A	Chile	C	3	f237
ATC220	GCA_001270975.1	1998	N/A	Chile	C	3	f237
ATCC_17802	GCA_001011015.1	1951	N/A	Japan	C	1	None
B-265	GCA_000516875.1	2004	N/A	Mozambique	C	3	f237
BB22OP	GCA_000328405.1	1982	N/A	Bangladesh	E	88	None
CTVP4C	In-house	2012	Connecticut	USA	C	36	None
CTVP5C	In-house	2012	Connecticut; New York	USA	C	36	None
CTVP6C	In-house	2012	New York	USA	C	36	None
CTVP13C	In-house	2012	N/A	N/A	C	36	Vipa36
CTVP19C	In-house	2013	Massachusetts	USA	C	34	None
CTVP20C	In-house	2013	N/A	N/A	C	36	Vipa26
CTVP20E	In-house	2013	Connecticut	USA	E	28	None
CTVP23C	In-house	2013	Massachusetts; Virginia	USA	C	36	None
CT24E	In-house	2013	Connecticut	USA	E	1136	Other
CTVP25C	In-house	2013	Connecticut	USA	C	36	Vipa26
CTVP29C	In-house	2013	Connecticut	USA	C	NF	None
CTVP30C	In-house	2013	Connecticut	USA	C	36	None
CTVP35C	In-house	2013	Washington; British Columbia; New Brunswick	USA, Canada	C	194	None
CTVP41C	In-house	2013	Massachusetts; Prince Edward Island	USA, Canada	C	36	Vipa36
CTVP44C	In-house	2013	Connecticut	USA	C	36	Vipa10290
CT4264	In-house	2013	Connecticut	USA	E	NF	Other
CT4287	In-house	2013	Connecticut	USA	E	674	Other
CT4291	In-house	2013	Connecticut	USA	E	674	Other
EKP-008	GCA_000510585.1	2007	N/A	Bangladesh	E	479	None
EKP-021	GCA_000571915.1	2008	N/A	Bangladesh	E	3	f237
EKP-026	GCA_000525005.1	2008	N/A	Bangladesh	E	3	f237
EKP028	GCA_000522005.1	2008	N/A	Bangladesh	E	3	f237
EN2910	GCA_000877685.1	2000	Washington	USA	C	36	Vipa10290

EN9701072	GCA_000877715.1	1997	Washington	USA	C	43	Other*
EN9701121	GCA_000877725.1	1997	Washington	USA	C	50	None
EN9701173	GCA_000877555.1	1997	Washington	USA	C	36	Vipa10290
EN9901310	GCA_000877565.1	1999	Washington	USA	C	36	Vipa10290
F11-3A	GCA_000707545.1	1988	Washington	USA	C	36	Vipa10290
FDA_R31	GCA_000430405.1	2007	Louisiana	USA	N/A	23	Other
FORC_004	GCA_001433415.1	N/A	N/A	N/A	N/A	N/A	None
FORC_006	GCA_001304775.1	2014	Gyeongnam	South Korea	N/A	N/A	Other
FORC_008	GCA_001244315.1	N/A	N/A	N/A	N/A	N/A	None
G8	In-house	2007	New Hampshire	USA	E	NF	None
G61	In-house	2007	New Hampshire	USA	E	1125	None
G79	In-house	2007	New Hampshire	USA	E	NF	None
G95	In-house	2007	New Hampshire	USA	E	NF	None
G145	In-house	2007	New Hampshire	USA	E	NF	None
G149	In-house	2007	New Hampshire	USA	E	631	None
G151	In-house	2007	New Hampshire	USA	E	83	None
G227	In-house	2007	New Hampshire	USA	E	1087	None
G320	In-house	2008	New Hampshire	USA	E	NF	Other*
G360	In-house	2008	New Hampshire	USA	E	NF	None
G363	In-house	2008	New Hampshire	USA	E	NF	Other*
G441	In-house	2008	New Hampshire	USA	E	NF	None
G445	In-house	2008	New Hampshire	USA	E	NF	None
G524	In-house	2008	New Hampshire	USA	E	NF	None
G640	In-house	2008	New Hampshire	USA	E	NF	None
G650	In-house	2008	New Hampshire	USA	E	NF	None
G653	In-house	2008	New Hampshire	USA	E	NF	None
G729	In-house	2008	New Hampshire	USA	E	380	Other*
G735	In-house	2008	New Hampshire	USA	E	NF	None
G747	In-house	2008	New Hampshire	USA	E	NF	None
G755	In-house	2008	New Hampshire	USA	E	NF	None
G756	In-house	2008	New Hampshire	USA	E	NF	None
G760	In-house	2008	New Hampshire	USA	E	NF	None
G1286	In-house	2009	New Hampshire	USA	E	107	Other*
G1334	In-house	2009	New Hampshire	USA	E	NF	None
G1350	In-house	2009	New Hampshire	USA	E	NF	None
G1355	In-house	2009	New Hampshire	USA	E	NF	None

G1386	In-house	2009	New Hampshire	USA	E	1356	Other
G1393	In-house	2009	New Hampshire	USA	E	NF	None
G1445	In-house	2009	New Hampshire	USA	E	NF	Vipa26
G1449	In-house	2009	New Hampshire	USA	E	NF	Vipa26
G1463	In-house	2009	New Hampshire	USA	E	NF	None
G1487	In-house	2009	New Hampshire	USA	E	NF	None
G3578	In-house	2013	New Hampshire	USA	E	674	Other
G3599	In-house	2013	New Hampshire	USA	E	674	Other
G3654	In-house	2013	New Hampshire	USA	E	1123	None
G3673	In-house	2013	New Hampshire	USA	E	34	None
G4026	In-house	2013	New Hampshire	USA	E	773	None
G4186	In-house	2013	New Hampshire	USA	E	34	None
G6928	In-house	2015	New Hampshire	USA	E	631	None
G6494	In-house	2015	New Hampshire	USA	E	NF	Other
G6499	In-house	2015	New Hampshire	USA	E	NF	Other
Gxw_7004	GCA_001541615.1	2007	Guangxi	China	C	N/A	f237
Gxw_9143	GCA_001541625.1	N/A	N/A	N/A	N/A	N/A	None
HS-06-05	GCA_001280705.1	N/A	N/A	N/A	N/A	N/A	None
HS-13-1	GCA_001270125.1	N/A	N/A	N/A	N/A	N/A	None
IDH02189	GCA_000522025.1	2009	N/A	India	C	3	f237
ISF-01-07	GCA_001267555.1	N/A	N/A	N/A	N/A	N/A	None
ISF-25-6	GCA_001267595.1	N/A	N/A	N/A	N/A	N/A	None
ISF-29-3	GCA_001273575.1	N/A	N/A	N/A	N/A	N/A	None
ISF-54-12	GCA_001280635.1	N/A	N/A	N/A	N/A	N/A	None
ISF-77-01	GCA_001270285.1	N/A	N/A	N/A	N/A	N/A	None
ISF-94-1	GCA_001280645.1	N/A	N/A	N/A	N/A	N/A	None
J-C2-34	GCA_000958655.1	N/A	N/A	N/A	N/A	NF	None
K23	GCA_001497485.1	N/A	N/A	N/A	N/A	N/A	None
K1198	GCA_001188035.1	2004	Alaska	USA	E	59	Other
K1203	GCA_000707585.1	2004	Alaska	USA	E	59	Other
K1461	GCA_000958575.1	N/A	N/A	N/A	N/A	36	None
K5030	GCA_000182465.1	2005	N/A	India	C	3	f237
M0605	GCA_000523375.1	2013	N/A	Mexico	E	539	Other
MA5	In-house	2013	Duxbury, MA	USA	E	NF	Other
MA58	In-house	2014	Wellfleet, MA	USA	E	NF	Other
MA59	In-house	2014	Wellfleet, MA	USA	E	NF	Other*

MA60	In-house	2014	Wellfleet, MA	USA	E	NF	Other
MA76	In-house	2014	Wellfleet, MA	USA	E	NF	Other
MA77	In-house	2014	Wellfleet, MA	USA	E	NF	Other
MA78	In-house	2014	Wellfleet, MA	USA	E	NF	Other
MA97	In-house	2014	Wellfleet, MA	USA	E	NF	Other
MA118	In-house	2014	Wellfleet, MA	USA	E	NF	None
MA137	In-house	2014	Duxbury, MA	USA	E	NF	None
MA143	In-house	2014	Duxbury, MA	USA	E	NF	Other*
MA145	In-house	2014	Duxbury, MA	USA	E	NF	Other
MA146	In-house	2014	Duxbury, MA	USA	E	NF	Other
MA147	In-house	2014	Duxbury, MA	USA	E	NF	Other
MA157	In-house	2014	Duxbury, MA	USA	E	771	None
MA161	In-house	2014	Duxbury, MA	USA	E	NF	Other
MA175	In-house	2014	Katama, MA	USA	E	1399	Other
MA239	In-house	2014	Katama, MA	USA	E	NF	Other
MA267	In-house	2015	Katama, MA	USA	E	1185	Other
MA271	In-house	2015	Katama, MA	USA	E	1185	Other
MA281	In-house	2015	Dennis, MA	USA	E	NF	None
MA303	In-house	2015	Dennis, MA	USA	E	NF	None
MA304	In-house	2015	Dennis, MA	USA	E	NF	None
MA371	In-house	2015	Katama, MA	USA	E	1185	Other
MA398	In-house	2015	Barnstable, MA	USA	E	NF	None
MA414	In-house	2015	Barnstable, MA	USA	E	NF	None
MA432	In-house	2015	Barnstable, MA	USA	E	1185	Other
MA441	In-house	2015	Barnstable, MA	USA	E	NF	None
MA448	In-house	2015	Barnstable, MA	USA	E	1185	Other
MA459	In-house	2015	Dennis, MA	USA	E	NF	None
MA505	In-house	2015	Dennis, MA	USA	E	NF	None
MA561	In-house	2015	Barnstable, MA	USA	E	631	None
MAVP1	In-house	2013	Virginia	USA	C	36	None
MAVP2	In-house	2013	Barnstable, MA; Plymouth, MA; Katama, MA; Virginia; Washington	USA	C	36	Vipa26
MAVP4	In-house	N/A	N/A	N/A	C	N/A	None
MAVP5	In-house	2013	New York	USA	C	NF	None
MAVP6	In-house	2013	Plymouth, MA	USA	C	36	Vipa26
MAVP7	In-house	2013	Duxbury, MA; Kingston, MA; New Brunswick; Maine	USA, Canada	C	36	None
MAVP8	In-house	2013	Duxbury, MA; Kingston, MA; Plymouth, MA; Little Harbor,	USA	C	36	Vipa26

			MA; Barnstable, MA; Maine					
MAVP9	In-house	N/A	N/A	N/A	C	36	Vipa26	
MAVP10	In-house	N/A	N/A	N/A	C	1346	None	
MAVP13	In-house	N/A	N/A	N/A	C	NF	None	
MAVP14	In-house	2013	Barnstable, MA; British Columbia; New Brunswick; Prince Edward Island; Washington	USA, Canada	C	324	None	
MAVP15	In-house	N/A	N/A	N/A	C	1127	None	
MAVP16	In-house	2013	N/A	N/A	C	36	Vipa26	
MAVP19	In-house	2013	Virginia	USA	C	36	None	
MAVP20	In-house	2013	Duxbury, MA	USA	C	36	None	
MAVP21	In-house	N/A	N/A	N/A	C	674	Other	
MAVP23	In-house	2013	Virginia	USA	C	36	None	
MAVP24	In-house	2013	Duxbury, MA; Maine	USA	C	36	Vipa26	
MAVP25	In-house	2013	N/A	N/A	C	1127	None	
MAVP26	In-house	2013	Plymouth, MA; Duxbury, MA	USA	C	36	Vipa26	
MAVP29	In-house	2013	Barnstable, MA; British Columbia; New Brunswick; Prince Edward Island; Washington	USA, Canada	C	36	None	
MAVP30	In-house	2013	N/A	N/A	C	631	None	
MAVP31	In-house	2013	Connecticut	USA	C	36	Vipa36	
MAVP36	In-house	2013	Katama, MA	USA	C	36	None	
MAVP37	In-house	2013	Kingston, MA	USA	C	36	Vipa26	
MAVP38	In-house	2013	Virginia; Westport, MA	USA	C	36	Vipa36	
MAVP39	In-house	N/A	N/A	N/A	C	631	None	
MAVP41	In-house	2013	N/A	N/A	C	36	Other	
MAVP45	In-house	2013	Duxbury, MA; Barnstable, MA; Katama, MA	USA	C	36	Vipa26*	
MAVP46	In-house	N/A	N/A	N/A	C	110	None	
MAVP48	In-house	2013	N/A	N/A	C	36	Vipa36	
MAVP50	In-house	2013	N/A	N/A	C	636	None	
MAVP51	In-house	N/A	N/A	N/A	C	636	None	
MAVP54	In-house	2013	Katama, MA	USA	C	36	None	
MAVP55	In-house	2013	Massachusetts; Canada		C	632	None	
MAVP56	In-house	2013	Prince Edward Island	Canada	C	631	Other	
MAVP57	In-house	2013	Prince Edward Island	Canada	C	636	None	
MAVP60	In-house	2014	Katama, MA	USA	C	36	Vipa36	
MAVP61	In-house	2014	Katama, MA	USA	C	36	Vipa36	
MAVP62	In-house	2014	Katama, MA	USA	C	36	Vipa36	
MAVP63	In-house	2014	Katama, MA	USA	C	36	Vipa36	

MAVP64	In-house	2014	Dennis, MA	USA	C	36	None
MAVP65	In-house	2014	Plymouth, MA	USA	C	NF	None
MAVP66	In-house	2014	Kingston, MA	USA	C	NF	None
MAVP67	In-house	2014	Oyster Pond River, MA	USA	C	308	None
MAVP69	In-house	2014	Buzzards Bay, MA; Eastham, MA; Wellfleet, MA; Katama, MA; Prince Edward Island; Connecticut; Rhode Island; Virginia	USA, Canada	C	43	Other
MAVP70	In-house	2014	Wellfleet, MA; Duxbury, MA; Virginia	USA	C	36	Vipa26
MAVP71	In-house	2014	Wellfleet, MA; Kingston, MA; Duxbury, MA; Popponesset Bay, MA; Stage Harbor, MA; Prince Edward Island; Maine	USA, Canada	C	43	Other
MAVP72	In-house	2014	Wellfleet, MA; Cotuit Bay, MA	USA	C	36	None
MAVP73	In-house	2014	Prince Edward Island; Connecticut	USA, Canada	C	NF	Other
MAVP74	In-house	2014	Prince Edward Island; Connecticut	USA, Canada	C	631	None
MAVP75	In-house	2014	Duxbury, MA; Connecticut	USA	C	631	None
MAVP76	In-house	2014	Barnstable, MA; Virginia	USA	C	614	None
MAVP77	In-house	2014	Wellfleet, MA; Kingston, MA; British Columbia	USA, Canada	C	36	None
MAVP78	In-house	2014	Duxbury, MA	USA	C	631	None
MAVP79	In-house	2014	Waquoit, MA; Katama, MA	USA	C	36	Vipa36
MAVP80	In-house	2015	Katama, MA; Menemsha Inlet and Pond, MA	USA	C	36	Vipa36
MAVP81	In-house	2015	Duxbury, MA; Connecticut; Prince Edward Island	USA, Canada	C	36	None
MAVP82	In-house	2015	Katama, MA	USA	C	36	Vipa36
MAVP83	In-house	2015	Barnstable, MA; Duxbury, MA	USA	C	36	Vipa26
MAVP84	In-house	2015	Katama, MA	USA	C	36	Vipa36
MAVP85	In-house	2015	Katama, MA	USA	C	36	Vipa36
MAVP-A	In-house	2010	N/A	N/A	C	631	None
MAVP-B	In-house	2011	N/A	N/A	C	1127	None
MAVP-E	In-house	2010	Connecticut	USA	C	631	None
MAVP-F	In-house	2011	N/A	N/A	C	NF	None
MAVP-G	In-house	2011	N/A	N/A	C	809	Other
MAVP-H	In-house	2011	Barnstable, MA	USA	C	636	None
MAVP-I	In-house	2011	N/A	N/A	C	N/A	None
MAVP-J	In-house	N/A	N/A	N/A	C	NF	None
MAVP-K	In-house	N/A	N/A	N/A	C	8	Other
MAVP-L	In-house	2011	Hyannis, MA	USA	C	631	None
MAVP-M	In-house	2011	Catuit, MA	USA	C	1127	None
MAVP-N	In-house	2011	N/A	N/A	C	NF	Other
MAVP-P	In-house	N/A	N/A	N/A	C	631	None
MAVP-Q	In-house	2011	Osterville, MA	USA	C	631	None

