Winter 2019

FILAMENTOUS BACTERIOPHAGE ASSOCIATED WITH SHAPING COMMUNITY STRUCTURE AND FITNESS OF INVASIVE VIBRIO PARAHAEMOLYTICUS ST36

Jillian Means
University of New Hampshire, Durham

FILAMENTOUS BACTERIOPHAGE ASSOCIATED WITH SHAPING COMMUNITY STRUCTURE AND FITNESS OF INVASIVE *VIBRIO PARAHAEMOLYTICUS* ST36

BY

JILLIAN MEANS

BS in Biology, Saint Michael's College, 2012

THESIS

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirements for the Degree of

Master of Science

in

Microbiology

December 2019

This thesis was examined and approved in partial fulfillment of the requirements for the degree of Master of Science in Microbiology by:

Thesis Director, Cheryl Whistler, Professor, Molecular, Cellular and Biomedical Sciences

Stephen Jones, Research Associate Professor, Natural Resources and the Environment

Loren Launen, Professor, Biology, Keene State College

On January 18th, 2019

Approval signatures are on file with the University of New Hampshire Graduate School

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER II

CHAPTER III

LIST OF FIGURES

INTRODUCTION

CHAPTER I

CHAPTER II

ABSTRACT

FILAMENTOUS BACTERIOPHAGE ASSOCIATED WITH SHAPING COMMUNITY STRUCTURE AND FITNESS OF INVASIVE *VIBRIO PARAHAEMOLYTICUS* ST36

By

Jillian Means

University of New Hampshire

Vibrio parahaemolyticus a ubiquitous coastal inhabitant, is the leading cause of bacterial seafood-borne illnesses in the United States. An increasing number of reported cases and rapid expansion into new areas has led to the classification of *V. parahaemolyticus* as an emergent pathogen. Most strains of *V. parahaemolyticus* are not virulent; however, the spread of virulent lineages from their native ranges to new locations has contributed drastically to the increase in vibriosis attributed to *V. parahaemolyticus* in recent years. In the United States (US), sequence type (ST) 36, a virulent strain endemic to the Pacific Northwest (PNW), spread from its native range up and down both coasts of North America even crossed the Atlantic to cause an outbreak in Spain in 2012, Specifically, the North Atlantic coast of the US traditionally did not have a major disease burden due to *V. parahaemolyticus;* however, the introduction of ST36 and the evolution of local pathogenic lineages have led to a sharp increase in the number of cases traced to product from this region.

Here we use genomics and phylogeographic analysis to examine the dynamics of the expansion of ST36 and its subsequent establishment in Northeast coastal waters. The impact of basal acquisition of two unique filamentous bacteriophages by distinct clonal clades within the Northeastern ST36 populations is also explored. We propose that the acquisition of these bacteriophages influenced the fitness of their hosts and enabled the establishment of robust local populations of pathogenic *V.*

parahaemolyticus, contributing greatly to the disease burden in the Northeast. Filamentous bacteriophages are distributed throughout many *V. parahaemolyticus* populations and may be important drivers of evolution amongst these strains. In direct competition under laboratory conditions, the bacteriophage associated with the Gulf of Maine clonal population, Vipa26, does not impact growth of persistently infected isolates and protects them from superinfection by similar phages. Upon new infection, the growth of susceptible isolates slows dramatically before the integration and down regulation of phage production. qPCR assays for integrated and replicative form of phage elucidate this dynamic during infection. This implicates Vipa26 as a potential sword and shield for this strain, possibly aiding the progenitor of the Gulf of Maine population of ST36 in its subsequent global expansion. Impact of phage on biofilm formation, resistance to predatory grazing and competitive fitness in natural seawater microcosms were also investigated. These studies indicate that phage integration is linked to environmental fitness of ST36 and further investigation into the phage-host relationship is warranted to shed light onto the dynamics of the establishment of novel *V. parahaemolyticus* populations.

INTRODUCTION

Jillian Means[1]

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

1. Justification

Food and water borne pathogens are a major concern for human health even in modern, developed nations. In the United States alone it is estimated that there are approximately 48 million cases of food borne illness, leading to about 3,000 deaths each year (1, 2). In addition to the heavy burden on human health, the economic cost of these diseases is immense, estimated between $51 billion to $77.7 billion annually in the USA (3). To combat this, national surveillance programs, food safety laws, wastewater management and clean drinking water are vital tools. Developing nations with less infrastructure and education devoted to curbing these diseases face a much higher burden on health and their economy. The World Health Organization (WHO) estimates that 31 major causes of foodborne disease lead to about 600 million cases of disease globally, with the majority of the burden on less developed regions (4). Understanding the ecology and virulence of these pathogens is important for informing management, prevention and treatment to reduce the severity of the impact these diseases have on human health.

2. Vibrios

Vibrios are a diverse aquatic group of halophilic bacteria, with over 100 species widely distributed through brackish and marine environments (5). They are Gram-negative facultative anaerobes, characterized by a curved rod shape, lateral flagella and a polar flagellum (6). Unusual in other bacteria, vibrios often contain two complete chromosomes, generally with one significantly larger

than the other (7, 8). Vibrios form diverse associations with other marine life, some such as the fish pathogen *Vibrio salmonicida* cause disease (9), whereas *Vibrio fischeri* form mutualistic symbioses with squid and fish species. Although most vibrios are harmless to humans, approximately twelve species are capable of causing disease. In the United States the three most common causes of vibriosis are *Vibrio parahaemolyticus*, *Vibrio vulnificus* and *Vibrio alginolyticus* (10); however, globally the vibrio species leading to the greatest health crisis is *Vibrio cholerae*, the pathogen responsible for deadly cholera outbreaks.

Cholera is a disease characterized by severe diarrhea leading to rapid dehydration and death if left untreated. According to the WHO, cholera affects around 1.3 – 4 million people every year, with estimates of fatalities between 21,000 and 143,000 (11). This devastating disease is caused by *V. cholerae*, a natural inhabitant of coastal and brackish water worldwide. Despite the abundance of cases of cholera reported, all are caused by only two of the many serotypes, O1 and O139. These are designated toxigenic as they carry a prophage, CTXϕ, which encodes the main cholera toxins, *ctxA* and *ctxB* (12). Lysogenic conversion of strains newly infected with CTXϕ can give rise to new toxigenic lineages. There have been six historical cholera epidemics with a seventh ongoing since 1961 (13). The severe diarrhea caused by cholera releases quantities of up to 10^7 shed *V. cholerae* per stool back into the environment, continuing the infective cycle when others consume the contaminated water (14, 15). Modern sanitation and access to clean drinking water have largely solved the issue of cholera in developed countries. Cholera is also a very treatable disease with access to proper medical facilities. Despite this, it remains a major burden in impoverished regions which lack access to these modern amenities, and in regions recently hit by a natural or man-made disaster, disrupting normal functioning of such facilities and allowing cholera to spread (16).

In the United States, cholera has not been a major issue since the advent of modern water safety systems and wastewater treatment plants. However, vibriosis is still a major source of disease

and economic loss in the USA, usually as a foodborne infection but in rare cases through contact with contaminated seawater. One of the less common but most severe causative agents of infection is *V. vulnificus*. *V. vulnificus* is a widely distributed inhabitant of warm brackish and coastal waters and like many vibrios, its abundance is closely linked with water temperatures, with highest concentrations found in warm waters. Though *V. vulnificus* infections most commonly cause gastroenteritis linked to consumption of oysters harvested in the warm waters of the Gulf of Mexico (17), more famously *V. vulnificus* can cause severe and often fatal wound infections. Despite its low clinical prevalence with only approximately 100 cases annually in the United States, the severity of wound infection leads to a 91% hospitalization rate (1). These skin infections are most common in people who are immune compromised and have an open wound exposed to brackish or salt water (17, 18). Amputation and intensive care are frequent treatments, despite such aggressive measures mortality associated with *V. vulnificus* infections is around 35% (1).

V. alginolyticus, which is the second most prevalent cause of vibriosis reported in the United States since 2007, when it overtook *V. vulnificus*, is also endemic to coastal waters around the world (19). This species is also a noted pathogen of fish and shellfish leading to major losses in aquaculture (20). Unlike many other vibrios, *V. alginolyticus* is usually not a foodborne pathogen, instead skin or ear infection caused by contact with seawater are the most common infections (19, 21). Unlike *V. vulnificus* infections which are usually comorbid with other conditions, 74% of patients did not have any pre-existing conditions, and less hospitalization was reported especially for ear infections and lower extremity skin infections (1 and, 11% respectively) (19). Despite fairly low rates of hospitalization and mortality, *V. alginotlyicus* infections can rarely lead to serious treatments such as amputation or surgical intervention (19). Although *V. vulnificus* and *V. alginolyticus* infections are serious, the majority of the burden of vibriosis in the USA is due to *V. parahaemolyticus*.

A ubiquitous inhabitant of marine and estuarine environments, *V. parahaemolyticus*, the focus of this work, is the most common cause of vibriosis in the USA and a leading cause of foodborne gastroenteritis causing an estimated 45,000 cases nationally each year (10). *V. parahaemolyticus* infection is usually due to consuming raw or undercooked shellfish, although wound infections after exposure to water also occur (1). First defined in Japan in 1950 after causing an outbreak of seafood-borne illness with 272 cases reported and 20 deaths, *V. parahaemolyticus* has since become one of the leading causes of seafood-borne illness (10, 22, 23).

3. Vibrio parahaemolyticus ecology

An environmental pathogen, *V. parahaemolyticus* is widespread in tropical and temperate estuaries, and coastal waters and can tolerate a broad range of conditions (24, 25). In the Great Bay in New Hampshire, isolates have been collected from the water column even in temperatures as low as 1°C (26). Under optimal conditions in the laboratory, *V. parahaemolyticus* grows extremely quickly, able to double in <10 minutes (27). A strong relationship between *V. parahaemolyticus* abundance and temperature and salinity drives a seasonal cycle with peak abundance in the summer months reaching concentrations of 10^2CFU/100mL in the water column (28–31). However, as the temperature decreases, so does the abundance and genetic diversity of the strains in the water (26). *V. parahaemolyticus* enters a viable but non culturable (VBNC) state when faced with difficult conditions including low temperature, and most strains persist in sediment over the winter months; proliferating back into the water column as temperatures warm (26, 28, 32, 33).

V. parahaemolyticus can be found free floating in the water column, in the sediment or living in association with, or as a pathogen of, marine life such as corals, plankton, shellfish and fish (29, 34–37). It attaches readily to many surfaces and *V. parahaemolyticus* abundance is often associated with high levels of plankton and chlorophyll-*a,* in addition to salinity and temperature (29, 33). The natural ability to use many substrates for energy and nitrogen enables a wide range of possible habitats. It

accumulates and becomes particularly concentrated in shellfish, including oysters, relative to the surrounding water.

V. parahaemolyticus naturally transforms exogenous DNA, facilitating horizontal gene transfer(HGT) (38)**.** Several vibrio species including *V. cholerae* are competent in the presence of chitin; however, the dependence on chitin in *V. parahaemolyticus* may be strain dependent based on mixed reports (39–41). Although not all strains of *V. parahaemolyticus* are pathogenic, due to this and other mechanisms of HGT such as transduction or conjugation, new pathogenic lineages can arise from the transfer of Vibrio Pathogenicity Islands (VPaIs), to a previously non-disease causing environmental strain (42).