MAVP-R	In-house	2011	Dennis Port, MA	USA	C	631	None
MAVP-S	In-house	N/A	N/A	N/A	C	NF	Other
MAVP-T	In-house	2010	N/A	N/A	C	631	None
MAVP-U	In-house	2011	N/A	N/A	C	749	Other
MAVP-V	In-house	N/A	N/A	N/A	C	36	None
MAVP-W	In-house	N/A	N/A	N/A	C	43?	Other
MAVP-X	In-house	2011	N/A	N/A	C	322	None
MAVP-Y	In-house	N/A	N/A	N/A	C	43	None
MDOH-04	In-house	2004	Florida	USA	C	3	None
MEVP1	In-house	2013	N/A	N/A	C	110	None
MEVP2	In-house	2013	Maine	USA	C	110	None
MEVP3	In-house	2013	N/A	N/A	C	632	None
MEVP4	In-house	2013	Duxbury, MA	USA	C	36	Vipa26
MEVP5	In-house	2013	N/A	N/A	C	NF	Other
MEVP6	In-house	2013	Maine	USA	C	NF	None
MEVP7	In-house	N/A	N/A	N/A	C	NF	Other
MEVP10	In-house	N/A	N/A	N/A	C	36	Other
NBRC_12711	GCA_000813305.1	N/A	N/A	N/A	N/A	1	None
NCKU_TN_S02	GCA_000736345.1	N/A	N/A	N/A	N/A	247	None
NCKU_TV_3HP	GCA_000736335.1	N/A	N/A	N/A	N/A	970	None
NCKU_TV_5HP	GCA_000736315.1	N/A	N/A	N/A	N/A	970	None
NHVP2	In-house	2013	Maine	USA	C	36	None
NHVP3	In-house	2013	Virginia	USA	C	36	Vipa26
NIHCB0603	GCA_000454265.1	2006	N/A	Bangladesh	C	3	f237
NIHCB0757	GCA_000477475.1	2006	N/A	Bangladesh	C	65	None
NSV_7536	GCA_001471485.1	N/A	N/A	N/A	N/A	N/A	None
NY-3438	GCA_000707565.1	1998	New York	USA	C	36	Vipa10290
Peru288	GCA_000522065.1	2001	N/A	Peru	C	3	f237
peru466	GCA_000182345.1	1996	N/A	Peru	C	3	f237
PMA109_5	GCA_001270805.1	2005	Puerto Montt	Chile	E	3	f237
PMA37.5	GCA_001270835.1	2005	Puerto Montt	Chile	E	3	f237
PMC14_7	GCA_001270895.1	2007	Puerto Montt	Chile	C	3	f237
PMC48	GCA_001270905.1	N/A	N/A	N/A	N/A	N/A	None
PMC58_5	GCA_001270815.1	2005	Puerto Montt	Chile	C	3	f237
PMC58_7	GCA_001270825.1	2007	Puerto Montt	Chile	C	3	f237
RIMD_2210633	GCA_001270945.1	1996	Kansai	Japan	C	3	f237

RM-13-3	GCA_001267965.1	N/A	N/A	N/A	N/A	N/A	N/A	None
RM-14-5	GCA_001273555.1	N/A	N/A	N/A	N/A	N/A	N/A	None
RM-17-6	GCA_001267655.1	N/A	N/A	N/A	N/A	N/A	N/A	None
S176-10	GCA_001280725.1	N/A	N/A	N/A	N/A	N/A	N/A	None
S195-7	GCA_001268005.1	N/A	N/A	N/A	N/A	N/A	N/A	None
S349-10	GCA_001268015.1	2010	N/A	N/A	Canada	E	1516	Other
S357-21	GCA_001273635.1	N/A	N/A	N/A	N/A	N/A	N/A	None
S372-5	GCA_001280655.1	N/A	N/A	N/A	N/A	N/A	N/A	None
S383-6	GCA_001267625.1	2011	N/A	N/A	Canada	N/A	N/A	Other
S439-9	GCA_001270155.1	N/A	N/A	N/A	N/A	N/A	N/A	None
S440-7	GCA_001270235.1	N/A	N/A	N/A	N/A	N/A	N/A	None
S448-16	GCA_001267635.1	2012	N/A	N/A	Canada	N/A	N/A	Other
S456-5	GCA_001268045.1	2012	N/A	N/A	Canada	N/A	N/A	Other
S487-4	GCA_001270215.1	N/A	N/A	N/A	N/A	N/A	N/A	None
S499-7	GCA_001270145.1	2013	N/A	N/A	Canada	E	N/A	Other
SBR10290	GCA_000522045.1	1997	Washington	USA	C	36	Vipa10290	
SG176	GCA_000958565.1	2006		USA	E	NF		
SNUVpS-1	GCA_000315135.1	2012	N/A	Korea	E	917	None	
T12739	GCA_000786835.1	N/A	N/A	N/A	N/A	546	None	
T9109	GCA_000786845.1	N/A	N/A	N/A	N/A	634	None	
TUMSAT_D06_S3	GCA_000591495.1	N/A	N/A	N/A	N/A	413	None	
TUMSAT_DE1_S1	GCA_000591455.1	N/A	N/A	N/A	N/A	114	Other	
TUMSAT_DE2_S2	GCA_000591475.1	N/A	N/A	N/A	N/A	970	None	
TUMSAT_H01_S4	GCA_000591515.1	N/A	N/A	N/A	N/A	698	None	
TUMSAT_H10_S6	GCA_000591555.1	N/A	N/A	N/A	N/A	977	None	
UCM-V493	GCA_000568495.1	2002	N/A	Spain	E	471	None	
V14-01	GCA_000558885.1	2001	N/A	Chile	C	3	f237	
v110	GCA_000388025.1	2010	Hong Kong	China	E	809	None	
V223-04	GCA_000558905.2	2004		Chile	C	NF	f237*	
VH3	GCA_001013435.1	N/A	N/A	N/A	N/A	NF	None	
VIP4-0219	GCA_000500525.1	2006	Hong Kong	China	E	937	Other	
VIP4-0395	GCA_000500505.1	2007	Hong Kong	China	C	3	f237	
VIP4-0407	GCA_000500405.1	2008	Hong Kong	China	C	3	f237	
VIP4-0430	GCA_000500445.1	2008	Hong Kong	China	E	507	Other	
VIP4-0434	GCA_000500425.1	N/A	N/A	N/A	N/A	332	None	
VIP4-0439	GCA_000500365.1	2008	Hong Kong	China	C	3	f237	

VIP4-0443	GCA_000500465.1	N/A		N/A		N/A	N/A	NF	None
VIP4-0444	GCA_000500485.1	N/A		N/A		N/A	N/A	NF	None
VIP4-0445	GCA_000500385.1	2008		Hong Kong		China	C	NF	f237
VIP4-0447	GCA_000500545.1	N/A		N/A		N/A	N/A	396	None
VP-48	GCA_000593285.1	1996		N/A		India	C	N/A	f237
VP-NY4	GCA_000454145.1	1997		N/A		India	C	3	f237
VP1	GCA_000707405.1	2012		Maryland		USA	C	631	None
VP2	GCA_000707165	2012		Maryland		USA	C	651	None
VP3	GCA_000707145.1	2012		Maryland		USA	C	652	None
VP4	GCA_000707025.1	2012		Maryland		USA	C	653	Other
VP5	GCA_000706945.1	2012		Maryland		USA	C	113	None
VP6	GCA_000707065.1	2012		Maryland		USA	C	677	Other*
VP7	GCA_000707305.1	2012		Maryland		USA	C	113	None
VP8	GCA_000707425.1	2012		Maryland		USA	C	631	None
VP9	GCA_000707385.1	2012		Maryland		USA	C	631	None
VP10	GCA_000707445.1	2012		Maryland		USA	C	43	Other
VP11	GCA_000707105.1	2012		Maryland		USA	C	113	None
VP12	GCA_000707225.1	2012		Maryland		USA	C	36	None
VP13	GCA_000707685.1	2012		Maryland		USA	C	678	None
VP14	GCA_000707705.1	2012		Maryland		USA	C	162	Other
VP15	GCA_000707725.1	2012		Maryland		USA	C	679	None
VP16	GCA_000707745.1	2012		Maryland		USA	C	3	f237
VP17	GCA_000707765.1	N/A		N/A		N/A	N/A	3	f237*
VP18	GCA_000707805.1	2012		Maryland		USA	C	3	f237
VP19	GCA_000707785.1	2010		Maryland		USA	C	8	Other
VP20	GCA_000707825.1	2010		Maryland		USA	C	8	Other*
VP21	GCA_000707645.1	2010		Maryland		USA	E	8	Other*
VP22	GCA_000707905.1	2010		Maryland		USA	E	676	None
VP23	GCA_000707665.1	2010		Maryland		USA	E	8	Other
VP24	GCA_000707265.1	2010		Maryland		USA	E	8	Other
VP25	GCA_000707285.1	2010		Maryland		USA	E	810	None
VP26	GCA_000707085.1	2010		Maryland		USA	E	811	None
VP27	GCA_000707365.1	2010		Maryland		USA	E	34	None
VP28	GCA_000707185.1	2010		Maryland		USA	E	768	Other
VP29	GCA_000707345.1	2010		Maryland		USA	E	8	Other
VP30	GCA_000706925.1	2013		Maryland		USA	C	36	None

VP31	GCA_000707445.1	2013	Maryland	USA	C	631	None
VP32	GCA_000707325.1	2013	Maryland	USA	C	36	None
VP33	GCA_000707845.1	2013	Maryland	USA	C	36	None
VP34	GCA_000707005.1	2012	Maryland	USA	C	653	None
VP35	GCA_000707465.1	2013	Maryland	USA	C	631	None
VP36	GCA_000707865.1	2013	Maryland	USA	C	36	None
VP38	GCA_000706845.1	2013	Maryland	USA	C	36	None
VP39	GCA_000706985.1	2013	Maryland	USA	C	896	None
VP40	GCA_000706865.1	2013	Maryland	USA	C	36	None
VP41	GCA_000707485.1	2013	Maryland	USA	C	631	None
VP42	GCA_000706965.1	2013	Maryland	USA	C	36	None
VP43	GCA_000707205.1	2013	Maryland	USA	C	36	None
VP44	GCA_000707505.1	2013	Maryland	USA	C	631	None
VP45	GCA_000706885.1	2013	Maryland	USA	C	631	None
VP46	GCA_000706905.1	2013	Maryland	USA	C	36	None
VP49	GCA_000662375.1	2008	Mangalore	India	E	1024	None
VP232	GCA_000454185.1	1998	N/A	India	C	3	f237
VP250	GCA_000454225.1	1998	N/A	India	C	3	f237
VP551	GCA_000877415.1	2007	Washington	USA	E	3	f237
VP766	GCA_000877605.1	2007	Washington	USA	E	133	None
VP2007-007	GCA_000558925.1	2007	Mississippi	USA	E	306	None
VP2007-095	GCA_000454165.1	2007	Florida	USA	C	631	None
VPCR-2009	GCA_000593305.1	N/A	N/A	N/A	N/A	NF	None
VPCR-2010	GCA_000454475.1	2010	N/A	USA	E	308	None
VPTS-2009	GCA_000593325.1	N/A	N/A	N/A	N/A	1013	None
VPTS-2010	GCA_000593345.1	N/A	N/A	N/A	N/A	6	None
VPTS-2010-2	GCA_000593365.1	N/A	N/A	N/A	N/A	NF	None

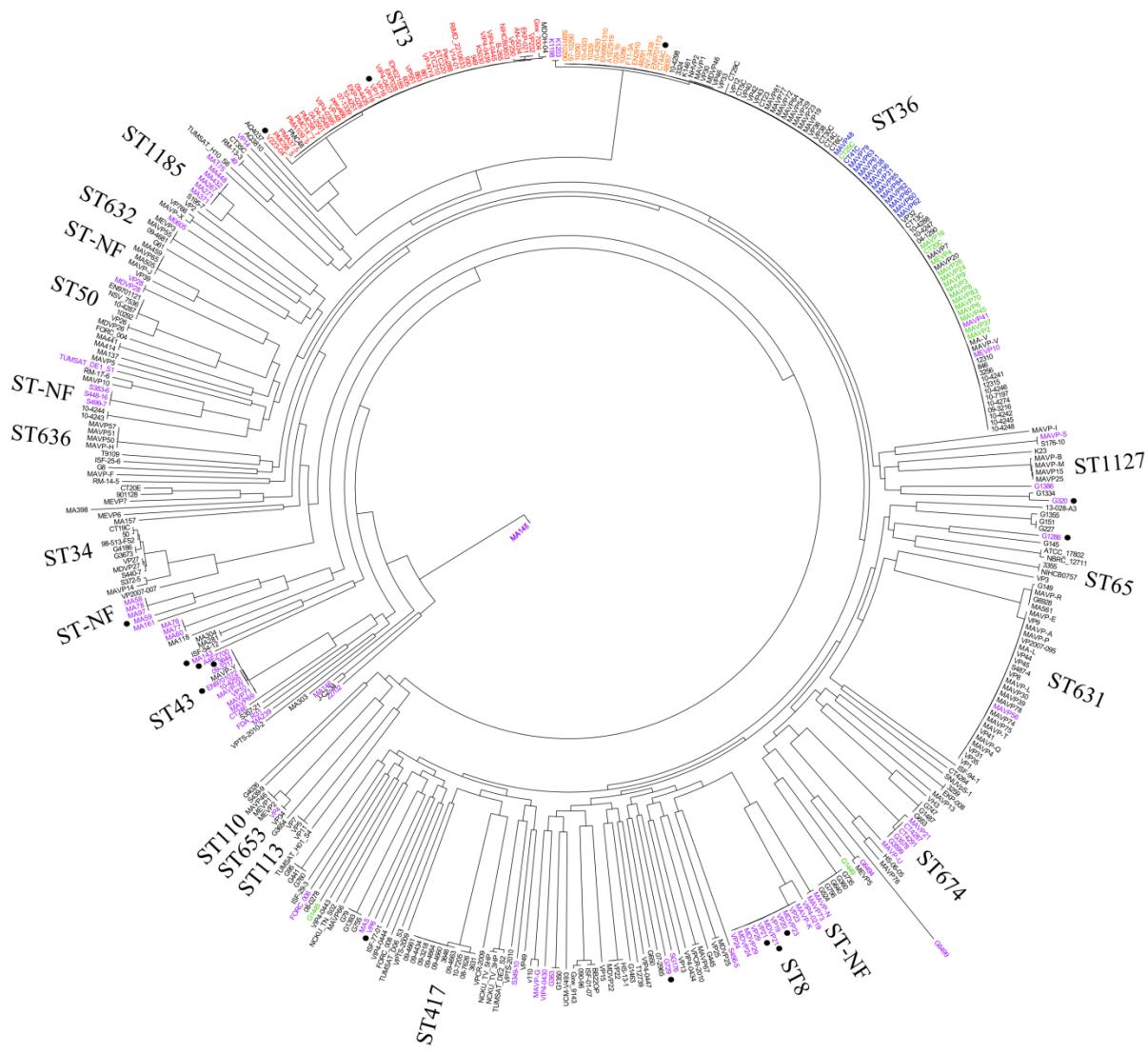


Figure 4.3. Maximum-likelihood tree of *V. parahaemolyticus* strains.

Strains are colored red if they contain f237, orange Vipa10290, green Vipa26, blue Vipa36, and purple for other phage. These other phage are not necessarily identical. The circle after the strain name represents phage content present over multiple scaffolds with complete coverage, or nearly complete coverage over one scaffold. The sequence type (ST) of clades of interest (ST3, ST36, clonal with phage, and clinical strains) is marked. ST-NF refers to sequence types that are not yet defined.

With the exception of two isolates, PMC 48 and MDOH-04, all ST3 strains contained f237. f237 is unique to the pandemic lineage, and is used as a diagnostic identifier [145, 147]. ST36 strains, on the other hand, harbored Vipa10290, Vipa36, and Vipa26 (Fig. 4.3). This indicates these phage were acquired after the evolution of this lineage, and the separation of strains by phage type suggested overall genetic similarity between strains harboring the same phage. Two ST36 strains, MAVP41 and MEVP10, harbored other phage types. The phage in MAVP41 aligned identically to 10593/10893bp of Vipa26 spread across three scaffolds, and no sequence aligned to the remaining 300bp. This strain assembled into 234 scaffolds, which is less than ideal. It is possible MAVP41 harbored Vipa26, but it was not captured in the assembly. MEVP10, on the other hand, assembled into 33 scaffolds and contained f237-like phage content on a single scaffold. This alignment captured the span of ORF10 through ORF9, and the hypothetical genes did not match Vipa26 or Vipa36, so MEVP10 indeed harbored another type of phage. G1445 and G1449, the environmental strains that harbored Vipa26, did not share ancestry with ST36 harboring Vipa26. The overall genetic relationship between *V. parahaemolyticus* strains was driven more by genetic similarities between strains of the same ST than phage type, and characterization of the other phage may reveal additional ST-specific phage and elucidate part of these strains' evolutionary past.

Phylogeography of V. parahaemolyticus ST36 strains and association of Vipa10290, Vipa26 and Vipa36 with subpopulation

Knowing the ST36 lineage acquired phage after evolving, we examined the phylogenetic relationship of just ST36 to further clarify the relationship between these strains. The maximum likelihood tree, which was based on 79.6% of the reference strains, clustered strains into three

main branches (Fig. 4.4). The bottommost branch contained the oldest strains, Pacific Northwest (PNW) strains from 1988-2002. This branch also contained NY-3438 from the Long Island Sound (LIS) in 1998. This suggests there was an initial invasion into the LIS from the PNW in 1998 or earlier, perhaps coinciding with the spread of ST3 that caused a multistate outbreak in 1998 [32, 33]. The second LIS strain in this branch, CT44C, was from 2013. This suggests one of two things: this lineage invaded the LIS twice, or this lineage persisted in the LIS after the initial invasion. Whole genome variant analysis revealed only 70 differences between NY-1998 and CT44C, all of which are single base pair indels or SNPs. However, this lineage did not cause any infections in Connecticut or southern Cape Cod prior to the outbreak in 2013 (NY data is unavailable). This may suggest two separate invasions of this lineage, but could also suggest something (perhaps increasing ocean temperatures [10, 136, 138, 215]) other than the mere presence of the ST36 strains in the environment triggered the outbreak.

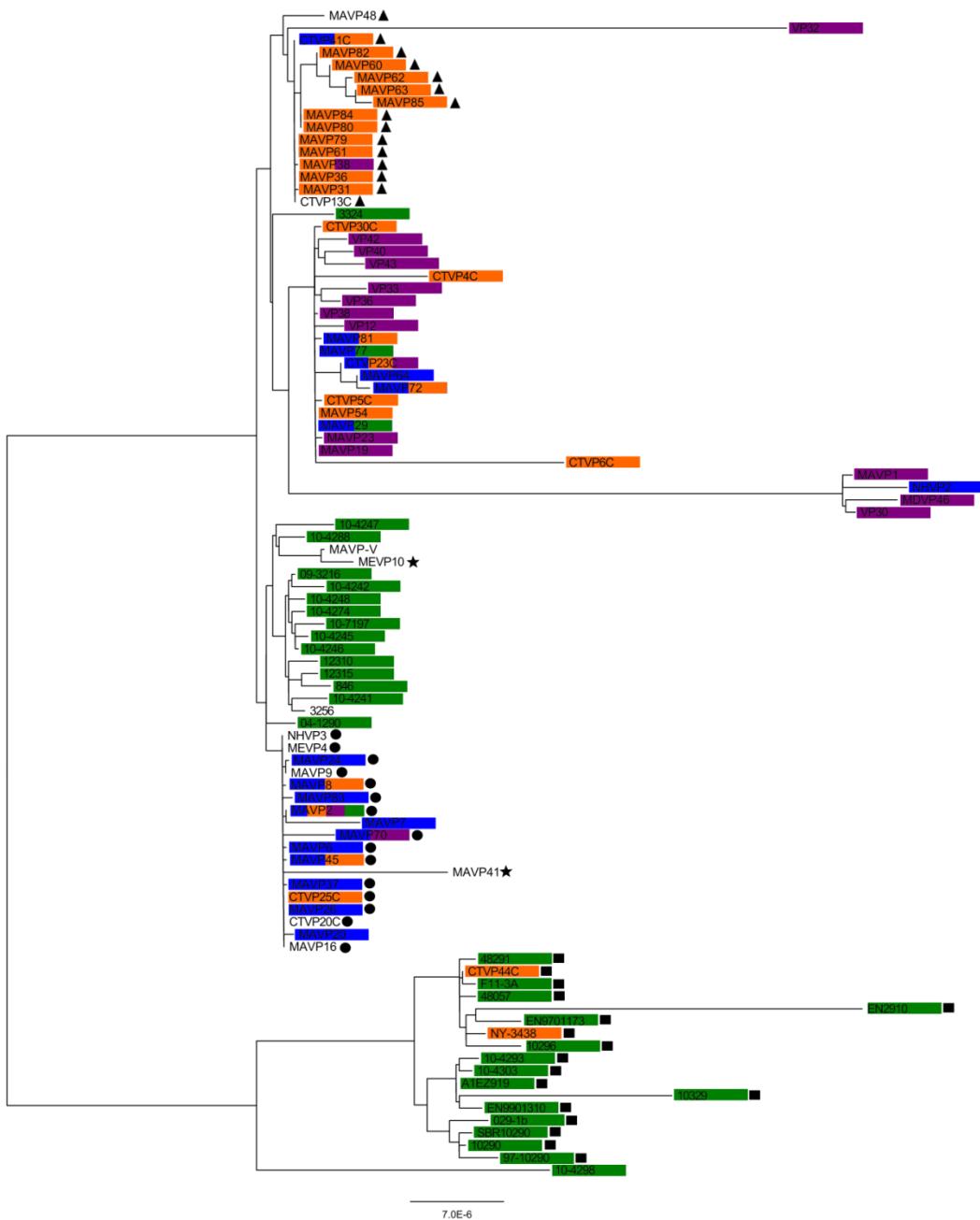


Figure 4.4. Maximum-likelihood tree of *V. parahaemolyticus* ST36 strains.
 Strains are colored based on geographic origin: green in the Pacific Northwest (British Columbia, Washington, Oregon), purple is the mid-Atlantic coast (Maryland, Virginia), orange is the Long Island Sound (New York, Connecticut, southern Cape Cod), and blue is the Gulf of Maine (northern Cape Cod, Maine, Prince Edward Island). Strains with no coloring had no traceback data, and strains with multiple colors were unable to be traced to a single region. Phage type is indicated by a shape following the strain name: the square is Vipa10290, the circle is Vipa26, the triangle is Vipa36, and the star is other. The strains with other do not have the same phage.

The middle branch in this phylogenetic tree contained PNW strains from 2003-2009, Gulf of Maine (GOM) strains from 2013-2015, and a few strains with mixed or no traceback data (Fig. 4.4). This tree topology suggests the population of strains that caused illness in the PNW shifted, supporting data showing the emergence of a new lineage in the PNW [136]. However, the ST36 strains causing illness in the August 2012 outbreak in Galicia, Spain are most closely related to the older PNW lineage [136], suggesting this population is still present in the environment. The PFGE pattern of ST36 outbreak strains from Galicia is identical to ST36 strains from NY from June 2012 [141], implying possible origin. The absence of data is by no means conclusive, but there were no publicly available strains from 2002-2012 that closely relate to this older lineage, from either the LIS or the PNW, so it is possible the outbreak in Spain was caused by this older PNW lineage persisting in the LIS.

The GOM ST36 strains were more closely related to the newer PNW lineage than the LIS strains, despite only being separated by Cape Cod. The LIS strains are clustered in the upper branch, most closely related to the mid-Atlantic coast (MAC) strains and a single PNW strain, 3324. While one strain is not enough to confidently draw conclusions about the ancestry of the MIC and LIS strains, the ancestry of the MAC and LIS strains was likely shared, and was a different ancestry than the GOM population, agreeing with previous work using fewer strains [231]. Distinct subpopulations of the PNW established within different regions of the Atlantic coast, perhaps indicative of separate invasion events or mechanisms of transport.

Both the geographic origin and phylogenetic relationship of ST36 strains correlated with phage content. All strains from the older PNW lineage with phage harbored Vipa10290 (Fig. 4.4) and this phage was not present in any other strains (ST36 or otherwise, Fig. 4.3). The alignment of the Vipa10290 sequences revealed high similarity between the strains; the only

differences were a few gaps in 48291, NY-3438, F11-3A, and 48057. Gaps are eliminated when generating both the distance matrix and phylogenetic tree, so the Vipa10290 sequences appeared identical in these analyses (data not shown).

ST36 strains with Vipa36 were all traced to the LIS, and the alignment and distance matrix revealed the sequences of Vipa36 were 100% identical (0 base substitutions per site). This phage was not in any other phage-harboring strain (Fig. 4.3). While the LIS and mid-Atlantic coast populations have shared ancestry, the ancestor that initially acquired Vipa36 remained localized enough to replicate into a distinct phage-harboring subpopulation in the LIS.

ST36 strains with Vipa26 were only traced to the GOM, with the exception of CT25C. This could be explained by inaccurate traceback, as it is at times quite difficult to pinpoint the exact source of the oysters, especially during a multistate outbreak or multi-area exposure. An alignment and distance matrix of Vipa26 revealed the phage in the ST36 strains were also 100% identical, whereas the two environmental strains harboring the phage had 0.002 base substitutions per site compared to the other strains. This was reflected in the Vipa26 neighbor-joining tree (Fig. 4.5). The assembly of Vipa26 in G1449 contained several ambiguous bases, which likely caused G1449 to appear more closely related to the clinical strains than G1445, despite several SNPs shared between the environmental strains and not the clinical strains. This identity between the clinical strains implies an early GOM ST36 ancestor cell encountered Vipa26 and replicated to produce a phage-harboring population. These three phage appear to be promising for use as a biomarker for inferred ancestry and more refined trace-back for managing disease and outbreaks.

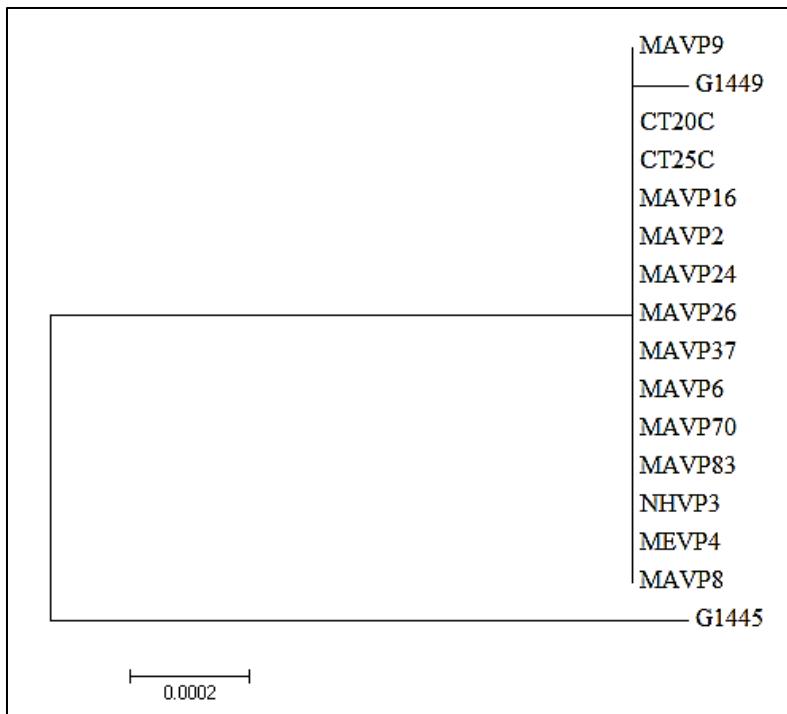


Figure 4.5. Phylogenetic tree of Vipa26.

Neighbor-joining tree of Vipa26 phage harbored in ST36 strains and environmental New Hampshire strains from the Gulf of Maine.

There were two ST36 strains possibly harboring f237-like phage other than Vipa10290, Vipa26, and Vipa36 (Fig. 4.4). MAVP41 was from an unknown source in 2013, and although somewhat divergent (possibly due to the poor assembly) grouped with the GOM ST36 strains and may actually harbor Vipa26. MEVP10, with no traceback data, grouped closely with the 2003-2009 PNW strains. Characterizing the MEVP10 phage, and other phage harbored in *V. parahaemolyticus* strains (Fig. 4.3), may provide other biomarkers assisting with strain traceback and/or strain relatedness.

CONCLUSIONS

In this study, we analyzed the genomes of *V. parahaemolyticus* strains to elucidate the origin of the United States Atlantic-coast ST36 population. We determined ST36 strains harbor three phage with similar architecture to f237: Vipa10290, Vipa26, and Vipa36. Of the conserved genes, two are secondary toxins in CTXΦ [218]. The accessory cholera toxin increases short-circuit current and causes fluid secretion [206]. The zona occludens toxin [68] modifies tight junctions to decrease intestinal tissue resistance [69], and is found in other enteric pathogens [207]. ST3 has increased cytotoxicity and adherence compared to non-ST3 strains, perhaps due to f237 [232]. While not all ST36 strains harbor f237-like phage, these phage, coupled with other virulence-associated traits, may increase virulence. This could be tested by introducing very closely related phage-harboring and phage-deficient strains (e.g., MAVP26 and MAVP20, respectively) to an animal model to assess differences in virulence.

V. cholerae virulence-associated traits, including CTXΦ, were horizontally acquired from a diverse environmental population [67]. Environmental New England *V. parahaemolyticus* populations harbored diverse phage, and whereas Vipa36 was not identified, Vipa26 was detected in two environmental strains from 2009. These strains pre-dated the outbreaks, implying that the GOM population acquired Vipa26 after arriving to this region. In New England, there have been far more phage-harboring than phage-deficient clinical ST36 strains, so harboring f237-like phage may provide not only increased virulence but a fitness advantage in the environment. Phage have the ability to alter bacterial community structure under various selective conditions [21, 24, 90]. Current attempts at demonstrating a fitness advantage in Vipa26 and Vipa36 phage-harboring strains were unsuccessful [163], but fitness advantage is dependent on the pressure source [231] and/or sensitivity of the competitor strain [108].