4. Vibrio parahaemolyticus pathogenesis

4.1 Clinical presentation and reporting

Human pathogenic *V. parahaemolyticus* typically causes self-limiting gastroenteritis linked to the consumption of raw or undercooked seafood. Symptoms generally include diarrhea, abdominal pain, nausea, vomiting, fever and chills (10). However, occasionally in immune compromised or elderly individuals, *V. parahaemolyticus* can lead to serious complications and, rarely, death (1). *V. parahaemolyticus* is also capable of infecting open wounds that occur in or are exposed to seawater (1). Classified as an emergent disease, it is one of the only foodborne pathogens which is increasing in abundance in the US over the last decade despite increasing management strategies, causing it to be a significant economic and health concern (43). Early surveillance of vibriosis including *V. parahaemolyticus* was dictated by individual states with some requiring clinical cases to be reported while others did not track clinical cases. However, in 2007 it became mandatory at a national level to report cases of vibriosis to the Cholera and Other Vibrio Illness Surveillance (COVIS) system. Information such as clinical data, potential sources of exposure to the patient, and potential sources of

contaminated food are collected where possible (44). Each region still has specific guidelines to govern the management of oyster harvesting, including potential closures under certain conditions.

Despite mandated reporting measures, illnesses caused by *V. parahaemolyticus* are severely underreported and occasionally misdiagnosed (45). As it is usually self-limiting, resolving in around 72 hours, those sickened will often not go to a hospital but self-care in their own homes. Even if they do, it is possible that the attending physician will not order a test to determine the infectious agent, and even if a test is ordered, not all labs are able to test for vibrios (45). One study estimated that in the case of underreporting for *V. parahaemolyticus*, a corrective factor of around 20 must be applied to calculate the true number of illnesses (46). One issue facing public health officials is tracking levels of virulent *V. parahaemolyticus* compared to the total abundance, as well as modeling conditions which are favorable for rapid expansion of the *V. parahaemolyticus* population (47). Currently, shellfish harvesting is only closed after a certain number of illnesses have been traced back to a harvest area within a defined time period.

4.2 *V. parahaemolyticus* virulence determinants

Most *V. parahaemolyticus* strains are not pathogenic and not all virulence factors are known in *V. parahaemolyticus*. Two of the major known virulence factors, thermostable direct hemolysin (*tdh)* and tdh-related hemolysin (*trh*), are sometimes used as genetic markers for the presence of pathogenic *V. parahaemolyticus* (48, 49). However, around 10% of clinical isolates exhibit neither of these classic virulence markers, increasing the difficulty of proactively monitoring danger levels (50, 51). Environmental and pathogenic strains alike may exhibit one, both or neither of these genes, with up to 48% of environmental isolates containing either *tdh* or *trh* in certain locations (52).

The TDH protein is a cytotoxic and hemolytic pore-forming tetramer, leading to the release of water and ions from the cell and eventually cell death (53, 54). TRH is 70% homologous to TDH and also

alters ion flow in the mammalian cells (55, 56). Despite being considered a marker for virulence, the deletion of *tdh* only reduced, but did not eliminate the ability of the virulent sequence type (ST) 3 *V. parahaemolyticus* to cause fluid to accumulate in the intestines of infected rabbits, the premier model for gastrointestinal vibriosis (57, 58). Pathogenic strains often have genetic islands (VPaIs) containing *tdh* and *trh* which are horizontally transferred between strains (42, 59, 60). Certain VPaIs contain Type 3 Secretion Systems (T3SS) in conjunction with these hemolysins (42, 61, 62). Other pathogenicity determinants including T3SS2 effectors, *vopV* and *vopZ,* have been identified, illustrating the complexity of virulence in *V. parahaemolyticus* (63, 64).

T3SSs are needle like structures able to deliver effectors directly into eukaryotic cells. All *V. parahaemolyticus* strains harbor a T3SS on chromosome I designated T3SS1, capable of cytotoxicity but not important for conferring virulence (65). Many pathogenic strains contain a second accessory T3SS called T3SS2 with three currently described variations occurring as part of a VPaI with either *tdh* (T3SS2α), *trh* (T3SS2β) or both (T3SS2γ) (42, 61, 62). It has been suggested that the T3SS2α island, containing two copies of *tdh* and no *trh,* may have increased resistance to predatory grazing by eukaryotes compared to strains which lack this island (65, 66). This fitness advantage may explain the persistence of virulence trait in the environment, given the general lack of significant selection for the human as a host. The mechanisms for virulence in *V. parahaemolyticus* are complex and not yet fully understood, with many processes including adhesion, biofilm formation, siderophore production and extracellular proteases important for colonizing a host and eliciting an enterotoxic response.

4.4 Epidemiology of disease

Prior to 1996, outbreaks were rare and *V. parahaemolyticus* infection was sporadic and caused by dominant local variants restricted in geographic range (67, 68). In 1996, ST3 (serotype O3:K6) clonal complex arose in Southeast Asia, and by 2006 this complex had become pandemic causing illness in

Europe, Asia, Africa and the Americas (69). ST3 is now the only pandemic clonal complex of *V. parahaemolyticus* and is a leading cause of *V. parahaemolyticus* infections globally. This marked a shift in the epidemiology of *V. parahaemolyticus*. The rapid spread of the highly virulent ST3 illustrated the mobile nature of *V. parahaemolyticus* and highlighted potential mechanisms which could introduce it to new areas. Natural currents could carry it to new shores, either free floating, or more likely associated with plankton (70, 71). It is also likely that the greater globalization of the seafood market and well as expanding shipping routes enabled the bacteria to hitch rides to new regions (72, 73). One potential factor contributing to its heightened virulence and fitness that could contribute to the many outbreaks and spread of ST3 is a filamentous prophage, f237, which encodes a putative adhesive protein and may facilitate attachment to the host during infection (74). Since the spread of ST3 and its continued presence as a pathogen on the global scale, another sequence type, ST36 has also begun to spread widely from its native region of the Pacific Northwest establishing populations down into California as well as up and down the East Coast of the US and causing an outbreak in Spain (75–78).

5. *Vibrio parahaemolyticus* ST36 expansion

ST36 (serotype O4:K12 and O4:K(unknown)) is a highly virulent strain of *V. parahaemolyticus* originating from the Pacific Northwest (PNW) where it has been causing illnesses since the 1980s (76, 79). It is *trh* and *tdh* positive and contains a T3SS2γ, and is found with high clinical prevalence (75). It was introduced into the Northeast US waters prior to 2012 by some unknown mechanism, and has since established robust local populations in the Gulf of Maine (GOM) and Long Island Sound (LIS) (Fig. 1.1, 77, 80). In 2012, ST36 caused an outbreak in Galicia, Spain and was detected on the East Coast of the US (Fig. 1.1, 76, 78). Then in 2013, it caused a major outbreak traced to Northeast product with 104 cases reported from multiple states (75).

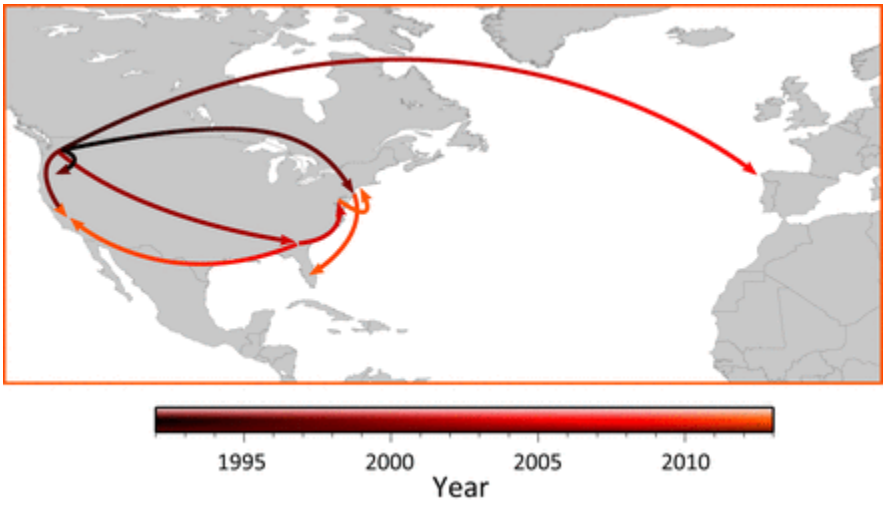

Figure 1.1. Proposed spatial phylogenetic reconstruction of ST36 expansion. Martinez-Urtaza *et al.* used BEAST and SPREAD to produce a model of the temporal spatial expansion of ST36 out of its native Pacific Northwest (76).

Since then ST36 has managed to persist in the Northeast (75, 80). Besides a similar outbreak that was caused by ST3 in 1998 and traced to New York product, *V. parahaemolyticus* infections from this area were quite rare and not cause for much concern (69, 81). Warming water globally due to climate change is a potential factor contributing to the spread of ST36, especially in the Gulf of Maine where sea surface temperatures have increased nearly 2°C since pre-1958 averages (82, 83). Warm water temperatures are closely correlated to *V. parahaemolyticus* abundance, increasing the likelihood of human exposure to elevated dosages (28, 31). Since the outbreak, ST36 has accounted for the majority of *V. parahaemolyticus* infections in the Northeast, making it a significant concern to the shellfish industry (80).

6. Bacteriophages

Bacteriophages are viruses specific to bacteria and archaea. Discovered in the early 20th century, they were first used as a method to cure or help prevent bacterial infections prior to the widespread usage of antibiotics (84, 85). The first paper on phage therapy was published in 1921 in Belgium, reporting excellent results in injecting staphylococcus-specific phage preparations into cutaneous boils

(86). Despite initial early success, the trials were often derided for being inconsistent and the results exaggerated. This controversy, combined with the rise of easily accessible and effective antibiotics in the aftermath of WWI, lead to phage therapy being largely abandoned for the greater part of the 20th century (87). In recent years the advent of multiple antibiotic resistant bacteria and the shortage of novel antibiotics to address these infections has led to renewed interest in bacteriophages (88, 89). Concurrently, they are also being recognized for having potentially huge, but often unaccounted for ecological roles, as the most abundant biological entities on earth with an estimated 10^31 on earth, 10 times more than their prokaryotic hosts (89).

6.1 Bacteriophage lifecycles

Phages are traditionally classified into two main lifecycles, lytic and lysogenic. The T4 *Escherichia coli* bacteriophage is a classic model of a lytic phage (90). The phage attaches, or adsorbs to the host, injects its genetic material into the cell and replicates in the cytoplasm, hijacking host cell machinery to build new phage particles (91). The new phage genomes are then packaged into the icosahedral heads, the whole virion is assembled, and the bacteria is lysed, releasing the progeny phage into the surroundings. The successful completion of the lytic cycle always leads to the lysis and death of the host cell. In the lysogenic cycle, exemplified by the *E. coli* specific lambda phage, the phage integrates its genetic material into the host cells genome after adsorption. The phage genome then transmits vertically with the cell as it replicates, until some trigger leads to the phage entering a lytic cycle of replication and release. Lysogenic phages in their integrated forms are known as prophages, and many bacterial genomes are littered with both active and inactive prophages.

Chronic temperate phages do not fit neatly into either the lysogenic or lytic lifecycles. These phages, including filamentous phages of the class Inovirus such as f237 in *V. parahaemolyticus* ST3, exhibit a long term infection of the host cell (74). Chronic infection is somewhat similar to the lysogenic

cycle, as the phage genome is often integrated as a prophage into the host cell chromosome. However, once integration has occurred, the phage is able to replicate with in the host cell without excising from the chromosome leaving a copy of the phage genome in the host chromosome. The virions then secrete from the host cell, usually without killing the cell. The infection can result in a slowing of growth as the phages use up valuable host cell resources even when the host is not lysed to release the progeny phage (92).

7. Bacteriophage impact on bacterial ecology

As the most abundant lifeform on the planet (93, 94), phages are incredibly important in impacting the ecology of the bacteria they infect. They have long been implicated in transfer of bacterial genes between species through transduction, as well as agents of population control (95, 96). More recent work has uncovered more ways in which phages influence bacterial ecology and host genomics. Microbial diversity can be maintained through transduction and the introduction of novel phage genes (97, 98). Some bacterial strains have been found with over 30% of the genome made up of phages and phage remnants (99, 100). Phages are also capable of changing the dynamics of direct bacterial competition in many ways such as differential predation, introduction of novel genetic content and placing selective pressure on populations (101, 102).