Modifications to the methodology (i.e., triggering phage replication prior to competition, using nutrient-poor media) may reveal an advantage.

Whole genome phylogenies revealed a phylogeographic relationship between ST36 strains that suggested multiple invasions into the Atlantic coast. Subpopulations of ST36 harbored phage unique to geographic origin that have promise for use in strain traceback. These phage can be detected with our PCR primers, but the primers should be modified and compared to the sequenced environmental f237-like phage for increased accuracy. Such tools permit refined shellfish management: having the ability to rapidly pinpoint which oyster beds are causing infections can help focus closures to prevent further illness and avoid broad impacts on the shellfish industry. As these phage were horizontally acquired, their usefulness as a biomarker may become limited as populations evolve. However, the phylogenetic trees we generated were reference-based, and not all ST36 strains harbored f237-like phage, so genetic content other than phage drove the phylogenetic groups that correlated with geographic origin. Further examination of the DNA content of these groupings will likely reveal further gene content distinct to subpopulations.

METHODS

Acquiring, characterizing, and sequencing *Vibrio parahaemolyticus* strains

This study utilized whole genome sequences of both publicly available and in-house *Vibrio parahaemolyticus* strains to determine relationships between sequence types and phage content. The genomes of all publicly-available *V. parahaemolyticus* strains were downloaded from NCBI [<https://www.ncbi.nlm.nih.gov/genome/691>] as of February 2016. Strains with poor assemblies were disregarded, leaving 221 *V. parahaemolyticus* isolates for analysis (Table 4.2).

The sequence type of these strains was inferred by cross-referencing the PubMLST database [www.pubmlst.org]. Assembled strains not in the database were cut into 50mers and sequence-typed using SRST2 [101].

In-house isolates were obtained from several sources. Clinical *V. parahaemolyticus* strains were received from Massachusetts, Maine, New Hampshire, and Connecticut State Departments spanning 2010-2015 [231]. Environmental strains were isolated from New Hampshire, Connecticut, and Massachusetts oysters between 2007-2015 during Most Probable Number enumeration as previously described [107, 132, 135, 184, 210]. All clinical and environmental isolates were confirmed as *V. parahaemolyticus* by detecting the thermolabile hemolysin gene (*tlh*) with PCR as published [161]. Reactions were then visualized on a 0.7% agarose gel and samples with a band at the expected size for *tlh* (450bp) were scored as positive for *V. parahaemolyticus*.

New Hampshire, Massachusetts, and Connecticut environmental isolates were screened with PCR for phage presence. Step-wise PCR primers were developed to detect f237-like phage content (1000bp), New England (NE)-like phage content (618bp) and Vipa26 and Vipa36 (1440bp and 854bp, respectively) that can be used in conjunction with *tlh* primers as an internal control (Table 4.3). The f237-like primers target ORF3-ORF5, part of the seven conserved genes in all f237-like phage. The NE primers amplify ORF10-*hypA*. *hypA* was present in Vipa26 and Vipa36, but not Vipa10290 or f237 (Fig. 4.1). The final set of primers amplifies the region of hypothetical protein genes in Vipa26 and Vipa36 between *hypD* and ORF9, with each phage producing bands of different lengths reflective of the respective number of genes. Each 10µL reaction contained 1x Accustart II Supermix, 1µL DNA template, 0.2µM of each forward and reverse primer, and nuclease-free water to volume. The first reaction used f237-like, NE-

like, and *tlh* primers at 94°C for 3 minutes; 30 cycles of 94°C for 1 minute, 55°C for 1 minute, and 72°C for 1 minute; and 72°C for 5 minutes. The second reaction used the Vipa26/36 primers and *tlh* primers at the same conditions, with the exception of a 1.5 minute elongation time.

Table 4.3. PCR primers used in this study.

Step	Primers	Genes amplified	5'-3' primer sequence	Reference
f237-like	ST36Phage F2	ORF3-ORF5	AGCAACGAAAACGCCTGT	This paper
	ST36Phage R2		ACCGTATCACCAATGGACTGT	This paper
NE-like	NEORF10F	ORF10- <i>hypA</i>	TTTCTTACTTCTGTGAGCATTGA	This paper
	NEHypR		GATTACTGAGCCTCTAAAGCCGTC	This paper
Vipa26/36	PhHypDF3	<i>hypD</i> -ORF9	AAGTGCTACATGAATGAAAGTGCT	This paper
	PhORF9R1		TCAATGAAGTATCACGAAATGACTA	This paper
<i>tih</i> Control	TLH-F2 TLH-R	<i>tih</i>	AGAACTTCATCTTGATGACACTGC GCTACTTCTAGCATTCTCTGC	This paper [161]

In-house isolates of interest (representing various results from the PCR phage screen) were sequenced at the Hubbard Center for Genome Studies (University of New Hampshire, Durham, NH, USA) on an Illumina HiSeq 2500 (Illumina, Inc., San Diego, California, USA) as previously described [231]. Reads were *de novo* assembled with the A5 pipeline [200] and sequence-typed with SRST2 [100] and/or PCR amplification and sequencing the house-keeping loci [231].

Analyzing genomic content and relationships of *V. parahaemolyticus* strains

Whole genome variant analysis was performed on sequence type (ST) 36 strains from the Pacific Northwest (10290) and Massachusetts (MAVP26, MAVP36) using *breseq* [49]. All sequenced *V. parahaemolyticus* strains were evaluated for f237-like phage content (ORF1 – zona occludens toxin) and specific phage type using BLAST [4]. To determine if the phage were

undergoing selective pressure, a Nei-Gojobori codon-based Z test [148] was performed in Mega 6 [199] on the seven core genes shared between f237, Vipa26, Vipa36, and Vipa10290.

The evolutionary relatedness of all *V. parahaemolyticus* strains, f237-like phage-harboring *V. parahaemolyticus* strains, and ST36 strains was determined with multiple-reference maximum-likelihood phylogenetic trees built with RealPhy [17] using bowtie2 [124] for reference mapping and PhyML [87] for generating the tree, visualized with FigTree 1.4.2 [172]. The references for the trees of all *V. parahaemolyticus* strains and the phage-harboring strains were RIMD-2210633, FDA-R31, and MAVP26, and the references for the ST36 strains were 10290, MAVP26, and MAVP36.

After aligning in MEGA 6 [199] with ClustalW [93] at default settings, pairwise distance matrices of Vipa10290, Vipa26, and Vipa36 were generated using the Jukes Cantor model [110] at default settings to determine similarity of the phage sequence. A neighbor-joining tree [175] was generated for Vipa26 in MEGA 6 [185] using Jukes Cantor model [110] at default settings.

ACKNOWLEDGEMENTS

We thank all the following for providing strains, traceback data, and oysters: S. Condon, K. Foley, T. Stiles, and J. Cyr from the Massachusetts Department of Public Health; M. Hickey, C. Schillaci, D. Murphy, D. Regan from the Massachusetts Division of Marine Fisheries; J.C. Mahoney (former) from the NH Department of Public Health; A. Robbins from the Maine CDC; J.K. Kanwit from the Maine Department of Marine Resources; K. DeRosia-Banick and J.A. DeCrescenzo from the Connecticut Department of Agriculture Bureau of Aquaculture.

We also thank Tiffany DeGroot, Bethany Parker, Kari Hartman, Malachi Hallee, Sabrina Pankey, Randi Foxall, and Marcus Dillon for laboratory assistance, and Crystal Ellis, Jennifer

Mahoney, Brian Schuster, Jong Yu, Rachel Donner, Matthew Gerding, and members of the Jones laboratory for ongoing environmental isolate collections, and Dr. Steve Jones for access to regional clinical and environmental strains.

SUPPLEMENTAL INFORMATION

Table 4S.1. All results of screening environmental strains for phage.

Strains are ordered by collection site and then strain. Primers sets (Table 4.3) that produced a band of expected size for a given strain is designated by the +, primers that did not produce a band are designated by the -, and primer sets untested on a given strain are designated by NT. Some of these results have been disproven through whole genome sequencing (see Table 4.2).

Collection Site	Strain	f237-like	NE phage	Vipa26	Vipa36
Connecticut	CT1E	-	-	NT	NT
	CT2E	-	-	NT	NT
	CT3E	-	-	NT	NT
	CT4E	+	+	-	+
	CT5E	-	-	NT	NT
	CT6E	-	-	NT	NT
	CT7E	-	-	NT	NT
	CT8E	-	-	NT	NT
	CT9E	-	-	NT	NT
	CT10E	+	-	NT	NT
	CT11E	-	-	NT	NT
	CT12E	-	-	NT	NT
	CT13E	-	-	NT	NT
	CT14E	+	-	NT	NT
	CT15E	-	-	NT	NT
	CT16E	-	-	NT	NT
	CT17E	-	-	NT	NT
	CT18E	-	-	NT	NT
	CT19E	-	+	-	+
	CT20E	-	-	NT	NT
	CT21E	-	-	NT	NT
	CT22E	-	-	NT	NT
	CT23E	-	-	NT	NT
	CT24E	+	-	+	-
	CT25E	-	-	NT	NT
	CT4238	-	-	NT	NT

	CT4240	-	-	NT	NT
	CT4242	-	-	NT	NT
	CT4243	-	-	NT	NT
	CT4245	-	-	NT	NT
	CT4247	-	-	NT	NT
	CT4250	-	-	NT	NT
	CT4252	-	-	NT	NT
	CT4254	-	-	NT	NT
	CT4258	-	-	NT	NT
	CT4259	-	-	NT	NT
	CT4261	-	-	NT	NT
	CT4262	-	-	NT	NT
	CT4264	+	+	+	-
	CT4266	-	-	NT	NT
	CT4267	-	-	NT	NT
	CT4270	-	-	NT	NT
	CT4280	-	-	NT	NT
	CT4285	-	-	NT	NT
	CT4286	-	-	NT	NT
	CT4287	-	-	NT	NT
	CT4290	-	-	NT	NT
	CT4291	-	-	NT	NT
	CT4296	-	-	NT	NT
	CT4300	-	-	NT	NT
Barnstable, MA	MA306	-	-	NT	NT
	MA307	-	-	NT	NT
	MA308	+	-	-	-
	MA309	+	-	NT	NT
	MA310	-	-	NT	NT
	MA311	-	-	NT	NT
	MA312	+	-	-	-
	MA313	-	-	NT	NT
	MA314	-	-	NT	NT
	MA315	-	-	NT	NT
	MA316	-	-	NT	NT
	MA317	-	-	NT	NT
	MA318	+	-	-	-
	MA319	+	-	NT	NT
	MA320	+	+	-	-
	MA321	+	-	-	-
	MA322	-	-	NT	NT
	MA323	-	-	NT	NT
	MA324	-	-	NT	NT

	MA325	+	+	-	-
	MA326	-	-	NT	NT
	MA327	-	-	NT	NT
	MA328	+	+	+	-
	MA329	-	-	NT	NT
	MA330	-	-	NT	NT
	MA331	-	-	NT	NT
	MA332	+	-	-	-
	MA333	-	-	NT	NT
	MA334	+	-	-	-
	MA335	-	-	NT	NT
Dennis, MA	MA336	-	-	NT	NT
	MA337	-	-	NT	NT
	MA338	-	-	NT	NT
	MA340	-	+	-	-
	MA341	-	-	NT	NT
	MA342	+	+	-	-
	MA343	-	-	NT	NT
	MA344	-	-	NT	NT
	MA345	-	-	NT	NT
	MA346	-	-	NT	NT
	MA347	-	-	NT	NT
	MA348	-	-	NT	NT
	MA349	+	-	-	-
	MA350	-	-	NT	NT
	MA351	-	-	NT	NT
	MA352	+	-	-	-
	MA353	+	-	-	-
	MA354	+	-	-	-
	MA355	+	-	-	-
	MA356	-	-	NT	NT
	MA358	-	-	NT	NT
	MA359	-	-	NT	NT
	MA360	-	-	NT	NT
	MA361	-	-	NT	NT
	MA362	-	-	NT	NT
	MA363	-	-	NT	NT
	MA364	-	-	NT	NT
	MA365	-	-	NT	NT
	MA366	-	-	NT	NT
	MA367	-	-	NT	NT

	MA368	-	-	NT	NT
	MA369	-	-	NT	NT
	MA370	-	-	NT	NT
Duxbury, MA	MA5	+	+	-	+
	MA35	-	-	NT	NT
	MA43	-	-	NT	NT
	MA48	-	-	NT	NT
	MA51	-	-	NT	NT
	MA54	-	-	NT	NT
	MA55	-	-	NT	NT
	MA104	-	-	NT	NT
	MA105	-	-	NT	NT
	MA106	-	-	NT	NT
	MA107	-	-	NT	NT
	MA109	-	-	NT	NT
	MA136	-	-	NT	NT
	MA138	-	-	NT	NT
	MA139	-	-	NT	NT
	MA143	+	+	+	-
	MA144	-	+	+	+
	MA145	+	+	+	-
	MA146	+	+	+	+
	MA147	+	+	+	-
	MA148	-	-	NT	NT
	MA155	+	+	-	-
	MA156	-	-	NT	NT
	MA157	-	-	+	-
	MA158	-	-	+	-
	MA159	-	+	-	+
	MA160	-	+	-	+
	MA224	+	+	-	+
Katama, MA	MA173	-	-	NT	NT
	MA175	+	-	NT	NT
	MA177	+	-	NT	NT
	MA185	-	-	NT	NT
	MA187	-	-	NT	NT
	MA189	-	-	NT	NT
	MA191	+	-	NT	NT
	MA193	+	-	NT	NT
	MA195	+	-	NT	NT
	MA199	-	-	NT	NT
	MA203	-	-	NT	NT
	MA205	-	-	NT	NT
	MA207	-	-	NT	NT
	MA214	+	-	NT	NT

	MA216	+	-	NT	NT
	MA220	-	-	NT	NT
	MA235	-	-	NT	NT
	MA237	-	-	NT	NT
	MA239	+	+	+	-
	MA240	-	-	NT	NT
	MA241	-	+	-	-
	MA242	-	-	NT	NT
	MA243	+	-	NT	NT
	MA244	-	-	NT	NT
	MA245	+	-	NT	NT
	MA248	-	-	+	-
	MA250	+	+	-	-
Wellfleet, MA	MA58	+	+	-	+
	MA59	-	-	-	+
	MA60	-	+	+	-
	MA61	-	+	+	-
	MA62	+	-	-	+
	MA63	-	-	NT	NT
	MA64	+	-	-	-
	MA66	+	-	-	+
	MA67	+	+	+	+
	MA68	-	-	NT	NT
	MA70	-	-	NT	NT
	MA71	+	-	-	+
	MA72	-	-	NT	NT
	MA73	-	-	NT	NT
	MA76	+	+	+	-
	MA77	+	+	+	-
	MA78	+	+	-	+
	MA82	+	-	NT	NT
	MA83	+	-	NT	NT
	MA84	-	-	NT	NT
	MA85	+	-	NT	NT
	MA86	-	-	NT	NT
	MA87	+	-	NT	NT
	MA88	-	-	NT	NT
	MA90	+	+	+	-
	MA91	-	-	NT	NT
	MA92	-	-	NT	NT
	MA94	-	-	NT	NT
	MA95	-	-	NT	NT

	MA96	-	-	NT	NT
	MA97	+	-	-	+
	MA98	+	+	-	+
	MA99	+	-	NT	NT
	MA100	-	-	NT	NT
	MA110	+	+	+	+
	MA112	-	-	NT	NT
	MA113	-	-	NT	NT
	MA114	-	-	NT	NT
	MA115	+	-	-	+
	MA116	-	-	NT	NT
	MA117	-	-	NT	NT
	MA118	+	+	-	+
	MA119	-	-	NT	NT
	MA121	-	-	NT	NT
	MA122	-	-	NT	NT
	MA123	-	-	NT	NT
	MA125	-	+	NT	NT
	MA126	-	-	NT	NT
	MA127	-	-	NT	NT
	MA128	-	-	NT	NT
	MA129	-	-	NT	NT
	MA130	-	-	NT	NT
	MA132	-	-	NT	NT
	MA133	-	-	NT	NT
	MA134	-	-	NT	NT
	MA135	+	-	NT	NT
	MA161	+	+	-	+
	MA162	-	-	NT	NT
	MA163	-	-	NT	NT
	MA167	-	-	NT	NT
	MA169	-	-	NT	NT
	MA170	-	-	-	-
New Hampshire	G1	-	+	-	-
	G4	-	-	NT	NT
	G6	-	-	NT	NT
	G7	+	-	-	-
	G8	-	-	NT	NT
	G10	-	-	NT	NT
	G12	-	-	NT	NT
	G23	-	-	NT	NT
	G25	-	+	-	-

	G26	-	-	NT	NT
	G31	-	-	NT	NT
	G43	-	-	NT	NT
	G61	-	-	NT	NT
	G62	-	-	NT	NT
	G69	-	+	-	-
	G74	-	-	NT	NT
	G79	+	-	-	-
	G95	-	-	NT	NT
	G227	-	-	NT	NT
	G235	-	-	NT	NT
	G242	-	-	NT	NT
	G316	-	-	NT	NT
	G317	-	-	NT	NT
	G319	-	-	NT	NT
	G325	+	+	+	-
	G365	-	-	NT	NT
	G377	-	-	NT	NT
	G387	-	-	NT	NT
	G389	-	-	NT	NT
	G401	-	-	NT	NT
	G407	+	+	-	-
	G409	-	-	NT	NT
	G412	-	-	NT	NT
	G416	-	-	NT	NT
	G420	-	-	NT	NT
	G425	-	-	NT	NT
	G426	-	-	NT	NT
	G438	-	-	NT	NT
	G439	-	-	NT	NT
	G441	-	-	NT	NT
	G454	-	-	NT	NT
	G478	-	-	NT	NT
	G486	-	-	NT	NT
	G487	-	-	NT	NT
	G498	-	-	NT	NT
	G500	-	-	NT	NT
	G524	-	-	NT	NT
	G535	-	-	NT	NT
	G575	-	-	NT	NT
	G632	-	-	NT	NT
	G633	-	-	NT	NT
	G650	-	-	NT	NT
	G747	-	-	NT	NT

	G1286	+	-	-	-
	G1386	+	-	-	+
	G1388	-	-	NT	NT
	G1389	-	+	-	-
	G1390	-	-	NT	NT
	G1391	-	-	NT	NT
	G1393	-	-	NT	NT
	G1394	-	-	NT	NT
	G1398	-	-	NT	NT
	G1401	-	-	NT	NT
	G1405	-	-	NT	NT
	G1425	-	-	NT	NT
	G1440	-	-	NT	NT
	G1445	+	+	+	-
	G1449	-	+	-	-
	G1463	-	+	-	-
	G1466	+	+	-	-
	G1470	-	-	NT	NT
	G1474	-	-	NT	NT
	G1480	-	-	NT	NT
	G1487	-	-	NT	NT
	G1488	-	-	NT	NT
	G1489	-	-	NT	NT
	G1491	-	-	NT	NT
	G1493	-	-	NT	NT
	G1498	-	-	NT	NT
	G1501	-	-	NT	NT
	G1502	-	+	-	-
	G1503	-	-	NT	NT
	G1505	-	-	NT	NT
	G1507	-	-	NT	NT
	G1508	+	-	-	-
	G1513	-	-	NT	NT
	G6353	-	-	NT	NT
	G6354	-	-	NT	NT
	G6362	+	+	-	-
	G6363	+	+	-	+
	G6364	-	-	NT	NT
	G6365	-	-	NT	NT
	G6367	-	-	NT	NT
	G6368	+	+	-	-
	G6369	-	-	NT	NT
	G6371	-	-	NT	NT
	G6372	-	-	NT	NT
	G6373	-	-	NT	NT

	G6374	-	-	NT	NT
	G6375	-	-	NT	NT
	G6376	-	-	NT	NT
	G6377	+	+	-	-
	G6379	-	-	NT	NT
	G6380	-	+	-	-
	G6381	-	-	NT	NT
	G6382	-	+	-	-
	G6383	-	-	NT	NT
	G6385	+	+	-	+
	G6386	-	-	NT	NT
	G6387	-	-	NT	NT
	G6388	-	-	NT	NT
	G6389	-	-	NT	NT
	G6390	-	-	NT	NT
	G6391	-	-	NT	NT
	G6392	-	-	NT	NT
	G6393	-	-	NT	NT
	G6454	-	-	NT	NT
	G6457	-	+	-	-
	G6459	+	-	-	-
	G6462	-	-	NT	NT
	G6464	-	-	NT	NT
	G6475	-	-	NT	NT
	G6484	-	-	NT	NT
	G6494	+	-	-	+
	G6496	-	-	NT	NT
	G6497	-	-	NT	NT
	G6498	-	-	NT	NT
	G6499	+	+	+	-
	G6500	-	-	NT	NT
	G6507	-	-	NT	NT
	G6509	-	+	-	-
	G6510	-	-	NT	NT
	G6511	-	-	NT	NT
	G6512	-	-	NT	NT
	G6513	-	-	NT	NT
	G6514	-	-	NT	NT

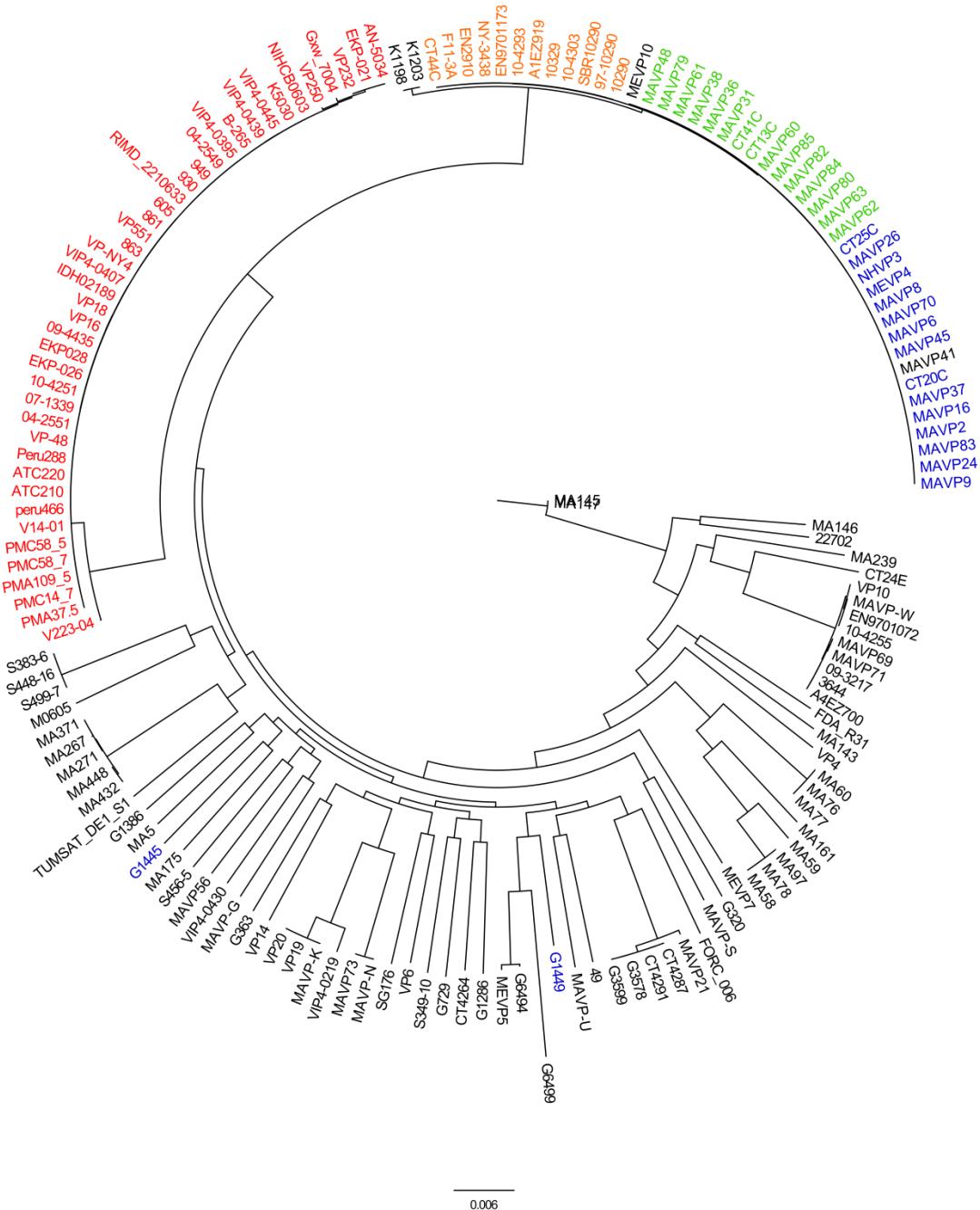


Figure 4S.1. Maximum-likelihood tree of phage-harboring *V. parahaemolyticus* ST36 strains.

Strains are colored red if they contain f237, orange Vipa10290, green Vipa36, blue Vipa26, and black for other phage. These other phage are not necessarily identical.

REFERENCES

- [1] Aagesen AM, Phuvasate S, Su Y-C, Häse CC. 2013. Persistence of *Vibrio parahaemolyticus* in the Pacific oyster, *Crassostrea gigas*, is a multifactorial process involving pili and flagellabut not type III secretion systems or phase variation. *Appl Environ Microbiol* **79**(10):3303-3305.
- [2] Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* **7**(12):8966-8969.
- [3] Ainsworth TD, Krause L, Bridge T, Torda G, Raina JB, Zakrzewski M, Gates RD, Padilla-Gamiño JL, Spaulding HL, Smith C, Woolsey ES, Bourne DG, Bonhaerts P, Hoegh-Guldberg, Leggat W. 2015. The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. *ISME J* **9**:2261-2274.
- [4] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**:403-410.
- [5] Asha Devi NK, Balakrishnan K, Gopal R, Padmavathy S. 2008. *Bacillus clausii* MB9 from the east coast regions of India: isolation, biochemical characterization, and antimicrobial potentials. *Curr Sci* **95**(5):627-636.
- [6] Asplund ME, Rehnstam-Holm A, Atnur V, Raghunath P, Saravanan V, Härnström K, B. Collin, Karunasagar I, Godhe A. 2011. Water column dynamics of *Vibrio* in relation to phytoplankton community opposition and environmental conditions in a tropical coastal area. *Environ Microbiol* **13**(10):2738-2751.
- [7] Ausubel F, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. 1990. Current protocols in molecular biology. Wiley and Sons, Inc., New York, N.Y.
- [8] Avila-Poveda OH, Torres-Ariño A, Girón-Cruz DA, Cuevas-Aquirre A. 2014. Evidence for accumulation of *Synechococcus elongatus* (Cyanobacteria: Cyanophyceae) in the tissues of the oyster *Crassostrea gigas* (Mollusca: Bivalvia). *Tissue Cell* **46**(5):379-387.
- [9] Baffone W, Tarsi R, Pane L, Campana R, Repetto B, Mariottini GL, Pruzzo C. 2006. Detection of free-living and plankton-bound vibrios in coastal waters of the Adriatic Sea (Italy) and study of their pathogenicity-associated properties. *Environ Microbiol* **8**(7):1299-1305.
- [10] Baker-Austin C, Trinanes JA, Taylor NGH, Hartnell R, Siitonen A, Martinez-Urtaza J. 2013. Emerging *Vibrio* risk at high latitudes in response to ocean warming. *Nature Clim Change* **3**:73-77.

- [11] **Balcázar JL, Rojas-Luna T.** 2007. Inhibitory activity of probiotic *Bacillus subtilis* UTM 126 against *Vibrio* species confers protection against vibriosis in juvenile shrimp (*Litopenaeus vannamei*). *Curr Microbiol* **55**(5):409-412.
- [12] **Balcázar JL, Rojas-Luna T, Cunningham DP.** 2007. Effect of the addition of four potential probiotic strains on the survival of Pacific white shrimp (*Litopenaeus vannamei*) following immersion challenge with *Vibrio parahaemolyticus*. *Journ Invert Pathol* **96**(2):147-150.
- [13] **Banerjee SK, Kearney AK, Nadon CA, Peterson C-L, Tyler K, Bakouche L, Clark CG, Hoang L, Gilmour MW, Farber JM.** 2014. Phenotypic and genotypic characterization of Canadian clinical isolates of *Vibrio parahaemolyticus* collected from 2000 to 2009. *J Clin Microbiol* **52**(4):1081-1088.
- [14] **Barriuso J, Valverde JR, Mellado RP.** 2011. Estimation of bacterial diversity using next generation sequencing of 16S rDNA: a comparison of different workflows. *BMC Bioinformatics* **12**:473.
- [15] **Bean NH, Goulding JS, Lao C, Angulo FJ.** 1996. Surveillance for foodborne-disease outbreaks - United States, 1988-1992. *MMWR* **45**(SS-5):1-55.
- [16] **Berg KA, Lyra C, Sivonen K, Paulin L, Suomalainen S, Tuomi P, Rapala J.** 2009. High diversity of cultivable heterotrophic bacteria in association with cyanobacterial water blooms. *ISME J* **3**(3):314-325.
- [17] **Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E.** 2014. Automated reconstruction of whole genome phylogenies from short sequence reads. *Mol Biol Evol* **31**(5):1077-1088.
- [18] **Beuchat, LR.** 1973. Interacting effects of pH, temperature, and salt concentration on growth and survival of *Vibrio parahaemolyticus*. *Appl Microbiol* **25**(5):844-846.
- [19] **Birkbeck TH, McHenery JG.** 1982. Degradation of bacteria by *Mytilus edulis*. *Marine Biol* **72**:7-15.
- [20] **Bohnhoff M, Drake BL, Miller CP.** 1954. Effect of streptomycin on susceptibility of intestinal tract to experimental *Salmonella* infection. *Proc Soc Exp Biol Med* **86**(1):132-137.
- [21] **Bossi L, Fuentes JA, Mora G, Figueroa-Bossi N.** 2003. Prophage contribution to bacterial population dynamics. *J Bacteriol* **185**(21):6467-6471.
- [22] **Boyd EF, Cohen ALV, Naughton LM, Ussery DW, Binnewies TT, Stine OC, Parent MA.** 2008. Molecular analysis of the emergence of pandemic *Vibrio parahaemolyticus*. *BMC Microbiol* **8**:110.