7.1 Phage distribution

Phages are believed to exist everywhere that bacteria can be found. In the gut of animals, coliphages – phages specific to coliform bacteria – are found associated with the microbiota, including 34% of humans, with much higher concentrations associated with ruminants (103, 104). This is likely due to the necessity of high bacterial loads for digesting their cellulose rich diets. In the environment, phages have been found at titers as high as 10^7 particles/gram of soil in agricultural fields (105). Marine sediment has one of the highest phage titers at 10^9 particles/gram (106). This figure is relatively higher

than in the water column above, at around 10^7 particles/mL in estuarine systems (107). Phage concentrations are generally 10 times higher than the associated bacterial load.

In some systems, phages are thought to ensure bacterial community diversity through the "kill the winner" model, which suggests that phages specific to the most successful strain in a community will proliferate, leading to a decrease in the population of the "winner", the strain that is otherwise most fit, allowing less competitive strains to survive (101, 108). This would kill the most abundant strain until the infective cycle was stopped due to a low abundance of susceptible bacteria, leading to cyclic boom/bust cycles in specific phage populations. This gives the more phage-defense oriented strains access to valuable resources and diversifies the overall community structure.

7.2 Phages in the aquatic environment

Phages in marine systems impact their hosts in a variety of ways at both the population and community level. Introduction of phages to a previously naïve population leads to the selection of resistant clones; in one study the overall density of the resulting culture was not affected, only the genetic composition (109). In this context, phages are drivers of evolution acting as a direct selective pressure on certain populations and effecting their genetic makeup and the overall structure of marine communities. The interplay between environmental cholera phages and cholera epidemics highlights this. During seasonal epidemics of cholera, cholera phages were rarely detected, with the inverse true during periods with few incidences of cholera reported (110, 111). Including cholera phage abundance is important when developing models of cholera epidemic severity and length (111). *V. cholerae* introduced into naïve regions which may not contain high abundance of their natural phage counterparts could also lead to severe epidemics that are not seasonally curbed by the increase in lytic phage in the environments (110). Interestingly, the lytic phages known to curb epidemics are selected

for in the host gut and may lead to increased predation in the environment after a period of many infections (112).

7.3 Bacteria vs phage arms race

This interplay between phage and bacteria has given rise to many varied and ingenious measures to protect bacteria from infection. This phage-bacteria arms race drives the evolution and ecology of both players. Bacteria have many weapons in their arsenal to protect against phage infection, preventing adsorption, blocking the injection of foreign DNA, abortive infection and restriction modification of foreign DNA found inside the cell. During adsorption the phages target a specific receptor on the target cell, so changes to this receptor or blocking it through production of extracellular matrix or competitive inhibitors will interfere with this step and neatly prevent the phage from infecting the cell (113). The plasticity of the phage genome leads to rapid evolution to overcome such barriers, thus causing a race between the bacteria and the phage.

A notable immunity system developed by bacteria is the clustered, regularly interspaced short palindromic repeats (CRISPR) - CRISPR associated (Cas) system, recently adapted as a powerful gene editing tool (114, 115). This type of system is found in about 40% of bacterial species, and is an adaptive way for the cell to respond to the incursion of foreign DNA in the form of phage genomes or plasmids. (116) The cell is able to recognize and selectively cut the foreign DNA, protecting itself from infection. CRISPR-Cas systems work alongside the more ubiquitous (found in ~90% of bacterial species) restriction modification (R-M) systems (117). Native restriction enzymes in the bacteria cut any foreign DNA using DNA modification differences to differentiate between host and alien DNA (117, 118).

Bacteria that are lysogenized by certain phage will also be protected by that phage from infection by similar phages. Many forms of lysogenic phage prevent the superinfection of their lysogenized hosts, protecting the infected bacteria from potentially greater harm from a more

aggressive phage by down regulating phage production (119). This relationship benefits the phage as well because it keeps the lysogen alive until a trigger to induce the phage, as well as interfering with competition from other phages.

Abortive infection is another well understood phage resistance mechanism. It has been described mostly on plasmids found in *Lactococci*, a group of bacteria important in cheesemaking where phage related cell death is a dire concern (120). This is an altruistic method of sacrificing the individual to save the population. The incoming phage triggers the abortive infection operon, leading to host cell death and the release of nonviable phage progeny (120, 121). These diverse systems are varied in their targets when activated, capable of acting on transcription, translation, replication, but are similar in their lethality to the cell.

7.4 Mutualistic coevolution of phage and host

Phage-host dynamics may seem mostly antagonistic at a glance; however, a deeper look will reveal a more complex relationship. Individual cells may benefit from harboring prophages, the novel genetic content of the phage giving the bacteria an edge in competition with other strains. In *E. coli*, phages can increase the survival of the cell after a severe stress through the upregulation of the stress response, in *Salmonella* species phages contribute directly to the gain of nitrate metabolism through a novel nitric oxide synthase encoded in the phage genome (122–124). Population level dynamics are even more complex, where phage mediated cell death is crucial to biofilm formation in several species (125, 126), whereas in others phage induced lysis releases bacteriocins, significantly boosting the competitive fitness of the population when in competition with susceptible strains (127). The balance between the bacterial species and their phages is intricate, covering the spectrum between true antagonists engaging in a never-ending arms race and a mutualistic relationship enhancing the fitness of the bacterial host.

7.5 Lysogenic conversion

Lysogenic conversion is another process of coevolution between the phage and the host. The prophage in this case confers a phenotype to the host cell beyond the normal effects expected from phage infection due to the presence of genes carried on the phage genome which are not involved in the phage lifecycle. These genes vary widely, from stress response regulators to antibiotic resistance, novel metabolic systems and most notoriously, virulence (128, 129). Prophage encoded virulence factors are found across many species, including toxins in *Corynebacterium diphtheriae,* and *Clostridium botulinum* the causative agents of diphtheria and botulism, respectively (130–132). The deletion or excision of the prophages attenuates the virulence of these strains (132). Although toxins are important for pathogen evolution due to lysogenic conversion, many less direct virulence factors that are equally crucial to pathogenesis are found on phage genomes, for example adhesins crucial for attachment and colonization of the host (133).

8. Phage classification - Inoviridae

Filamentous phages of the family Inoviridae are found across the spectrum of bacterial species, with the majority occurring in Gram negative species. They can vary greatly in genetic content; however, they maintain a similar core genome arrangement (Fig. 1.2). The coat of the phage is comprised of many repeating copies of a single protein down the length of the particle, with differing protein content in the head and tail structure (134, 135). They contain circular ssDNA genomes and are non-enveloped. These filamentous phage particles tend to have dimensions of around 7nm in width and 1000-2000nm in length (135). Filamentous phages can either be lysogenic or replicate solely in the cytoplasm; however, they do not follow a typical lytic life cycle (134). Instead the phage particle is extruded at a certain rate from the host cell without necessarily killing the cell (Fig. 1.2). Doing so does enact a significant drain on cellular resources and they are known to slow the growth of host cells.

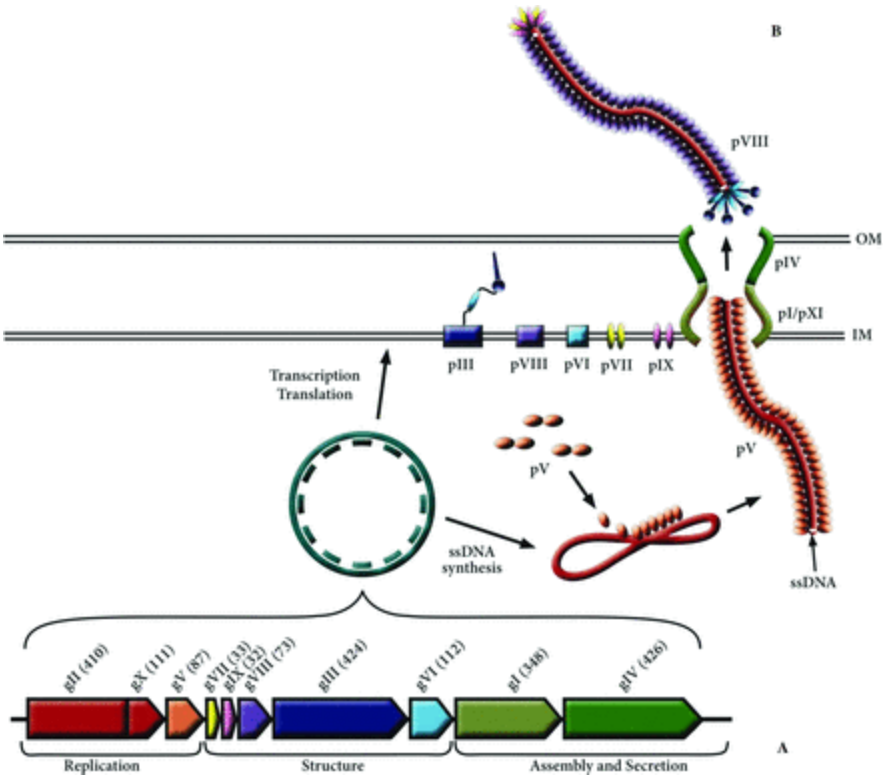

Figure 1.2. Genome organization, replication and assembly of Inovirus. A. Colors of genes are grouped and labelled based on function. **B.** Transcription/translation of the circular replicative form and assembly of the phage particle. Colors of proteins match the corresponding open reading frames in A (92).

9. Direct effects of select Inoviridae on hosts

9.1 Pf *Pseudomonas aeruginosa* phages

These filamentous phages are receiving attention for the impact they can have on the host bacteria. Some encode toxins, whereas others impact host motility and biofilm formation. For example the Pf phages of *Pseudomonas aeruginosa,* a medically important opportunistic pathogen, were found to stabilize the biofilm matrix and contribute to the formation and stability of the biofilm (136). The filamentous structure of the phage itself worked to strengthen the biofilm, while cell death increased due to superinfection, a condition frequently deadly to the host. Cell death is an important part of biofilm formation in Pseudomonas (137), and the extracellular DNA released is a crucial component of the biofilm matrix (138). *P. aeruginosa* causes difficult to treat infections in immune compromised

patients in part due to its ability to form dense biofilms in surgical wounds or the lungs of cystic fibrosis patients. These biofilms are largely antibiotic resistant, seriously narrowing potential treatment options. This is only one of many documented cases of phages being important players in the establishment of biofilms (125, 139).

9.2 *Shewanella piezotolerans* WP3 phage transcriptional regulation

Shewanella piezotolerans WP3 has decreased swarming motility after infection by the SW1 filamentous phage (140, 141). Lateral flagellar proteins are downregulated in part by phage encoded transcriptional regulators, and deletion of the phage leads to higher production of the flagella. The phage impacted the expression of many host genes, a majority of which were involved in the production of lateral flagella (140). The M13 coliphage, likely the most studied of the filamentous phages, also regulates the transcription of host cell genes during lysogeny, using a native phage transcriptional regulator (142). These regulatory systems can have crucial impacts on the fitness of the host cell and many indirect effects may result from lysogeny by this phage.