- [23] **Brady YJ, Vines CR, Delaney MA.** 1998. Microbial flora of offshore relayed oysters, *Crassostrea virginica*, in Mobile Bay, Alabama. In: Aquaculture 98 Book of Abstracts, February 15–19, 1998, Las Vegas, Nevada. p. 63.
- [24] **Brockhurst MA, Fenton A, Roulston B, Rainey PB.** 2006. The impact of phages on interspecific competition in experimental populations of bacteria. *BMC Ecol* **6**:19.
- [25] **Brooks JP, Edwards DJ, Harwich Jr MD, Rivera MC, Fettweis JM, Serrano MG, Reris RA, Sheth NU, Huang B, Girerd P.** 2015. The truth about metagenomics: quantifying and counteracting biases in 16S rRNA studies. *BMC Microbiol* **15**:66.
- [26] **Brown CT, Sharon I, Thomas BC, Castelle CJ, Morowitz MJ, Banfield JF.** 2013. Genome resolved analysis of a premature infant gut microbial community reveals a *Varibaculum cambriense* genome and a shift towards fermentation-based metabolism during the third week of life. *Microbiome* **1**:30.
- [27] **Burkhardt III W, Calci KR.** 2000. Selective accumulation may account for shellfish-assocaited viral illness. *Appl Environ Microbiol* **66**(4):1375-1378.
- [28] **Cabrera-García ME, Vázquez-Salinas C, Quiñones-Ramírez EI.** 2004. Serologic and molecular characterization of *Vibrio parahaemolyticus* strains isolated from seawater and fish products of the Gulf of Mexico. *Appl Environ Microbiol* **70**(11): 6401-6406.
- [29] **Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R.** 2010. PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**(2):266-267.
- [30] **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Gonzalez-Peña A, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koeing JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsuneko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* **7**(5): 335-336.
- [31] **Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R.** 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* **108**(S1):4516-4522.
- [32] **CDC.** 1998. Outbreak of *Vibrio parahaemolyticus* infections associated with eating raw oysters - Pacific Northwest, 1997. *MMWR* **47**(22):457-462.
- [33] **CDC.** 1999. Outbreak of *Vibrio parahaemolyticus* infection associated with eating raw oysters and clams harvested from Long Island Sound – Connecticut, New Jersey, and New York, 1998. *MMWR* **48**(03):48-51.

[34] **CDC.** Cholera-*Vibrio cholerae* infection. <http://www.cdc.gov/cholera/index.html>. Accessed 14-Dec-15.

[35] **CDC.** Foodborne Diseases Active Surveillance Network (FoodNet): FoodNet Reports. <http://www.cdc.gov/foodnet/reports/index.html>. Accessed 11-Dec-15.

[36] **CDC.** National Surveillance of Bacterial Foodborne Illnesses (Enteric Diseases): National Cholera and Vibrosis Surveillance. <http://www.cdc.gov/nationalsurveillance/cholera-vibrio-surveillance.html>. Accessed 11-Dec-15.

[37] **CDC.** *Vibrio* illness (Vibrosis): *Vibrio vulnificus*. <http://www.cdc.gov/vibrio/vibriov.html>. Accessed 14-Dec-15.

[38] **Chakraborty S, Nair GB, Shinoda S.** 1997. Pathogenic vibrios in the natural aquatic environment. *Rev Environ Health* **12**(2):63-80.

[39] **Chauhan A, Wafula D, Lewis DE, Pathak A.** 2014. Metagenomic assessment of the eastern oyster-associated microbiota. *Genome Announc* **2**(5):e01083-14.

[40] **Chen H, Liu Z, Wang M, Chen S, Chen T.** 2013. Characterization of the spoilage bacterial microbiota in oyster gills during storage at different temperatures. *J Sci Food Agric* **93**(15):3748-3754.

[41] **Chen W, Zhang CK, Cheng Y, Zhang S, Zhao H.** 2013. A comparison of methods for clustering 16S rRNA sequencing into OTUs. *PLoS ONE* **8**(8):e70837.

[42] **Chen X, Li S, Aksoy S.** 1999. Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *Journ Mol Evol* **48**(1):49-58.

[43] **Chimetto LA, Brocchi M, Thompson CC, Martins RCR, Ramos HR, Thompson FL.** 2008. Vibrios dominate as culturable nitrogen-fixing bacteria of the Brazilian coral *Mussismilia hispida*. *Syst Appl Microbiol* **31**(4):312-319.

[44] **Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen A.S., McGarrell DM, Marsh T, Garrity GM, Tiedje JM.** 2008. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids R* **37**(s1):D141-D145.

[45] **Colwell, RR.** 1970. Polyphasic taxonomy of the genus *Vibrio*: Numeric taxonomy of *Vibrio cholerae*, *Vibrio parahaemolyticus*, and related *Vibrio* species. *J Bacteriol* **104**(1):410-433.

[46] **Consortium, The Human Microbiome.** 2012. Structure, function, and diversity of the healthy human microbiome. *Nature* **486**:207-214.

[47] **Cox AM, Gomez-Chiarri M.** 2012. *Vibrio parahaemolyticus* in Rhode Island coastal ponds and the estuarine environment of Narragansett Bay. *Appl Environ Microbiol* **78**(8):2996-2999.

- [48] Daniels NA, MacKinnon L, Bishop R., Altekruse S, Ray B, Hammond RM, Thompson S, Wilson S, Bean NH, Griffin PM, Slutsker L. 2000. *Vibrio parahaemolyticus* infections in the United States, 1973-1998. *J Infect Dis* **181**(5):1661-1666.
- [49] Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory evolved microbes from next-generation sequencing data using *breseq*. *Methods Mol Biol* **1151**:165-188.
- [50] Deepanjali A, Kumar HS, Karunasagar I, Karunasagar I. 2005. Seasonal variation in abundance of total and pathogenic *Vibrio parahaemolyticus* bacteria in oysters along the southwest coast of India. *Appl Enviro Microbiol* **71**(7):3575-3580.
- [51] DePaola A, Hopkins LH, Peeler JT, Wentz B, McPhearson RM. 1990. Incidence of *Vibrio parahaemolyticus* in U.S. coastal waters and oysters. *Appl Environ Microbiol* **56**(8):2299-2302.
- [52] DePaola A, Nordstrom JL, Bowers JC, Wells JG, Cook DW. 2003. Seasonal abundance of total and pathogenic *Vibrio parahaemolyticus* in Alabama oysters. *Appl Environ Microbiol* **69**(3):1521-1526.
- [53] DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**(7):5069-5072.
- [54] Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP, Raoult D. 2000. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical bacterial isolates. *J Clin Microbiol* **38**(10):3623-3630.
- [55] Duan J, Su Y. 2005. Occurrence of *Vibrio parahaemolyticus* in two Oregon oyster-growing bays. *J Food Sci* **70**(1):M58-M63.
- [56] Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**(19):2460-2461.
- [57] Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimaera detection. *Bioinformatics* **27**(16):2194-2200.
- [58] Evans J, Sheneman L, Foster JA. 2006. Relaxed Neighbor-Joining: A Fast Distance-Based Phylogenetic Tree Construction Method. *J Mol Evol* **62**(6):785-792.
- [59] Eiler A, Gonzalez-Rey C, Allen S, Bertilsson S. 2007. Growth response of *Vibrio cholerae* and other *Vibrio* spp. to cyanobacterial dissolved organic matter and temperature in brackish water. *FEMS Microbiol Ecol* **60**(3):411-418.

- [60] **Eiler A, Johansson M, Bertilsson S.** 2006. Environmental influences on *Vibrio* populations in northern temperate and boreal coastal waters (Baltic and Skagerrak Seas). *Appl Environ Microbiol* **72**(9):6004-6011.
- [61] **Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky, Chen F, Ochman H, Hugenholtz P.** 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J* **4**(5):642-647.
- [62] **Faith DP.** 1992. Conservation evaluation and phylogenetic diversity. *Biol Conserv* **61**:1-10.
- [63] **Faith DP, Baker AM.** 2006. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evol Bioinform Online* **2**:121-128.
- [64] **Falcioni T, Papa S, Campana R, Manti A, Battistelli M, Baffone W.** 2008. State transitions of *Vibrio parahaemolyticus* VBNC cells evaluated by flow cytometry. *Cytometry B Clin Cytom* **74**(5):272-281.
- [65] **Farmer III JJ.** 1979. *Vibrio ("Beneckeia") vulnificus*, the bacterium associated with sepsis, septicaemia, and the sea. *Lancet*: **314**(8148):903.
- [66] **Faruque SM, Kamruzzaman M, Asadulghani M, Sack DA, Mekalanos JJ, Nair GB.** 2002. CTXΦ-independent production of the RS1 satellite phage by *Vibrio cholerae*. *Proc Natl Acad Sci USA* **100**(3):1280-1285.
- [67] **Faruque SM, Mekalanos JJ.** 2003. Pathogenicity islands and phages in *Vibrio cholerae* evolution. *Trends Microbiol* **11**(11):505-510.
- [68] **Fasano A, Baudry B, Pumplin DW, Wasserman SS, Tall BD, Ketley JM, Kaper JB.** 1991. *Vibrio cholerae* produces a second enterotoxin, which affects intestinal tight junctions. *Proc Natl Acad Sci USA* **88**:5242- 5246.
- [69] **Fasano A, Fiorentini C, Donelli G, Uzzau S, Kaper JB, Margaretten K, Ding X, Guandalini S, Cornstock L, Goldblum SE.** 1995. Zona occludens toxin modulates tight junctions through protein kinase C-dependent actin reorganization, *in vitro*. *J Clin Invest* **96**(2):710-720.
- [70] **FDA.** Appendix 5: FDA and EPA Safety Levels in Regulations and Guidance. <http://www.fda.gov/downloads/Food/GuidanceRegulation/UCM252448.pdf>. Accessed 18-Dec-15.
- [71] **FDA.** Bacteriological analytical manual. <http://www.fda.gov/Food/FoodScienceResearch/LaboratoryMethods/ucm2006949.htm>. Accessed 8-Mar-16.

- [72] **FDA.** 2013. Guide for the Control of Molluscan Shellfish 2013 Revision. <http://www.fda.gov/downloads/Food/GuidanceRegulation/FederalStateFoodPrograms/UCM415522.pdf>. Accessed 5-Jan-16.
- [73] **Feinstein LM, Sul WJ, Blackwood CB.** 2009. Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Appl Environ Microbiol* **75**(6):5428-5433.
- [74] **Fox GE, Magrum LJ, Balch WE, Wolfe RS, Woese CR.** 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc Natl Acad Sci* **74**(10):4537-4541.
- [75] **Froelich B, Ayrapetyan M, Oliver JD.** 2013. Integration of *Vibrio vulnificus* into marine aggregates and its subsequent uptake by *Crassostrea virginicia* oysters. *Appl Environ Microbiol* **79**(5):1454-1458.
- [76] **Froelich BA, Oliver JD.** 2013. Increases in the amounts of *Vibrio* spp. in oysters upon addition of exogenous bacteria. *Appl Environ Microbiol* **79**(17):5208-5213.
- [77] **Froelich BA, Ringwood A, Sokolova I, Oliver JD.** 2009. Uptake and depuration of the C- and E-geneotypes of *Vibrio vulnificus* by the Eastern oyster (*Crassostrea virginicia*). *Environ Microbiol Rep* **2**(1):112-115.
- [78] **Froelich BA, Williams T, Noble R, Oliver JD.** 2012. Apparent loss of *Vibrio vulnificus* in North Carolina oysters coincides with a drought-induced increase in salinity. *Appl Environ Microbiol* **78**(11):3885-3889.
- [79] **Fujino T, Okuno Y, Nakada D, Aoyama A, Fukai K, Murai K, Ueho T.** 1953. On the bacteriological examination of “shirasu” food poisoning. *Med J Osaka Univ* **4**:299-304.
- [80] **Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Kazutoshi Y, Tobe T, Clarke JM, Topping DL, Suzuki T, Taylor TD, Itoh K, Kikuchi J, Morita H, Hattori M, Ohno H.** 2011. Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**(7331):543-547.
- [81] **Gaspar JM, Thomas WK.** 2015. FlowClus: Efficiently filtering and denoising pyrosequenced amplicons. *BMC Bioinformatics* **16**:105.
- [82] **Genther FJ, Volety AK, Oliver LM, Fisher WS.** 1999. Factors influencing *in vitro* killing of bacteria by hemocytes of the Eastern oyster (*Crassostrea virginica*). *Appl Enviro Microbiol* **65**(7):3015-3020.
- [83] **Givens CE, Bowers JC, DePaola A, Hollibaugh JT, Jones JL.** 2014. Occurrence and distribution of *Vibrio vulnificus* and *Vibrio parahaemolyticus* – potential roles for fish, oyster, sediment, and water. *Letters in Appl Microbiol* **58**(6):503-510.

- [84] Givens CE, Ransom B, Bano N, Hollibaugh JT. 2015. Comparison of the gut microbiomes of 12 bony fish and 3 shark species. Mar Ecol Prog Ser **518**:209-223.
- [85] González-Escalona N, Cachicas V, Acevedo C, Rioseco ML, Vergara JA, Cabello F, Romero J, Espejo RT. 2005. *Vibrio parahaemolyticus* diarrhea, Chile, 1998 and 2004. Emerg Infect Dis **11**(1):129-131.
- [86] Gould SJ. 1996. Planet of the Bacteria. *Washington Post Horizon*, **119** (344): H1.
- [87] Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol **59**(3):307-321.
- [88] Haendiges J, Timme R, Allard MW, Myers RA, Brown EW, González-Escalona N. 2015. Characterization of *Vibrio parahaemolyticus* clinical strains from Maryland (2012-2013) and comparisons to a locally and globally diverse *V. parahaemolyticus* strains by whole-genome sequence analysis. Front Microbiol **6**:125.
- [89] Hara-Kudo Y, Takatori K. 2011. Contamination level and ingestion dose of foodborne pathogens associated with infections. Epidemiol Infect **139**(10):1505-1510.
- [90] Harcombe WR, Bull JJ. 2005. Impact of phages on two-species bacterial communities. Appl Environ Microbiol **71**(9):5254-5259.
- [91] Hauben L, Vauterin L, Swings J, Moore ERB. 1997. Comparison of 16S ribosomal DNA sequences of all *Xanthomonas* species. Int J Syst Evol Microbiol **47**(2):328-335.
- [92] Hentges DJ, Freter R. 1962. In vivo and in vitro antagonism of intestinal bacteria against *Shigella flexneri*. I. Correlation between various tests. J Infect Dis **110**(1):30-37.
- [93] Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22**:4673-4680.
- [94] Hoashi K, Ogata K, Taniguchi H, Yamashita H, Tsuji K, Mizuguchi Y, Ohomo N. 1990. Pathogenesis of *Vibrio parahaemolyticus*: intraperitoneal and orogastric challenge experiments in mice. Microbiol Immun **34**(4):355-366.
- [95] Hoffmann M, Fischer M, Ottesen A, McCarthy PJ, Lopez JV, Brown EW, Monday SR. 2010. Population dynamics of *Vibrio* spp. associated with marine sponge microcosms. ISME J **4**(12):1608-1612.
- [96] Hollis DG, Weaver RE, Baker CN, Thornsberry C. 1976. Halophilic *Vibrio* species isolated from blood cultures. Journ Clin Microbiol **3**(4):425-431.