9.3 Virulence effects of filamentous phages

9.3.1 CTXϕ *Vibrio cholerae* phage

Within the *Vibrio* genus there is an even more medically important filamentous phage, CTX found in *Vibrio cholerae,* the causative agent of toxigenic conversion in cholera (143). The phage integrates into the *dif* site of the genome and encodes several of the major toxins which lead to severe illness, CtxA and CtxB, as well as Ace and Zot (12). These toxins act on the epithelial cells of the human intestine to cause severe loss of fluids which can lead to dehydration and death if left untreated (11). The presence of a phage encoding the principle toxins that cause cholera means virulence can be spread horizontally through the population, leading to outbreaks. The phage can infect and convert previously

non-toxigenic environmental isolates (144). However, phage mediated effects can also attenuate an epidemic due to the induction of phage after infection by a satellite phage RS1 (145).

9.3.2 RSS1 and RSM1 *Ralstonia* phages

Sometimes different phages can have drastically differing impacts on their host's virulence. For example, in *Ralstonia solanacearum*, a soil inhabiting plant pathogen, there are two well studied filamentous phages impacting its virulence. RSS1 enhances host virulence through a multitude of transcriptional regulatory changes to the host cell, including early induction of the global virulence regulator *phcA* (146, 147). On the other hand, RSM1, another somewhat distantly related filamentous phage actually has the opposite impact, instead reducing host virulence (146, 148). The interplay of bacteria and the phages determines the behavior and pathogenicity of the lysogen.

10. Inoviridae in *V. parahaemolyticus*

In *V. parahaemolyticus*, the filamentous phage f237 is notably nearly universally associated with the pandemic complex ST3 (74, 149). This phage was found to integrate site specifically into the host chromosome targeting the *dif* site, an important region for dimer resolution (150). CTXϕ contains similar genomic structure and also integrates into the *dif* site of *V. cholerae*, indicating that this is a common target of integration for filamentous prophages in vibrios (150). Although the core architecture of the f237 phage in *V. parahaemolyticus* is similar CTXϕ, the cholera toxins *ctxA* and *B,* are located in the variable region which lack any structural or sequence homology to f237 (149). ORF8 is instead encoded by f237 in this region and shares motifs with an adhesin protein found in Drosophila, leading to speculation that the presence of ORF8 could increase attachment of ST3 in the human gut contributing to colonization and disease (74). However, the true impact of f237 on the success and virulence of ST3 remains unclear, with some clinical isolates lacking phage reported (151).

11. Research Objectives

11.1 Chapter 2: Do genomic and phage content of *V. parahaemolyticus* strains correlate to their geographic region of origin?

The introduction and persistence of *V. parahaemolyticus* ST36 along the Atlantic coast has led to serious increases in vibriosis throughout the region. Elucidating the spread and population structure of ST36 incursion to this region is important for understanding the second inter-oceanic expansion of *V. parahaemolyticus* strains. Here we use a diverse library of clinical and environmental strains with robust traceback data spanning the period prior to the invasion of ST36 through the present. In this study we hypothesize that whole genome phylogeny, accurate traceback and phage content are crucial to understanding the ecological context of trans-oceanic expansion and uncover distinct sub-populations of ST36 relating directly to region of origin.

11.2 Chapter 3: What is the impact of novel inovirus phage on the ecology and virulence of strains from clonal sub-populations of invasive ST36?

ST36 is the second pathogenic sequence type of *V. parahaemolyticus* since ST3 to spread so far beyond its traditional range and establish long-term stable populations. Analysis of ST36 genomes reveals the presence of filamentous prophages associated with distinct regional subpopulations. These phages are very similar to f237 found universally in ST3 and are classed in a family known for diverse impacts on the host cell function. Phages are drivers of microbial ecology and evolution, so the presence of these phages suggests that they could be influencing ST3 and ST36 subpopulations. Due to the success and pathogenicity of both of these phage harboring sequence types, in this study we investigate the potential effect of this phage on the host cell using molecular and culture based techniques. We hypothesize that the acquisition of the phage provides a demonstrable fitness or virulence benefit to the host compared to phage deficient isolates.

CHAPTER ONE

Using phylogenetic relationships paired with traceback data and phage content to elucidate the invasion of *Vibrio parahaemolyticus* Sequence Type 36 into the United States North Atlantic Coast

Jillian Means[2], Ashley L. Marcinkiewicz[4], Randi L. Foxall[1,2], Stephen H. Jones[1,2], Vaughn S. Cooper[3], 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]Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[4]Division of Infectious Disease, Wadsworth Center, New York State Department of Health, Albany, NY

Abstract

Vibrio parahaemolyticus is a ubiquitous resident in coastal seawater bacterial communities. Most strains are harmless; however, *V. parahaemolyticus* is still the most common cause of bacterial seafood-borne gastroenteritis in the United States. In recent years, the epidemiology of *V. parahaemolyticus* outbreaks has shifted with the emergence and spread of virulent sequence types (ST) across the globe. ST3, the only pandemic clonal complex of pathogenic *V. parahaemolyticus,* is not a major health concern in the United States because it has not established robust local populations, including in the Northeast, where *V. parahaemolyticus* infections are not traditionally common. A strain endemic to the Pacific Northwest, ST36, has since spread into the Atlantic and causes disease up and down the East Coast, as well as one outbreak in Spain in 2012. Martinez *et al.* previously described this expansion; however, here we leverage our unique collection of environmental and clinical strains with robust traceback data to the geographic location of origin to further elucidate the invasion of ST36 into the waters of the Northeast. The current understanding of ST36 population dynamics and spread is based off of location

of reported illness, without accounting for the movement of shellfish and people. Here we elucidate the population structure and unique phage content of the Northeast ST36 population and relate it to the body of water from which each strain originated, providing unprecedented insight into movement and evolution of ST36 through its natural environment as it expands far beyond its native range. We find that much of the diversity of this lineage arose in the PNW and was subsequently introduced into new habitats, potentially influenced by phage acquisition or loss. Investigating the spread of pre-pandemic ST36 furthers our understanding of the virulent *V. parahaemolyticus* lineages and their evolution and expansion.

Introduction

Cases of *V. parahaemolyticus*, an emergent pathogen and the leading cause of bacterial seafood-borne gastroenteritis, have increased in the Northeast of the United States over the last decade (33, 80, 82). Although a global increase in disease is in part due to the rise of a pandemic clonal complex (sequence type 3 or ST3) in Asia in 1996 and subsequent world-wide spread by 2006, this strain is not prevalent and does not contribute greatly to the rise in cases in the Northeast US region (69, 80). Warming water temperatures, a growing aquaculture industry, the emergence of new pathogenic lineages and expansion of highly virulent strains are all drivers of this trend (42, 77, 82). The introduction of ST36, a virulent lineage endemic to the Pacific Northwest, is particularly important to the shift in *V. parahaemolyticus* epidemiology in the region. Originally limited to Pacific Northwest where it began causing infections in 1979, ST36 began to spread beyond its traditional range into California in the early 2000s (68, 76). Subsequently in 2012, ST36 was traced to product from the Long Island Sound (LIS) and caused the largest outbreak of *V. parahaemolyticus* in Europe on a cruise ship off the coast of Galicia, Spain, marking a interoceanic expansion, the first since the pandemic ST3 (77, 78). The following year, ST36 caused an outbreak in 2013 traced to Northeast product with 104 illnesses reported across

multiple states (75, 77). The interoceanic expansion of ST36 highlights its pandemic potential and provides a unique opportunity to broaden our understanding of how pandemic strains arise and spread.

Martinez-Urtaza *et al*. investigated the spatial-temporal and genomic context of the dispersal of ST36; however, in this study we include more strains from the North Atlantic Coast of the US to gain insight into the genetic content and phylogeographic traceback of ST36 in the Northeast United States (76). The previous study into the expansion of ST36 focused on explaining its incursion into new habitats through the location of disease occurrence (76). Whereas location of the reported illness yields valuable data on the epidemiology and disease incidence of the strain, it leaves a gap in knowledge of the ecological context of the expansion. *V. parahaemolyticus* is an environmental pathogen; therefore, understanding the phylogeny of pathogenic lineages linked to the bodies of water from which they arise provides valuable insight into the ecological dynamics of the recent spread of these clades. Robust connection of clinical samples with potential source where the bacteria originated is critical to building a better understanding of ST36 population dynamics. Here we are uniquely situated to shed light on the ecological context of the incursion of ST36, due to our extensive library of both clinical and environmental strains with associated traceback information available for analysis.

One potential factor influencing the spread and establishment of local ST36 populations is filamentous bacteriophage. These viruses infecting bacteria, classified in the family *Inoviridae,* cause chronic infections of the host, which can lead to unexpected influences on host phenotype. One example of the dramatic effect these phages can have on host phenotype and evolution is CTXϕ, a phage responsible for toxigenic conversion in *Vibrio cholerae* (143). Despite the well documented impacts on *V. cholerae* competition dynamics and virulence, filamentous phages in *V. parahaemolyticus* remain largely uncharacterized (112, 152). This is a critical gap in our understanding of *V. parahaemolyticus* diversity, especially considering that nearly every isolate of the pandemic lineage of *V. parahaemolyticus,* ST3, contains a filamentous bacteriophage named f237 integrated into the host (74,

149). Using comparative genomics to identify novel content, we describe novel f237-like filamentous phages closely associated geographic origin of the host population. We propose that these distinct phages can elucidate the spread and establishment of ST36 along the east coast and provide valuable insight into the influence of these phages on population dynamics of *V. parahaemolyticus*.

The incursion of ST36 into the Northeast is a critical concern for shellfish growers and distributers, therefore robust tracing and management is important to limit outbreaks. Monitoring the overall *V. parahaemolyticus* load is often a poor indicator of the amount of pathogenic *V. parahaemolyticus* present in a region, therefore rapid ways to identify strains of particular concern such as ST36 can inform management strategies. More robust traceback methods informed by *V. parahaemolyticus* population structure and genomic markers for use in PCR can improve current protocols. Unique phage content has been previously used to identify certain *V. parahaemolyticus* strains (74, 149). In this context, we propose that phages can also be linked to region of origin of the strain as well as its identity.

In this study, we elucidate the ecological context of the expansion of ST36 by using phylogeny in conjunction with data on geographic origin. We are uniquely situated to shed light on limitations of prior studies, due to our extensive library of both clinical and environmental strains with associated traceback information available for analysis. We describe three novel *Inoviridae* phages closely associated geographic origin of the host population and investigate how the distinct phages found in these populations can help elucidate the spread and establishment of ST36 along the east coast. We propose the phage as regional markers to facilitate rapid tracing of clinical ST36 isolates.

Methods

Acquiring, characterizing, and sequencing Vibrio parahaemolyticus *strains*

This study utilized quality whole genome sequences from strains of broad distribution, including those available through NCBI and our own collection from the Northeast US region to determine

relationships between lineages and phage content. The genomes of 221 quality assemblies of *V. parahaemolyticus* were obtained from NCBI (https://www.ncbi.nlm.nih.gov/genome/681, August, 2019)(Table S1). The sequence type of these strains was obtained by cross-referencing the PubMLST database (www.pubmlst.org). The sequence type of assembled strains not in the database was determined either by using SRST2 or by PubMLST (153).

The clinical *V. parahaemolyticus* strains in our collection were received from Massachusetts, Maine, New Hampshire, and Connecticut State public health labs spanning 2010-2017 (80). Environmental strains were isolated from New Hampshire, Connecticut, and Massachusetts oysters, water, sediment and plankton between 2007-2015 during Most Probable Number analyses as described previously (33, 154–156). All clinical and environmental isolates were confirmed as *V. parahaemolyticus* by detecting the thermolabile hemolysin gene (*tlh*) with PCR as published (157) and many determined by whole genome sequencing. Regional isolates of interest 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 (80). Reads were d*e novo* assembled with the A5 pipeline (158) and sequence-typed with SRST2 (153) and/or PCR amplification and sequencing the house-keeping loci (80).

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

Whole genome variant analysis was performed on ST36 strains from the Pacific Northwest (10290) and Massachusetts (MAVP26, MAVP36) using *breseq* (159). One genome was selected from the PNW (10329) and compared to one from GOM (MAVP-26) using Mauve and visualized using EasyFig (160, 161). The unique prophage subsequently discovered in MAVP-26, MAVP-36 and 10290 were classified belonging to the family *Inoviridae* proposed subfamily Protoinoviridae (162) and named as *Vibrio* phage according to current recommendations, following confirmation that the chosen names were unique (163, 164). The family *Inoviridae* viruses infecting bacteria (vB) were named for the species

in which they infected, *Vibrio parahaemolyticus* (Vipa), and each were given a unique numeric designation representing the strain identity: vB_Vipa26 (in MAVP-26), vB_Vipa36 (in MAVP-36) and vB_Vipa10290 (in 10290). All sequenced *V. parahaemolyticus* strains were evaluated for f237-like phage content (ORF1 – zona occludens toxin) and specific phage type using BLAST (165).

To determine if phage evolved under selective pressure, a Nei-Gojobori codon-based Z test (166) was performed in Mega 6 (167) on the seven core genes shared between the filamentous phage f237, and the related Vipa26, Vipa36, and Vipa10290 described in this paper. Protein sequences for other filamentous phages Vf33 (NC_005948.1), CTXΦ (MF155889.1) and the type strain for *Inoviridae,* M13 (GCF_000845205.1), were retrieved from NCBI database for comparison with Vipa26 using BLASTp (4). Proteins with similar size and location were included although no significant sequence similarity was found.

The evolutionary relatedness of ST36 strains was determined using reference-free, whole genome maximum likelihood phylogenetic trees built by kSNP 3.1 (168) and visualized with iTOL (169). Phage relatedness was found using the Genome-BLAST Distance Phylogeny (GBDP) method (170) to conduct pairwise comparisons of the nucleotide sequences under settings recommended for prokaryotic viruses (171). Branch support was inferred from 100 pseudo-bootstrap replicates each. Trees were rooted at the midpoint (172) and visualized using iTOL (169).

The evolutionary relatedness of strains was determined using reference-free, whole genome alignments using kSNP 3.1 (168) and maximum likelihood phylogenetic trees built by RAxML (173) and visualized with iTOL (169). Whole genome trees were rooted using *Vibrio alginolyticus* ARGOS_108. Visualization of gene/feature gain and loss and recombination in MAVP-26 was determined by comparison with PNW strain 10329 using Mauve and visualized using EasyFig 2.2.0 (160, 161).

Phage typing by PCR

All ST36 clinical isolates were screened with PCR for the presence of phage related to f237 and subsequently, these same primers were employed for surveying environmental strains of unknown sequence type. Step-wise PCR primers were developed to detect f237-like phage content (1000bp), f237-like phage that had content unique within ST36 phages to those traced to New England, (NE-like phage content (618bp)) and primers for Vipa26 and Vipa36 (1440bp and 2870bp, respectively) that can be used in conjunction with *tlh* primers as an internal control (Table 2.1). f237-like primers are specific to the core conserved region of f237 and related phages, specifically ORF3 to ORF5. NE-like primers bind specifically to the hypothetical protein, HypA common to the NE phages and exclude Vipa10290. It is important to note that the NE-like amplicon is present in isolates lacking phage entirely so it is only useful in conjunction with f237-like primers. Isolates positive for f237-like and NE-like phage were screened for Vipa26 or Vipa36 using the Vipa26/36 PCR which captured size differences between the variable regions of the two phage. These primers are utilized to screen for Vipa26 and Vipa36-like phage content in the general population and to identify isolates of interest for further analysis. In ST36 isolates, this screening differentiates between isolates containing Vipa10290 associated with the PNW population and members of the New England populations.

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. Reactions were then visualized on a 0.7% agarose gel for the expected band sizes.

Table 2.1. PCR primers used in this study. Step wise primers used to rapidly screen isolates for potential phage content.

Step	Primers	Genes amplified	5'-3' sequence	Reference
f237-like	ST36Phage F2	ORF3-ORF5	AGCAACGAAAACGCCTGT	This paper
	ST36Phage R2		ACCGTATCACCAATGGACTGT	This paper

NE-like	NEORF10F NEHypR	*HypA*	TTTCTTACTTCTGTGAGCATTTGA GATTACTGAGCCTCTAAAGCCGTC	This paper This paper
Vipa26/36	PhHypDF3 PhORF9R1	*HypD*-ORF9	AAGTGCTACATGAATGAAAGTGCT TCAATGAAGTATCACGAAATGACTA	This paper This paper
tlh Control	TLH-F2 TLH-R	*tlh*	AGAACTTCATCTTGATGACACTGC GCTACTTTCTAGCATTTTCTCTGC	(174) (157)

Results

Phylogeny of ST36 and distribution of filamentous phage within the population

To elucidate the phylogeographic relationship of both clinical and environmental isolates, we constructed a phylogeny with associated traceback data where possible. Martinez-Urtaza *et al*. identified two lineages tracing to the Pacific Northwest (PNW) designated the "old" and "new" clades (both Green); here we will use the same terminology to differentiate the two major clades representing distinct lineages traced to the PNW throughout our analysis (76). The topology of the phylogenetic tree illustrates two main branches containing several distinct clades of ST36 strains, and excepting a few isolates these clades link closely to the geographic traceback of the isolates (Fig. 2.1). One major branch includes three clonal clades, all derived from the "new" PNW lineage that are traced predominantly to North Atlantic sources of the US including the Gulf of Maine (most blue isolates), Long Island Sound (most light orange isolates), and Katama Bay (most dark orange isolates). The Gulf of Maine (GOM) clonal clade (Fig. 2.1; blue) shared its most recent common ancestor with the new PNW lineage, and was not derived from within the other ST36 populations traced to sources of the Atlantic, including those from the Mid-Atlantic and Long Island Sound, or from California. Surprisingly, one Long Island Sound (LIS) isolate (CTVP44C) with a definitive trace to a local source in 2013 is a member of the old PNW lineage (Fig. 2.1; green with black square), (Fig. 2.1; light orange) on the second major branch of the tree. The isolates from the outbreak caused by ST36 in Galicia, Spain in 2012 also group closely with this old population (Fig 2.1). Interestingly, 10-4298, a clinical isolate from the PNW collected in 2001 is a part of the Spanish clade (Table S2.2, Fig. 2.1) Two isolates from the PNW group (VpG-11 and 3324) group

closely with strains from California and the LIS. Adjacent to the LIS clade, there is a clonal group tracing to Katama Bay (KB) (Fig. 2.1; dark orange) and possibly another clade tracing to the Mid-Atlantic Coast (MAC) (Fig. 2.1; purple); however, more strains with robust geographic traceback are needed to confirm this lineage.

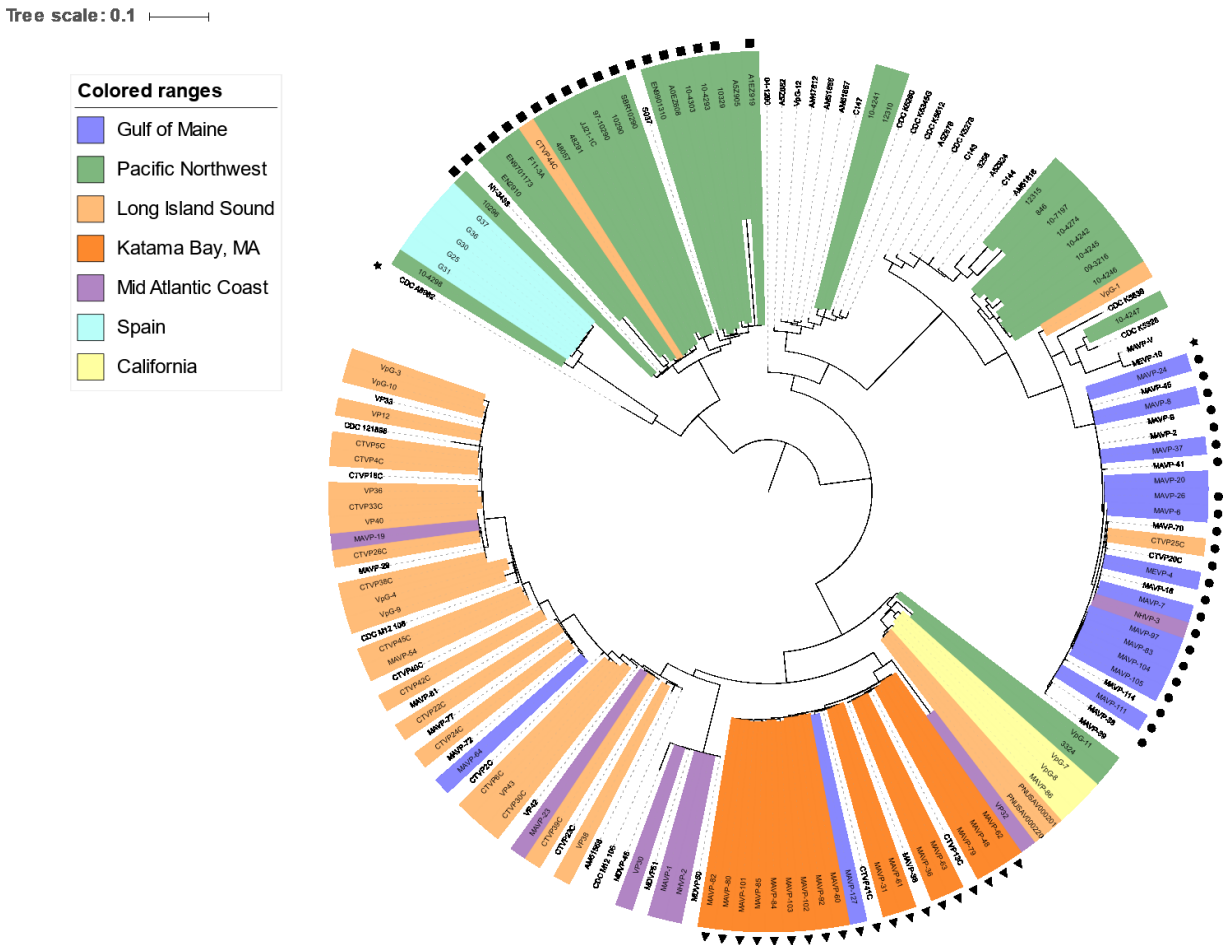

Figure 2.1. Maximum-likelihood tree of select *Vibrio parahaemolyticus* ST36 isolates. Isolates are colored to correspond to geographic origin of infection source, no color unknown traceback or several different potential source locations, due to inaccuracies inherent in traceback of clinical isolates. The shape next to each isolate indicates a specific but highly similar phage: square is Vipa10290, triangle is Vipa36, circle is Vipa26, and star is other which are not necessarily identical.

Although members the GOM clade and the LIS clade have established populations that reside in close proximity to each other in the Atlantic and share a common ancestor with the new PNW clade, this whole genome phylogeny suggests they are distinct lineages. The GOM and KB clades are both currently

clinically prevalent, whereas clinical isolates from the old PNW lineage are no longer highly clinically prevalent since the 1990s and early 2000s. To investigate differences in genetic content, we compared the complete, closed genomes of 10329 (PNW), MAVP-36 (KB) and MAVP-26 (GOM). Several major insertions, duplications and deletions occurred during evolution and divergence of the strains (Figure S2.1) and could be the focus of future study. Here we investigate more closely one difference between strains; the phage content unique to each clonal population, specifically the type of filamentous prophages found in the genome (Figure S2.1). The phages are classified in the family *Inoviridae* (inoviruses) and are designated according to current recommendations as viruses infecting bacteria (vB) Vipa26 in the GOM clonal population and Vipa10290 in the old PNW lineage. Further investigation of ST36 revealed a third similar prophage, subsequently named Vipa36, in the KB population (175). Since the outbreak in 2013, all but two clinical isolates (MAVP-20, MAVP-89, both GOM) in the GOM or KB clonal clades harbor their respective phage (Fig. 2.1). MAVP-41 contains a unique filamentous phage (Fig. 2.2); however, further analysis of genetic content reveals a high degree of sequence identity to Vipa26 across multiple scaffolds.