- [97] **Hosokawa T, Nikoh N, Koga R, Satô M, Tanahashi M, Meng X-Y, Fukatsu T.** 2012. Reductive genome evolution, host-symbiont co-speciation and uterine transmission of endosymbiotic bacteria in bat flies. *ISME J* **6**:577-587.
- [98] **Howard-Jones N.** 1984. Robert Koch and the cholera vibrio: a centenary. *Br Med J (Clin Res Ed)* **288**(6414): 379-381.
- [99] **Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM.** 2007. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* **8**(7):R143.
- [100] **Iida T, Makino K, Nasu H, Yokoyama K, Tagomori K, Hattori A, Okuno T, Shinagawa H, Honda T.** 2002. Filamentous bacteriophages of vibrios are integrated into the *dif*-like site of the host chromosome. *J Bacteriol* **184**(17):4933-4935.
- [101] **Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE.** 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* **6**:90.
- [102] **Islam MS, Goldar MM, Morshed MG, Khan Mn, Islam MR, Sack RB.** 2002. Involvement of the *hap* gene (mucinase) in the survival of *Vibrio cholerae* O1 in association with the blue-green alga, *Anabaena* sp. *Can J Microbiol* **48**(9):793-800.
- [103] **Islam MS, Mahmuda S, Morshed MG, Bakht HB, Khan MN, Sack RB, Sack DA.** 2004. Role of cyanobacteria in the persistence of *Vibrio cholerae* O139 in saline microcosms. *Can J Microbiol* **50**(2):127-131.
- [104] **Janda JM, Abbot SL.** 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* **45**(9):2761-2764.
- [105] **Jay E, Bambara R, Padmanabhan R, Wu R.** 1973. DNA sequence analysis: a general, simple and rapid method for sequencing large oligodeoxyribonucleotide fragments by mapping. *Nucleic Acids R* **1**(3):331-354.
- [106] **Jones SH, Howell TL, O'Neill KR, Langan R.** 1995. Strategies for removal of indicator and pathogenic bacteria from commercially-harvested shellfish, p 69-77. *In* R. Poggi and J-Y Le Gall (ed), *Shellfish Depuration: Second International Conference on Shellfish Depuration*. IFREMER, Brest, France.
- [107] **Jones SH, Striplin MJ, Mahoney JC, Cooper VS, Whistler CA.** 2010. Incidence and abundance of pathogenic Vibrio species in the Great Bay Estuary, New Hampshire, p 127-134. *In* Lassus P (ed), *Proceedings of the Seventh International Conference on Molluscan Shellfish Safety*. Quae Publishing, Nantes, France.
- [108] **Joo J, Gunny M, Cases M, Hudson P, Albert R, Harvill E.** 2006. Bacteriophage-mediated competition in *Bordetella* bacteria. *Proc Biol Sci* **273**(1595):1843-1848.

- [109] **Jordan DC.** 1982. NOTES: Transfer of *Rhizobioum japonicum* Buchanan 1980 to *Brazyrhizobium* gen. nov., a genus of slow-growing, root nodule bacteria from leguminous plants. *Int J Syst Evol Microbiol* **32**:136-139.
- [110] **Jukes TH, Cantor CR.** 1969. Evolution of protein molecules. In Munro HN, editor, *Mammalian Protein Metabolism*, pp. 21-132, Academic Press, New York.
- [111] **Kamada N, Chen GY, Onohara N, Núñez G.** 2013. Control of pathogens and pathobionts by the gut microbiota. *Nat Immunol* **14**(7):685-690.
- [112] **Kaneko T, Cowell RR.** 1973. Ecology of *Vibrio parahaemolyticus* in Chesapeake Bay. *J Bacteriol* **113**(1):24-32.
- [113] **Kapareiko D, Lim HJ, Schott EJ, Hanif A, Wikfors GH.** 2011. Isolation and evaluation of new probiotic bacteria for use in shellfish hatcheries: II. Effects of a *Vibrio* sp. probiotic candidate upon survival of oyster larvae (*Crassostrea virginica*) in pilot-scale trials. *J Shellfish Res* **30**(3):617-625.
- [114] **Kaufman GE, Bej AK, DePaola A.** 2003. Oyster-to-oyster variability in levels of *Vibrio parahaemolyticus*. *J Food Prot* **66**(1):125-129.
- [115] **Keenan SW, Engel AS, Elsey RM.** 2013. The alligator gut microbiome and implications for archosaur symbioses. *Sci Rep* **3**:2877.
- [116] **Kennedy J, Flemer B, Jackson SA, Morrisey JP, O'Gara F, Dobson AD.** 2014. Evidence of a putative deep sea specific microbiome in marine sponges. *PLoS ONE* **9**(7):e103400.
- [117] **Kennedy K, Hall MW, Lynch MDJ, Moreno-Hagelsieb G, Neufield JD.** 2014. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Appl Environ Microbiol* **80**(18):5717-5722.
- [118] **Kennedy NA, Walker AW, Berry SH, Duncan SH, Farquharson FM, Louis P, Thomson JM.** 2014. The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS ONE* **9**(2):e88982.
- [119] **King GM, Judd C, Kuske CR, Smith C.** 2012. Analysis of stomach and gut microbiomes of the eastern oyster (*Crassostrea virginica*) from coastal Louisiana, USA. *PLoS One* **7**(12):e51475.
- [120] **Kinross JM, Darzi AW, Nicholson JK.** Gut microbiome-host interactions in health and disease. *Genomic Med* **3**:14.
- [121] **Koch R.** 1884. Sechster Bericht der deutschen Wissenschaftlichen Commission

zur Enforschung der Cholera, Geh Regierungsraths Dr Koch. Dtsch Med Wochenschr 10:191-192.

[122] Krishnamurthi S, Chakrabarti T, Stackebrandt E. 2009. Re-examination of the taxonomic position of *Bacillus silvestris* Rheims et al. 1999 and proposal to transfer it to *Solibacillus* gen. nov. as *Solibacillus silvestris* comb. nov. Int J Syst Evol Microbiol 59(Pt 5):1054-1058.

[123] Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Thurber RLV, Knight R, Beiko RG, Huttenhower C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31:814-821.

[124] Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9(4):357-359.

[125] Lemos LN, Fulthorpe RR, Roesch LFW. 2012. Low sequencing efforts bias analyses of shared taxa in microbial communities. Folia Microbiol 57(5):409-413.

[126] Levy P-Y, Fenollar F, Stein A, Borrione F, Raoult D. 2009. *Finegoldia magna*: a forgotten pathogen in prosthetic joint infection rediscovered by molecular biology. Clin Infect Dis 49(8):1244-1247.

[127] Li J, Chen Q, Long L-J, Dong J-D, Yang J, Zhang S. 2014. Bacterial dynamics within the mucus, tissue, and skeleton of the coral *Porites lutea* during different seasons. Sci Rep 4:7320.

[128] Liu XF, Li Y, Li JR, Cai LY, Li XX, Chen JR, Lyu SX. 2015. Isolation and characterization of *Bacillus* spp. antagonistic to *Vibrio parahaemolyticus* for use as probiotics in aquaculture. World J Microbiol Biotechnol 31(15):795-803.

[129] Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids R 35(18):e120-.

[130] Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. 2011. UniFrac: an effective distance metric for microbial community comparison. ISME J 5(2):169-172.

[131] Lydon KA, Farrell-Evans M, Jones JL. 2015. Evaluation of ice slurries as a control for postharvest growth of *Vibrio* species in oysters and potential for filth contamination. J Food Prot 78(7):1375-1379.

[132] Mahoney JC, Gerding MJ, Jones SH, Whistler CA. 2010. Comparison of the pathogenic potentials of environmental and clinical *Vibrio parahaemolyticus* strains indicates a role for temperature regulation in virulence. Appl. Environ. Microbiol 76(22):7459-7465.

- [133] **Main CR, Salvitti LR, Whereat EB, Coyne KJ.** 2015. Community-level and species-specific associations between phytoplankton and particle-associated *Vibrio* species in Delaware's inland bays. *Appl Environ Microbiol* **81**(17):5703-5713.
- [134] **Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, Nalin R, Jarrin C, Chardon P, Marteau P, Roca P, Dore J.** 2006. Reduced diversity of faecal microbiota in Chron's disease revealed by a metagenetic approach. *Gut* **55**:205-211.
- [135] **Marcinkiewicz AL, Schuster BM, Jones SH, Cooper VS, Whistler CA.** 2016. Bacterial community profiling of individual oysters: taxonomic and functional associations with *Vibrio parahaemolyticus* abundance. In preparation.
- [136] **Martinez-Urtaza J.** 2015. Bacterial emerging risk to seafood safety. Presented at Seafood and Emerging Food Safety Issues in Madrid, Spain, 25 Nov. 2015.
http://www.aecosan.msssi.gob.es/AECOSAN/docs/documentos/eventos/2015/seafood/3_MARTINEZ-URTAZA.pdf. Accessed 26-Mar-16.
- [137] **Martinez-Urtaza J, Baker-Austin C, Jones JL, Newton AE, Gonzalez-Aviles GD, DePaola A.** 2013. Spread of Pacific Northwest *Vibrio parahaemolyticus* strain. *N Engl J Med* **369**:1573-1574.
- [138] **Martinez-Urtaza J, Boers JC, Trinanes J, DePaola A.** 2010. Climate anomalies and the increasing risk of *Vibrio parahaemolyticus* and *Vibrio vulnificus* illnesses. *Food Res Int* **43**(7):1780-1790.
- [139] **Martinez-Urtaza J, Huapaya B, Gavilan RG, Blanco-Abad V, Ansede-Bermejo J, Cadarso-Suarez C, Figueiras A, Trinanes J.** 2008. Emergence of Asiatic diseases in South America in phase with El Niño. *Epidemiology* **19**(6):829-837.
- [140] **Martinez-Urtaza J, Lozano-Leon A, Varela-Pet J, Trinanes J, Pazos Y, Garcia-Martin O.** 2008. Environmental determinants of the occurrence and distribution of *Vibrio parahaemolyticus* in the rias of Galicia, Spain. *Appl Environ Microbiol* **74**(1):265-274.
- [141] **Martinez-Urtaza J, Powell A, Jansa J, Rey JLC, Montero OP, Campello MG, López MJZ, Pousa A, Valles MJF, Trinanes J, Hervio-Heath D, Keay W, Bayley A, Hartnell R, Baker-Austin C.** 2016. Epidemiological investigation of a foodborne outbreak in Spain associated with U.S. west coast genotypes of *Vibrio parahaemolyticus*. *SpringerPlus* **5**:87.
- [142] **Mignard S, Flandrois JP.** 2006. 16S rRNA sequencing in routine bacterial identification: a 30-month experiment. *J Microbiol Methods* **67**(3):574-581.
- [143] **Mouriño-Pérez RR, Worden AZ, Azam F.** 2003. Growth of *Vibrio cholerae* O1 in red tide waters off California. *Appl Environ Microbiol* **69**(11):6923-6931.
- [144] **Murphree RL, Tamplin ML.** 1331. Uptake and retention of *Vibrio cholerae* O1 in the Eastern oyster, *Crassostrea virginica*. *Appl Environ Microbiol* **61**(10):3656-3660.

[145] **Myers ML, Panicker G, Bej AK.** 2003. PCR detection of a newly emerged pandemic *Vibrio parahaemolyticus* O3:K6 pathogen in pure cultures and seeded waters from the Gulf of Mexico. *Appl Environ Microbiol* **69**(4):2194-2200.

[146] **Nair GB, Ramamurthy T, Bhattacharya SK, Dutta B, Takeda Y, Sack DA.** 2007. Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. *Clin Microbiol Rev* **20**(1):39-48.

[147] **Nasu H, Iida T, Sugahara T, Yamaichi Y, Park KS, Yokoyama K, Makino K, Shinagawa H, Honda T.** 2000. A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:K6 strains. *J Clin Microbiol* **38**(6):2156-2161.

[148] **Nei M, Gojobori T.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Bioland Evol* **3**(5):418-426.

[149] **Newton AE, Garrett N, Stroika SG, Halpin JL, Turnsek M, Mody RK.** 2014. Notes from the field: increase in *Vibrio parahaemolyticus* infections associated with consumption of Atlantic coast shellfish-2013. *MMWR* **63**(15):335-336

[150] **Newton AE, Kendall M, Vugia DJ, Henao OL, Mahon BE.** 2012. Increasing rates of vibriosis in the United States, 1996–2010: Review of surveillance data from 2 systems. *Clin Infect Dis* **54**(5):S391–S395.

[151] **Nishiguchi MK, Ruby EG, McFall-Ngai M.** 1998. Competitive dominance among strains of luminous bacteria provides an unusual form of evidence for parallel evolution in sepiolid squid-Vibrio symbioses. *Appl Environ Microbiol* **64**(9):3209-3213.

[152] **Noh ES, Kim Y-S, Kim D-H, Kim K-H.** 2013. Bacterial diversity in the guts of sea cucumbers (*Apostichopus japonicas*) and shrimps (*Litopenaeus vannamei*) investigated with tag-encoded 454 pyrosequencing of 16S rRNA genes. *Korean Journal od Microbiology* **49**(3):237-244.

[153] **Nordstrom JL, Vickery MCL, Blackstone GM, Murray SL, DePaola A.** 2007. Development of a multiplex real-time PCR assay with an internal amplification control for the detection of total and pathogenic *Vibrio parahaemolyticus* bacteria in oysters. *Appl Enviro Microbiol* **73**(18):5840-5847.

[154] **O'Neill KR, Jones SH, Grimes DJ.** 1990. Incidence of *Vibrio vulnificus* in northern New England water and shellfish. *FEMS Microbiol Lett* **60**(1-2):163-167.

[155] **Okuda J, Ishibashi M, Hayakawa E, Nishino T, Takeda Y, Mukhopadhyay AK, Garg S, Bhattacharya SK, Nair GB, Nishibuchi M.** 1997. Emergence of a unique O3:K6 clone of *Vibrio parahaemolyticus* in Calcutta, India, and isolation of strains from the same clonal group from Southeast Asian travelers arriving in Japan. *J Clin Microbiol*. **35**(12):3150-3155.

- [156] Olafsen JA, Mikkelsen HV, Giaevers HM, Hansen GH. 1993. Indigenous bacteria in hemolymph and tissues of marine bivalves at low temperatures. *Appl Environ Microbiol* **59**(6):1848–1854.
- [157] Oliver JD, Nilsson L, Kjelleberg S. 1991. Formation of nonculturable *Vibrio vulnificus* cells and its relationship to the starvation state. *Appl Environ Microbiol* **57**(9):2640-2644.
- [158] Oliveros JC. 2007-2015. Venny, an interactive tool for comparing lists with Venn's diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.
- [159] Osawa N, Mitsuhashi S. INFECTION OF GERMERE MICE WITH SHIGELLA FLEXNERI 3A. *Jpn J Exp Med* **34**:77–80.
- [160] Pacini F. 1854. Osservazioni microscopiche e deduzioni patologiche sul colera asiatico [Microscopic observations and pathological deductions on cholera]. *Gaz Med Ital* **6**:397-405.
- [161] Panicker G, Call DR, Krug MJ, Bej AK. 2004. Detection of Pathogenic *Vibrio* spp. in Shellfish by Using Multiplex PCR and DNA Microarrays. *Appl Environ Microbiol* **70**(12):7436-7444.
- [162] Paranjpye RN, Johnson AB, Baxter AE, Strom MS. 2007. Role of type IV pilins in persistence of *Vibrio vulnificus* in *Crassostrea virginica* oysters. *Appl Environ Microbiol* **73**(15):5041-5044.
- [163] Parker, B. 2015. Analysis of possible competitive advantage in *Vibrio parahaemolyticus* strains harboring newly-discovered bacteriophages Vf26 or Vf36. In partial fulfillment of Biochemistry, Molecular and Cellular Biology Honors Thesis at the University of New Hampshire, Durham, NH.
- [164] Parks DH, Tyson GW, Hugenholtz P, Beiko RG. 2014. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**(1):3123-3124.
- [165] Pearce MM, Zilliox MJ, Rosenfeld AB, Thomas-White KJ, Richter HE, Nager CW, Visco AG, Nygaard IE, Barber MD, Schaffer J, Moalli P, Sung VW, Smith AL, Rogers R, Nolen TL, Wallaca D, Meikle SF, Gai X, Wolfe AJ. 2015. The female urinary microbiome in urgency urinary incontinence. *Am J Obstet Gynecol* **213**(3):347.e1-347.e11.
- [166] Pfeffer C, Larsen S, Song J, Dong M, Besenbacher F, Meyer RL, Kjeldsen KU, Schriber L, Gorby YA, El-Naggar MY, Man Leung K, Schramm A, Risgaard-Petersen N, Nielsen LP. 2012. Filamentous bacteria transport electrons over centimetre distances. *Nature* **491**(218-221).
- [167] Pierce ML, Ward JE, Holoman BA, Zhao X, Hicks RE. 2016. The influence of site and season on the gut and pallial fluid microbial communities of the eastern oyster, *Crassostrea virginica* (Bivalvia, Ostreidae): community-level physiological profiling and genetic structure. *Hydrobiologia* **765**(1):97-113.