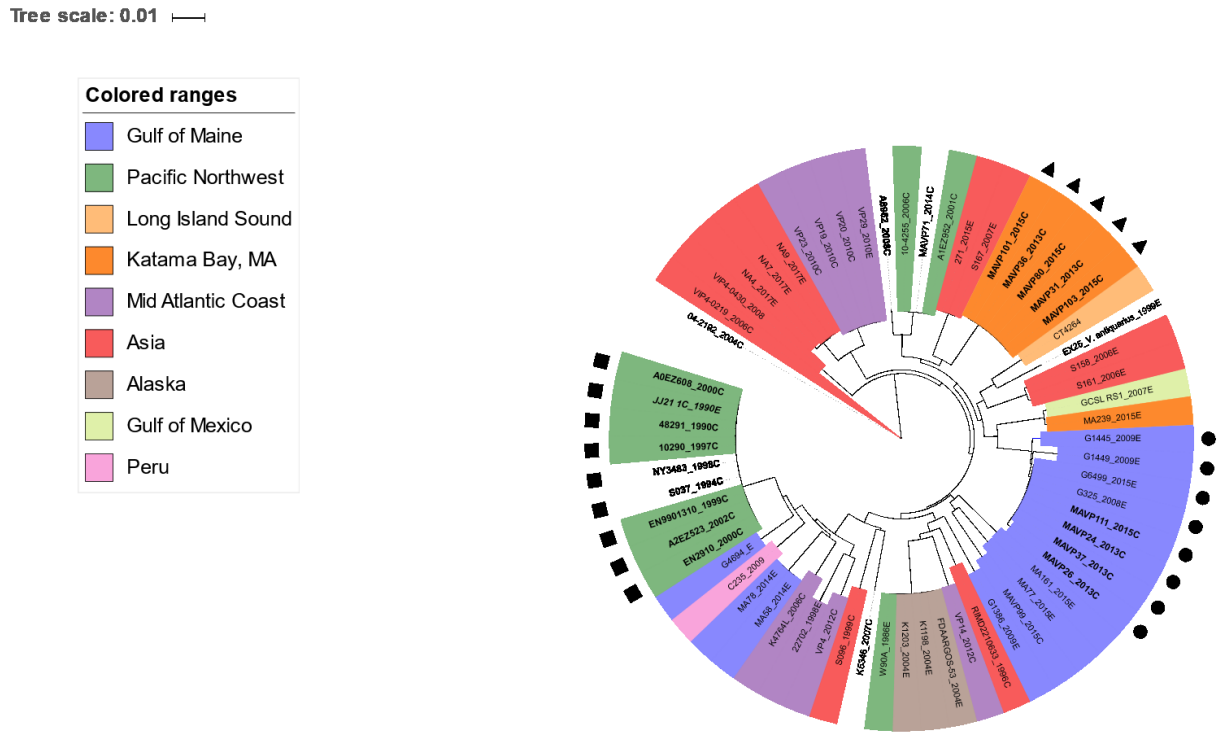

Figure 2.2. Phylogenetic tree of filamentous phages found in diverse STs. Genome-BLAST Distance Phylogeny (GBDP) of whole phage genomes of environmental and clinical *V. parahaemolyticus* isolates with date of isolation. Clinical and environmental isolates are indicated by a C or E, respectively. Traceback color coding relates to Figure 2.1. Shape next to isolate name corresponds to phage type (Fig. 2.1); circle is Vipa26, square is Vipa10290 and triangle is Vipa36. Isolates in bold are ST36 except for ST38 JJ21-1C (italicized and bold), which is a part of the ST36 clonal complex.

The presence of Vipa26 and Vipa36 in almost every clinical isolate from two highly successful and localized clades suggests the progenitors of each population acquired and maintained these inoviruses through clonal expansion. To determine whether these phages originated from local populations prior to expansion or the invading progenitor of the clade already contained these phage, we examined the evolutionary relatedness of filamentous phages in *V. parahaemolyticus*. A Genome Blast Distance Tree of whole phage genomes was constructed from phages found in diverse *V. parahaemolyticus* strains.

Although Vipa26 is common in the ST36 GOM clade, it also occurred in non-ST36 environmental isolates from the GOM (G1445, G1449, G325 and G6499) (Fig. 2.2) one of which (G325) was isolated in 2008 before ST36 was detected in areas of the Gulf of Maine. Vipa26 and Vipa36 are more closely related with each other than they are to Vipa10290, in contrast to the arrangement of the host phylogeny (Fig. 2.1). Together this suggests that the phage in these ST8 strains were likely horizontally acquired after the invasion of ST36 into the Atlantic. Every ST8 isolate that caused an outbreak in Maryland in 2010 (VP19 and VP20) carry the same filamentous phage that was identical to phage contained by ST8 environmental isolates at the same time (VP21, VP23, VP24 and VP29) (Fig. 2.2) (176). Interestingly, the most closely related phage to the Vipa10290 clade was from a GOM environmental isolate (G4694) (Fig. 2.2). Additionally, a phage infecting the deep-sea species *Vibrio antiquarius* (EX25) is closely related to an inovirus from an environmental isolate collected in Connecticut, CT4264 (Fig. 2.2).

Comparison of inoviruses of V. parahaemolyticus *reveal unique content*

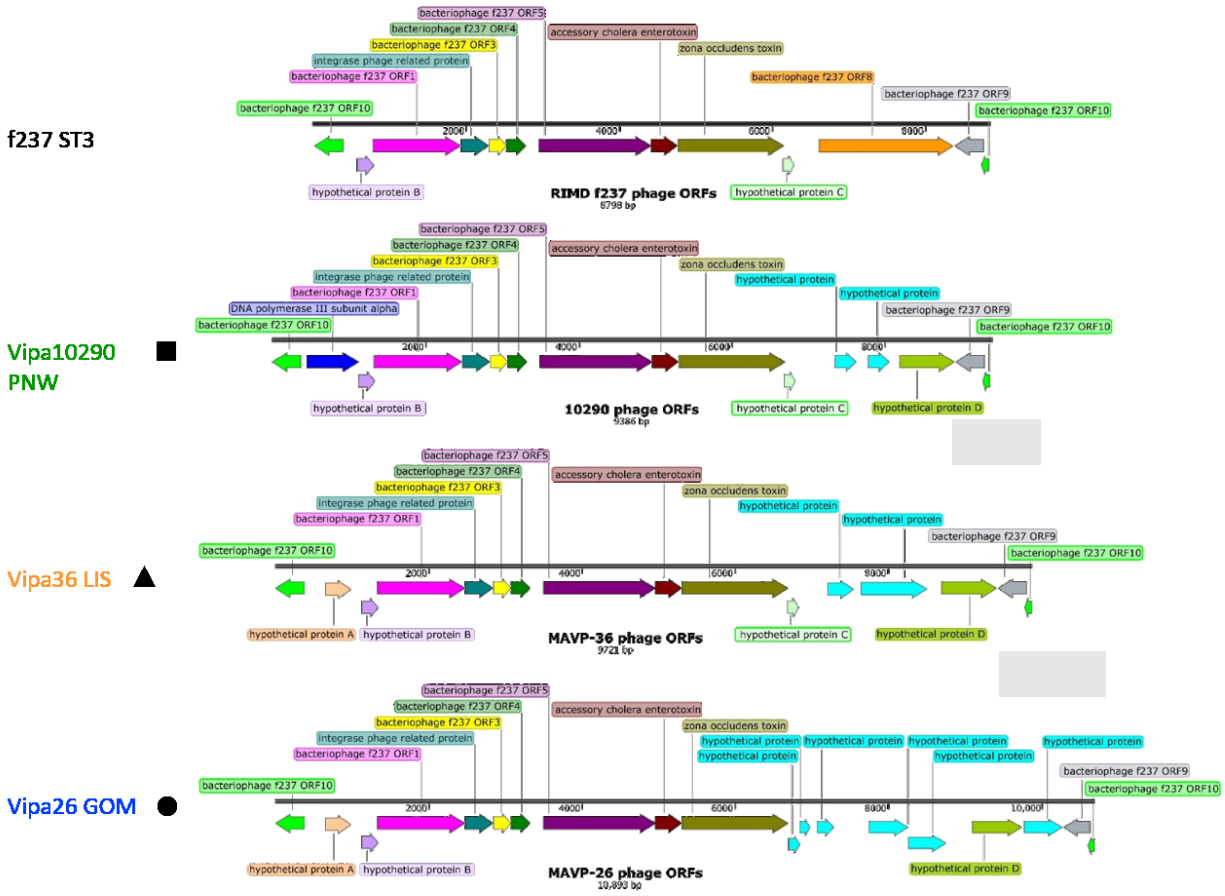

Figure 2.3. Shared core and distinctive content of filamentous f237-like phages. ORFs with the same color are homologous between the different phages with the exception of blue hypothetical proteins. The core genome (highlighted in gray) is highly conserved, while accessory regions vary, with some ORFs conserved in two or more of the phages. Color of the phage names and shapes correspond to those in Figures 2.1 & 2.2 (175).

A comparison of the genomes of Vipa26, Vipa36 and Vipa10290 elucidated their grouping with distinct geographic clades. This revealed similarity in structure to the ST3-associated f237 filamentous phage with a conserved core genome and two regions of highly variable content. . The ST36 phages lack ORF8 that is present in f237, and a useful diagnostic marker of the ST3 pandemic clonal complex. The core genome of ORF1-7 (bright pink through green, Fig. 2.3) is also flanked by the conserved ORF9 and ORF10 (gray and light green, Fig. 2.3). These seven core genes (ORF1-7) have evolved under purifying selection (codon-based Z test; dS – dN =9.425, $p < 0.001$) and as is expected for genes encoding essential

phage functions in an active phage. The Northeast ST36 population associated phages, Vipa26 and Vipa36, both share variable region content HypA (light pink in Fig. 2.3) whereas Vipa10290 lacks this hypothetical protein. Vipa26, Vipa36 and Vipa10290 all have in common the HypD hypothetical proteins (light green, Fig2.3), which is absent in f237. The ST36 phages as well as f237 share HypB (bright purple, Fig. 2.3), another hypothetical protein in the variable region, whereas all but Vipa26 contain HypC (light blue-green, Fig. 2.3). Comparison of translated nucleotide sequence using BLAST; however, reveals similarity in amino acid sequence of the hypothetical protein syntenic to HypC in Vipa26 (E-value<0.000001) with no relatedness of the nucleotide sequence (no alignment). All Vipa26 and Vipa36 prophages in sequenced isolates were integrated into the *dif* site of chromosome I as expected of inoviruses (150).

Analysis of amino acid sequence of the phage proteins elucidated the potential function of these through comparison with previously described phages, Vf33, CTXϕ, and the type strain for *Inoviridae*, M13 (Table 2.2). Vf33 infects other *V. parahaemolyticus* strains, CTXϕ infects *V. cholerae* and M13 is specific to *Escherichia coli*. The inoviruses encode their own integrase, whereas M13 replicates only extrachromosomally. Interestingly, the phages in ST36, as well as f237, harbored accessory cholera enterotoxin (*ace*) and zonula occludens toxin (*zot*) homologs as well as a putative transcriptional regulator homologous to RstR (ORF9) (Table 2.2). ORF10 was not homologous to ORFs in any phages other than Vf33 and its function is unknown.

Table 2.2. Homologs and potential functions of conserved core region of Vipa26. Vipa26 core protein homology compared to Vf33, a similar phage also found in *Vibrio parahaemolyticus*, CTXΦ from *Vibrio cholera* El Tor biotype and M13 of *Escherichia coli.* Core and accessory proteins without homology to other phages are not shown.

Vipa26 Proteins	Vf33 (E-value/Identity)	CTXΦ (E-Value/Identity)	M13 (E-Value/Identity)	Predicted Function Based on Homology[a]

ORF1	Vpf402 (0.0/99%)	RstA (7e-66/36%)	gII (size and location)	replication initiation protein
Integrase	Vpf117 (1E-85/92.5%)	RstB (0.04/13.3%)	-	Integration
ORF3	Vpf81 (0.004/7.8%)	RstC (size and location)	gV (size and location)	ssDNA binding protein, helix destabilizing
ORF4	Vpf77 (5.3/3.7%)	Cep (7.0/7.3%)	gVIII (0.45/4.9%)	major coat protein
ORF5	Vpf491 (0.0/59%)	OrfU (0.54/11.9%)	gIII (0.6/11.3%)	adsorption, termination of assembly, tail protein
Ace	Vpf104 (0.2/19%)	Ace (5e-10/18.3%)	gVI (4.1/8.7%)	minor coat protein, termination of assembly
Zot	Vpf380 (1E-4/13.9%)	Zot (2e-27/17.8%)	gI (5e-04/8.5%)	assembly protein, maturation
ORF9	Vpf122 (4e-66/68%)	RstR (0.17/5.7%)	-	transcriptional repressor, regulator

[a] *Functions summarized from Mai-Prochnow et al., 2015* (92) *and Chang et al., 1998* (177)

Screening for inoviruses in genomes and environmental isolates

Screening of available genomes using BLAST revealed filamentous phages distributed throughout diverse sequence types of *V. parahaemolyticus* (Fig 2.2). In order to better understand the ecological context and phage abundance of the communities invaded by KB and GOM ST36 founders, we screened environmental isolates from several locations in the Northeast for phage content using primers designed to identify 237-like filamentous phages and to detect and differentiate Vipa26, and Vipa36 in clinical ST36 isolates (see methods). Many isolates containing f237-like prophage were

identified (Table 2.3). Isolates positive for f237-like, and NE-like amplicons were further tested for Vipa26/36 content. Only if all three markers were present was an isolate putatively identified as containing Vipa26 or Vipa36 using this PCR method. Vipa26 was only identified in 11 of the isolates from GOM and 2 from LIS, whereas putative VIpa36 positive isolates were encountered less frequently, with 9 from GOM and 1 from LIS (Table 2.3).

A selection of the isolates was sequenced and checked for phage content using BLAST. Although Vipa36-positive amplicon profiles were identified in 10 isolates this phage was not present in any environmental isolates (Table 2.3, Table S2.2, Table S2.3). Of these isolates available with robust sequence data, genomic analysis uncovered phages similar to Vipa36 with variable regions of similar size, possibly explaining the false positive identification considering the PCR assay relies on size differentiation. Additionally, some isolates harbored two filamentous phages in their genomes (MA145 and G1286, Fig. S2.2), which could complicate identification with PCR with multiple bands or false positives. Genomic analysis confirmed at least four environmental isolates contained Vipa26, and the strains were of diverse STs.

Further analysis of the distribution of HypA, one of the markers that distinguishes Vipa-containing ST36 isolates established in the Atlantic from those of the native ST36 population (NE-like phage), revealed its presence in *V. parahaemolyticus* strains isolated from other locations, even in the absence of phage. Although the core region of the inoviruses we compared is highly conserved, the variable regions are divergent with no conserved structures between diverse phages. These regions varied greatly in size as well, with a larger variable region located downstream of the central core (between ~1500-4000bps) and a smaller variable region upstream of the central core (<1000bps). Several strains had phages with a core region similar to the f237-like core, but were divergent and did not align to the middle ~300bp. PCR detection is highly accurate and sensitive in ST36 isolates; however,

in other STs an extremely high rate of false positives (10/10 of detected Vipa36 and 9/13 of detected Vipa26) points to genomic analysis as more accurate.

Table 2.3 Survey of the distribution of f237-like inoviruses among Northeast regional environmental isolates. Environmental isolates from the GOM and LIS populations screened for phage similarity to identify possible Vipa26 and Vipa36 in non-ST36 populations. f237-like phage contain the core region typical of Inoviridae, as determine by BLAST alignment or PCR amplification. Vipa26/36 are scored positive only if f237-like, NE-like and Vipa26/36 bands are present when using PCR screening.

	PCR DETECTION				GENOME ANALYSIS			
	GOM		LIS		GOM		LIS	
	MA	NH	MA	CT	MA	NH	MA	CT
f237-like	46	19	12	5	13	8	2	2
Vipa26	7	4	1	1	0	4	0	0
Vipa36	7	2	0	1	0	0	0	0
Total Screened	154	138	28	50	14	10	2	2

Distribution of phage throughout V. parahaemolyticus populations

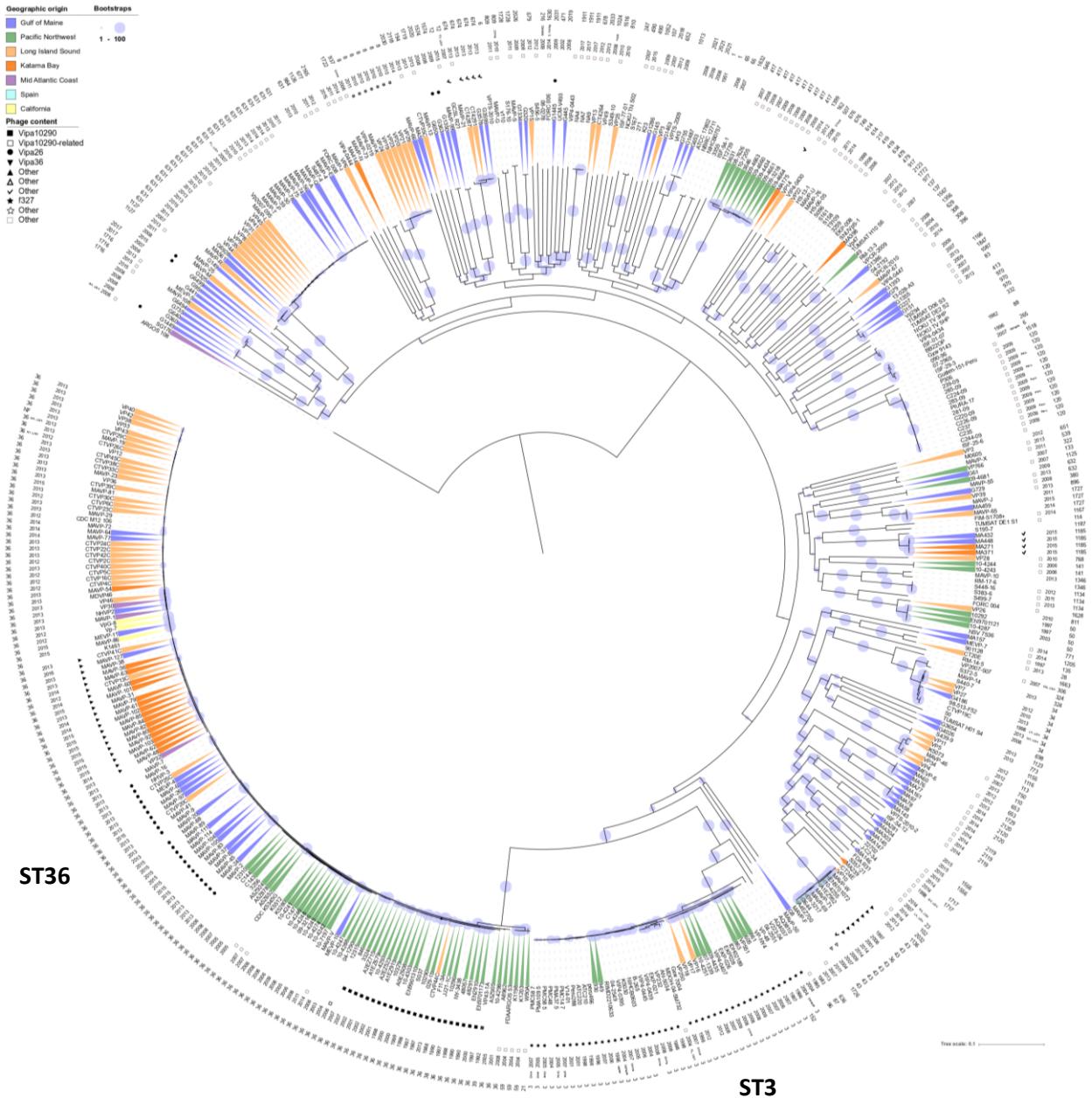

Figure 2.4. Distribution of inoviruses across diverse *V. parahaemolyticus* populations. Strain branches are colored if traceback data is known, colors match Figure 2.1. Year collected is indicated in the middle ring and the ST is listed around the outer edge. Specific filamentous phages are assigned different shapes – Vipa26 is a circle, Vipa10290 is a dark square and Vipa36 is a upside down triangle. There are other lineages of related phage each assigned a unique shape but are all referred to as Other in the legend. Other phages marked with an empty square with a grey outline are diverse and not necessarily related (Figure courtesy of Randi Foxall).

Inoviruses are common throughout ST36 environmental and clinical isolates, suggesting robust environmental reservoirs of phage present in Northeast waters. To further investigate the relationship of *V. parahaemolyticus* strains with filamentous phages, inovirus distribution was assessed in the global population represented in NCBI genomes (https://www.ncbi.nlm.nih.gov/genome/681, August, 2019) and our own collection of strains (Table S2.2, Table S2.3). Filamentous phages are fairly common with about 46% of isolates analyzed containing some type of phage (Table S2.1). Several clades (ST1185, ST3, ST8; Fig. 2.4) appear to have basally acquired a filamentous phage and held them through the expansion of the lineage whereas other closely related isolates have diverse phage content (ST36, ST631, ST43; Fig. 2.4). Interestingly, even in lineages with apparent basal phage acquisition the occasional phage deficient isolate appears (PMC48 in ST3; MAVP-20 in ST36 - GOM clade; Fig. 2.4, Table S2.3 and S2.4). Several distinct phages, including Vipa26, appear across multiple highly divergent sequence types supporting that these phages are active and move horizontally through populations.

Discussion

V. parahaemolyticus ST36, a strain of high clinical prevalence, expanded extensively into new habitats from its native PNW (76, 77). Understanding the diversification and spread of this lineage provides valuable insight into the evolution and ecology of newly arising pre-pandemic strains. Previous studies characterizing the spread of ST36 use only reporting state of illness to trace its interoceanic movement (76). A lack of available sequence data on clinical isolates from the USA PNW in recent years compounds this gap in knowledge of the ST36 expansion. To address this, we examine more closely the ecological context of diversity in ST36, leveraging a large library of strains with robust traceback data. Taken together, the topology and associated traceback of ST36 phylogeny implicates the PNW as a hot bed of evolution for ST36, generating diverse populations able to establish successfully in new habitats. The distribution of unique inoviruses in distinct clades alludes to a role of these phages in the evolution of *V. parahaemolyticus.*

Although the old PNW clade no longer causes illness in the PNW after the new PNW lineage replaced it in the early 2000s, one isolate containing Vipa10290 belonging to the old PNW population was traced to LIS in 2013 (CTVP44C, Fig 2.1) arguing that the introduction and establishment of ST36 occurred earlier than previously thought or the old PNW population is still present and simply no longer clinically prevalent contrary to previous reports (76). Martinez-Urtaza *et al.* also postulated that some of the diversification of the North Atlantic clades occurred in the Atlantic after introduction from the PNW (76). The distinct histories of the GOM and LIS clades contends otherwise. These clades are close geographically; however, the GOM clade shares a more recent common ancestor with the new PNW clade than the LIS, KB or MAC. These clades south of Cape Cod are more closely related and may have diversified in the Atlantic. It is difficult to fit the California clade (Yellow, Fig. 2.1) although the presence of a PNW isolate within this clade may support that this branch first arose on the West Coast.