- [168] **Polz MP, Cavanaugh CM.** 1998. Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64**(10):3724-3730.
- [169] **Porwal S, Lal S, Cheema S, Kalia VC.** 2009. Phylogeny in aid of the present and novel microbial lineages: diversity in *Bacillus*. *PLoS One* **4**(2):e4438.
- [170] **Prusse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids R* **35**(21):7188-7196.
- [171] **Pujalte MJ, Ortigosa M, Maciá MC, Garay E.** 1999. Aerobic and facultative anaerobic heterotrophic bacteria associated to Mediterranean oysters and seawater. *Int Microbiol* **2**(4):259-266.
- [172] **Rambaut A.** 2012. *FigTree v1. 4: Tree figure drawing tool*. Available online at: <http://tree.bio.ed.ac.uk/software/figtree>.
- [173] **Rey MS, Miranfa JM, Aubourg S, Barros-Velàquez J.** 2012. Improved microbial and sensory quality of clams (*Venerupis rhomboideus*), oysters (*Ostrea edulis*), and mussels (*Mytilus galloprovincialis*) by refrigeration in a slurry ice packing system. *Int J Food Sci Tech* **47**(4):861-869.
- [174] **Rosche TM, Yano Y, Oliver JD.** 2013. A rapid and simple PCR analysis indicates there are two subgroups of *Vibrio vulnificus* which correlate with clinical or environmental isolation. *Microbiol Immunol* **49**(4):381-389.
- [175] **Saitou N, Nei M.** 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**(4):406-425.
- [176] **Sanger F, Nicklen S, Coulson AR.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**(12):5463-5467.
- [177] **Sanyal SC, Sen PC.** 1974. Human volunteer study on the pathogenicity of *Vibrio parahaemolyticus*, p227-230. In Fujino T, Sakaguchi G, Sakazaki R, Takeda Y (ed), International symposium on *Vibrio parahaemolyticus*. Saikon Publishing, Tokyo.
- [178] **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PJ.** 2011. Foodborne illness acquired in the United States – Major pathogens. *Emerg Infect Dis* **17**(1):7-15.
- [179] **Schloss PD, Gevers D, Westcott SL.** 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PloS ONE* **6**:e27310.

- [180] **Schloss PD, Westcott SL.** 2011. Assessing and improving methods used in operational taxonomic unit-based approaches and 16S rRNA gene sequence analysis. *Appl Environ Microbiol* **77**(10):3219-3226.
- [181] **Schloss PD, Westcott 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, Weber CF.** 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**(23):7537-7541.
- [182] **Schreier HJ, Schott EJ.** 2014. Draft genome sequence of the oyster larval probiotic bacterium *Vibrio* sp. strain OY15. *Genome Announc* **2**(5):e01006.14.
- [183] **Schrimer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C.** Insight into bias and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids R* **46**(6):e37.
- [184] **Schuster BM, Tyzik AL, Donner RA, Striplin MJ, Almagro-Moreno S, Jones SH, Cooper VS, Whistler CA.** 2011. Ecology and genetic structure of a northern temperate *Vibrio cholerae* population related to toxigenic isolates. *Appl Enviro Microbiol* **77**(21):7568-7575.
- [185] **Segata, N., J. Izard, L. Waldron, D. Gevers, L. Miropolsky, W.S. Garrett, and C. Huttenhower.** 2011. Metagenomic biomarker discovery and explanation. *Genome Biol* **12**(6):R60.
- [186] **Sheneman L, Evans J, Foster JA.** 2006. Clearcut: A fast implementation of relaxed neighbor joining. *Bioinformatics* **22**(22):2823-2824.
- [187] **Shumway SE.** 1996. Natural environmental factors. In: V.S. Kennedy, R.I.E. Newell and A.F.Eble, editors. *The Eastern Oyster Crassostrea virginica*. Maryland Sea Grant College, University of Maryland, College Park, Maryland. pp. 467-513.
- [188] **Sinha R, Abnet CC, White O, Knight R, and Huttenhower C.** 2015. The microbiome quality control project: baseline study design and future directions. *Genome Biol* **16**:276.
- [189] **Sobrinho P de S, Destro MT, Franco BDGM, Landgraf M.** 2010. Correlation between environmental factors and prevalence of *Vibrio parahaemolyticus* in oysters harvested in the southern coastal area of Sao Paulo State, Brazil. *Appl Environ Microbiol* **76**(4):1290-1293.
- [190] **Soizick LG, Robert A, Haifa M, Jacques LP.** 2013. Shellfish contamination by norovirus: strain selection based on ligand expression?. *Clin Virol* **41**(1):3-18.
- [191] **Srivastava M, Tucker MS, Gulig PA, Wright AC.** 2009. Phase variation, capsular polysaccharide, pilus, and flagella contribute to uptake of *Vibrio vulnificus* by the Eastern oyster (*Crassostrea virginica*). *Environ Microbiol* **11**(8):1934-1944.

- [192] **Stackebrandt E, Goebel BM.** 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Evol Microbiol* **44**(4):846-849.
- [193] **Staely C, Chase E, Harwood VJ.** 2013. Detection and differentiation of *Vibrio vulnificus* and *V. sinaloensis* in water and oysters of a Gulf of Mexico estuary. *Environ Microbiol* **15**(2):623-633.
- [194] **Star B, Haverkamp THA, Jentoft S, Jakobsen KS.** 2013. Next generation sequencing shows high variation of the intestinal microbial species composition in Atlantic cod caught at a single location. *BMC Microbiol* **13**:248.
- [195] **Suzuki MT, Giovannoni SJ.** 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62**(2):625-630.
- [196] **Szabo G, Preheim SP, Kauffman KM, David LA, Shapiro J, Alm EJ, and Polz MF.** 2012. Reproducibility of *Vibrionaceae* population structure in coastal bacterioplankton. *ISME J* **7**(3):509-519.
- [197] **Takaichi S, Maoka T, Sasikala C, Ramana CV, Shimada K.** 2011. Genus specific unusual carotenoids in Purple Bacteria, *Phaeospirillum* and *Roseospira*: structures and biosyntheses. *Curr Microbiol* **63**(1):75-80.
- [198] **Takemura AF, Chien DM, Polz MF.** 2014. Associations and dynamics of Vibrioaceae in the environment, from the genus to the population level. *Front Microbiol* **5**: 38.
- [199] **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**(12):2725-2729.
- [200] **Taylor RK, Miller VL, Furlong DB, Mekalanos JJ.** 1987. Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. *Proc Natl Acad Sci USA* **84**(9):2833-2837.
- [201] **Thomas IV JC, Wafula D, Chauhan A, Green SJ, Gragg R, Jagoe C.** 2009. A survey of deepwater horizon (DWH) oil-degrading bacteria from the Eastern oyster biome and its surrounding environment. *Front Microbiol* **5**:149.
- [202] **Trabal-Fernandez N, Mazon-Suastegui JM, Vazquez-Juarez R, Ascencio-Valle F, Romero J.** 2013. Changes in the composition and diversity of the bacterial microbiota associated with oysters (*Crassostrea corteziensis*, *Crassostrea gigas* and *Crassostrea sikamea*) during commercial production. *FEMS Microbiol Ecol* **88**:69-83.
- [203] **Trindade-Silva AE, Rua C, Silva GGZ, Dutilh BE, Moreira APB, Edwards RE, Hajdu E, Lobo-Hajdu G, Vasconcelos AT, Berlinck RGS, Thompson FL.** 2012. Taxonomic and functional microbial signatures of the endemic marine sponge *Arenosclera brasiliensis*. *PLoS ONE* **7**(7):e39905.

- [204] **Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM.** 2005. Comparative metagenomics of microbial communities. *Science* **308**(5721):554-557.
- [205] **Tritt A, Eisen JA, Facciotti MT, Darling AE.** 2012. An integrated pipeline for *de novo* assembly of microbial genomes. *PLoS ONE* **7**(9):e42304.
- [206] **Trucksis M, Galen JE, Michalski J, Fasano A, Kaper JB.** 1993. Accessory cholera enterotoxin (Ace), the third toxin of a *Vibrio cholera* virulence cassette. *Proc Natl Acad Sci USA* **90**(11):5267-5271.
- [207] **Tschape H.** 1994. Toxigenic factors of enteritis-Salmonellae. 2nd Asian Pacific Symposium on Typhoid Fever and other Salmonellosis. 47a. (Abstr.)
- [208] **Turner JW, Good B, Cole D, Lipp EK.** 2009. Plankton composition and environmental factors contribute to *Vibrio* seasonality. *ISME J* **3**(9):1082-1092.
- [209] **Turner JW, Paranjpye RN, Landis ED, Biryukov SV, González-Escalona N, Nilsson WB, Strom MS.** 2013. Population structure of clinical and environmental *Vibrio parahaemolyticus* from the Pacific Northwest coast of the United States. *PLoS One* **8**(2):e55726.
- [210] **Urquhart EA, Jones SH, Yu JW, Schuster BM, Marcinkiewicz AL, Whistler CA, Cooper VS.** 2015. Environmental conditions associated with elevated risk conditions for *Vibrio parahaemolyticus* in the Great Bay Estuary, NH. *PLoS One*. Accepted with revisions.
- [211] **Van de Peer Y, Chapelle S, De Wachter R.** 1996. A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Res* **24**(17):3381-3391.
- [212] **Van Rossum T, Pylatuk MM, Osachoff HL, Griffiths EJ, Lo R, Quach M, Palmer R, Lower N, Brinkman FSL, Kennedy CJ.** 2016. Microbiome analysis across a natural copper gradient at a proposed northern Canadian mine site. *Front Environ Sci* **3**:84.
- [213] **Velazquez-Roman J, León-Sicairos N, de Jesus Hernández-Díaz L, Canizalez-Roman A.** 2014. Pandemic *Vibrio parahaemolyticus* O3:K6 on the American continent. *Front Cell Infect* **3**:110.
- [214] **Vazquez-Baeza Y, Pirrung M, Gonzalez A, Knight R.** 2013. EMPeror: a tool for visualizing high-throughput microbial community data. *Gigascience* **2**(1):16.
- [215] **Vezzulli L, Colwell R, Pruzzo C.** 2013. Ocean warming and spread of pathogenic vibrios in the aquatic environment. *Microb Ecol* **65**(4):817-825.
- [216] **Vezzulli L, Pezzati E, Moreno M, Fabiano M, Pane L, Pruzzo C.** 2009. Benthic ecology of *Vibrio* spp. and pathogenic *Vibrio* spp. in a coastal Mediterranean environment (La Spezia Gulf, Italy). *Microbial Ecol* **58**(4):808-816.

- [217] **Volety AK, McCarthy SA, Tall BD, Curtis SK, Fisher WS, Gentner FJ.** 2001. Responses of oyster *Crassostrea virginica* hemocytes to environmental and clinical isolates of *Vibrio parahaemolyticus*. *Aquat Microb Ecol* **25**:11-20.
- [218] **Waldor MK, Mekalanos JJ.** 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**(5270):1910-1914.
- [219] **Ward DM, Weller R, Bateson MM.** 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **345**(6270):63-65.
- [220] **Ward JE, Shumway SE.** 2004. Separating the grain from the chaff: particle selection in suspension- and deposit-feeding bivalves. *J Exp Mar Biol Ecol* **300**:83-130.
- [221] **Warner E, Oliver JD.** 2007. Refined medium for direct isolation of *Vibrio vulnificus* from oyster tissue and seawater. *Appl Environ Microbiol* **73**(9):3098-3100.
- [222] **Wegner KM, Volkenborn N, Peter H, Eiler A.** 2013. Disturbance induced decoupling between host genetics and composition of the associated microbiome. *BMC Microbiol* **13**:252.
- [223] **Werner JJ, Koren O, Hugenholtz P, DeSantis TZ, Walters WA, Caporaso JG, Angenent LT, Knight R, Ley R.** 2012. Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. *ISME J* **6**:94-103.
- [224] **Whistler CA, Hall JA, Xu F, Ilyas S, Siwakoti P, Cooper VS, Jones S.** 2015. Use of whole-genome phylogeny and comparisons for development of a multiplex PCR assay to identify sequence type 36 *Vibrio parahaemolyticus*. *J Clin Microbiol* **53**(6):1864-1872
- [225] **WHO, FAO.** 2011. Risk assessment of *Vibrio parahaemolyticus* in seafood: interpretive summary and technical report. <http://www.fao.org/3/a-i2225e.pdf> . Accessed 5-Jan-16.
- [226] **Woese CR, Gutell R, Gupta R, Noller HF.** 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev* **47**(4):621-669.
- [227] **Wong HC, Wang P.** 2004. Induction of viable but nonculturable state in *Vibrio parahaemolyticus* and its susceptibility to environmental stresses. *Journ Appl Microbiol* **96**(2):359-366.
- [228] **Worden AZ, Seidel M, Smriga S, Wick, Malfatti AF, Bartlett D, Adam F.** 2006. Trophic regulation of *Vibrio cholerae* in coastal marine waters. *Environ Microbiol* **8**(1):21-29.
- [229] **Wu HJ, Sun LB, Li CB, Li ZZ, Zhang Z, Wen XB, Hu Z, Zhang YL, Li SK.** 2014. Enhancement of the immune response and protection against *Vibrio parahaemolyticus* by indigenous probiotic *Bacillus* strains in mud crab (*Scylla paramamosain*). *Fish Shellfish Immunol* **41**(2):156-162.

- [230] **Xu D, Wang Y, Sun L, Liu H, Li J.** 2013. Inhibitory activity of a novel antibacterial peptide AMPNT-6 from *Bacillus subtilis* against *Vibrio parahaemolyticus* in shrimp. *Food Cont* **30**(1):58-61.
- [231] **Xu F, Ilyas S, Hall JA, Jones SH, Cooper VS, Whistler CA.** 2015. Genetic characterization of clinical and environmental *Vibrio parahaemolyticus* from the Northeast USA reveals emerging resident and non-indigenous pathogen lineages. *Front Microbiol* **6**:272.
- [232] **Yeung PSM, Hayes MC, DePaola A, Kaysner CA, Kornstein L, Boor KJ.** 2002. Comparative phenotypic, molecular, and virulence characterization of *Vibrio parahaemolyticus* O3:K6 isolates. *Appl Environ Microbiol* **68**(6):2901-2909.
- [233] **Yu JW, Howell TL, Whistler CA, Cooper VS, Jones SH.** 2010. Oyster processing strategies for eliminating *Vibrio parahaemolyticus* and *Vibrio vulnificus* from New Hampshire oyster, abstr. Public Health and Related Factors Session Abstracts, Vibrios in the Environment 2010. Biloxi, Mississippi.
- [234] **Zabala B, García, Espejo RT.** 2009. Enhancement of UV light sensitivity of a *Vibrio parahaemolyticus* O3:K6 pandemic strain due to natural lysogenization by a telomeric phage. *Appl Environ Microbiol* **76**(6):1697-1702.
- [235] **Zurel D, Benayahu Y, Or A, Kovacs A, Gophna U.** 2011. Composition and dynamics of the gill microbiota of an invasive Indo-Pacific oyster in the eastern Mediterranean Sea. *Environ Microbiol* **13**(6):1467-1476.