Although the new PNW lineage does not contain an f237-like phage, both the GOM and KB clades have unique phage content. These clades likely acquired these phages from the environment prior to the clonal expansion of the progenitor of each clade. Environmental isolates from the GOM harbor Vipa26 and one of these isolates was collected several years prior to the first detection of ST36 on the US East Coast, minimizing the likelihood that the phage was transferred from ST36 into the local populations. In addition, all the environmental isolates harboring Vipa26 trace to the Great Bay in New Hampshire, an area where ST36 has never been detected despite robust surveillance. Although environmental screening did not detect Vipa36, this is likely due to considerably fewer available environmental isolates from KB and the LIS than from the GOM and none pre-dating the incursion of ST36 into the region.

Considering the explosion of diversity from the PNW population and subsequent expansion of ST36 into new territory was concurrent with the loss of Vipa10290 from a portion of the PNW population, these phages may have a stabilizing influence on the host population. ST36 was fairly stable

and successful when associated with Vipa10290; however, when the infection was no longer beneficial, it led to considerable diversification and several novel lineages arising uniquely suited to new ecological niches. Subsequently, the acquisition of Vipa26 and 36 in the GOM and KB respectively, led to the establishment of persistent clonal clades. This hypothesis argues that phage content is in part a driving factor behind the rapid diversification and trans-Atlantic expansion of ST36. The complex relationship between host, phage and environment warrant investigation to further our understanding of the establishment of pre-pandemic *V. parahaemolyticus* lineages.

Diversity of phages does not correlate closely with diversity of their hosts but rather, the number of closely related phages present in *V. parahaemolyticus* and one from *V. antiquarius* suggests a large pool of phage moving horizontally through populations and even potentially between closely related species (Fig. 2.2). On the other hand, basal lysogenic conversion by filamentous phage is sometimes propagated through a clade, leading to strong correlation between specific phages with certain clades. *Inoviridae* are capable of forming mutualistic relationships with their hosts; however, the balance is delicate with the same phage having different impacts on closely related strains (178). Vipa26 and Vipa36 could be the perfect mutualists for the GOM and KB populations respectively, whereas the upset of the balance between Vipa10290 and the old PNW clade lead to diversification of ST36. The phage deficient isolates arising occasionally in closely phage-associated populations could occur when the ecological context changed for that isolate enough to cause the phage to become detrimental. Alternatively, it could reflect a background level of prophage excision and re-infection amongst these populations.

In ST36, the correlation between clade and geographic origin provides a unique opportunity to improve the accuracy of traceback for ST36 clinical isolates. Traceback is important for managing the spread of *V. parahaemolyticus* during outbreaks by identifying high risk areas and closing them to prevent further infections. Due to the movement of shellfish and people, relying on the reported

location of clinical cases does not provide enough information for meaningful phylogeographies. The PCR primers designed for this study are a rapid and accurate way of determining likely origin of clinical ST36 isolates for more targeted management strategies. While valuable for aiding in discovering the origin of ST36, non-specific binding of the NE-like primers in non-phage harboring isolates of other STs limits the primers to just ST36. Genomic analysis is more accurate but also more expensive and time consuming although costs are decreasing as technology improves.

Additionally, Vipa26, Vipa36 and Vipa10290 all display the highly conserved core and variable regions typical of the family *Inoviridae*, a diverse class of phage known for the potential impacts they may have on the lysogen (92). Alongside the highly conserved core genome are highly variable regions with widely differing content between phages. In CTXϕ, the *ctxA* and *ctxB* enterotoxins are located in this variable region (12, 143), as well as ORF8 in f237, speculated to be involved in adhesion in the host (74, 149). Differing variable content could cause these phages to have highly divergent impacts on the host cell behavior. The transduction of novel DNA through acquisition of these bacteriophages is a noted mechanism of bacterial evolution and thus may be linked to the evolution of these pathogenic lineages (179, 180). Vipa26 and Vipa36 were both acquired by the progenitors of two successful clonal clades and held through expansion, whereas f237 is universally associated with the only pandemic clonal complex, ST3, suggesting the presence of phage in these lineages is not a coincidence (74, 149). Future studies into the potential impact of these phages on these populations may elucidate potential fitness or virulence benefits associated with acquisition of filamentous phages.

The discovery of diverse phage in many *V. parahaemolyticus* populations indicates a potentially broader application of filamentous phage content as a marker of geographic origin as well as identifying the particular clade the isolate belongs to as is already done with ORF8 of f237 in ST3 (149). The use of accurate traceback data and phage content provides valuable ecological context to *V. parahaemolyticus* spread and phylogeography. This could allow for better understanding of the evolution and subsequent

movement of these pathogens, informing management practices including more rapid and targeted closures of effected oyster farms to ensure limited economic impact on growers and distributers.

Supplementary

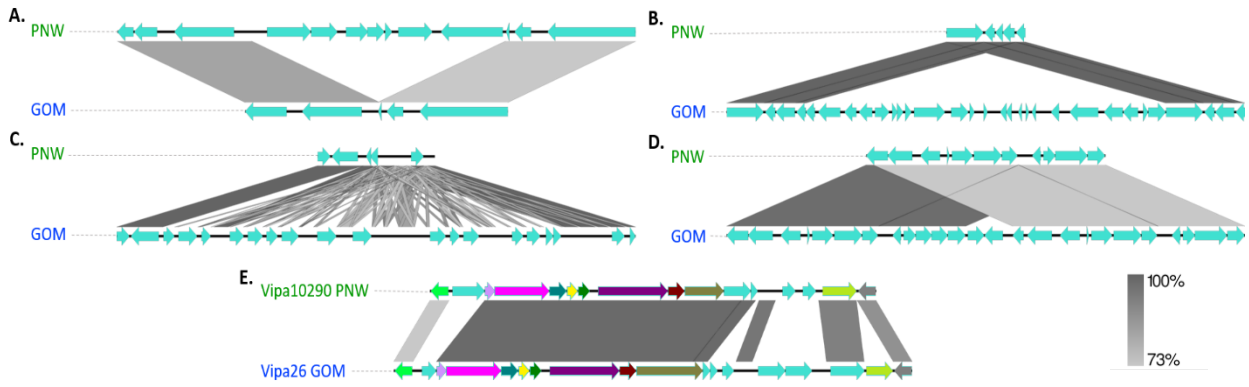

Figure	Chromosome	Change	Cause	Notable Elements
2 A.	1	Δ6,945bp	-	-
2 B.	1	Insert 22,338bp	Possible duplications	Motor proteins Metabolic enzymes Transcriptional regulator
2 C.	2	Insert 14,910bp	-	Toxin/Antitoxin Plasmid stabilizing protein
2 D.	1	Insert 24,780bp	Transposase and duplication	Maltose transport Capsular polysaccharides
2 E.	1	-	Phages	Filamentous phages (Inoviridae)

Figure S2.1. Comparison between PNW and GOM genomes. A-D. Depictions of insertions, duplications and deletions visualized with EasyFig2.2.2 (161), **E**. Filamentous phages integrated into chromosome 1 are colored consistently with Fig. 2.3. BLAST identity gradient is shown on the bottom right of the figure. Notable genetic content and variation in the GOM isolate (MAVP-26) compared to a historic PNW isolate (10329) is summarized and the associated figure is referenced.

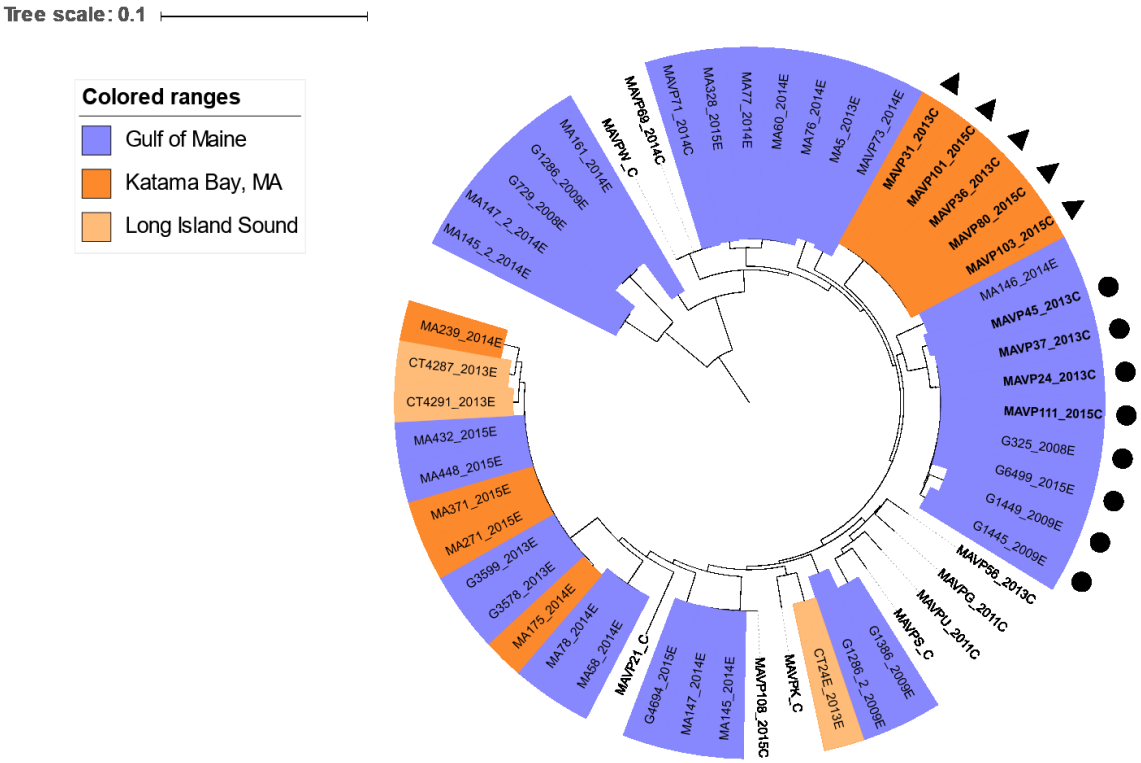

Figure S2.2. Phylogeny of filamentous phages found in New England isolates. Genome-BLAST Distance Phylogeny (GBDP) of whole phage genomes of environmental and clinical *V. parahaemolyticus* isolates with date of isolation. Clinical and environmental isolates are indicated by a C or E after the year of isolation. Colors of geographic traceback match with Figure 2.1 and 2.2, and bold indicates ST36 isolates. Shapes next to isolate names highlight Vipa26 (circle) and Vipa36 (triangle).

Table S2.1. Summary of phage content in all isolates used in this study. Number of isolates found harboring each specific phage according to genomic analysis. Other phages are not necessarily the same.

Phage Type	Count
Vipa26	23 (4.67%)
Vipa36	20 (4.06%)
Vipa10290	21 (4.26%)
f237	43 (8.72%)
Other Phages	121 (24.54%)
Total Isolates with Filamentous Phage	227 (46%)
Total Isolates Screened	493

Table S2.2. List of ST36 strains used in this study. All strains used in this study including NCBI assembly ID unless sequenced in-house, year and geographic region collected, clinically or environmentally isolated (C/E respectively) sequence type and phage type. N/A for any category indicates the information was not available.

NAME	ASSEMBLY	YEAR	REPORT-ING STATE	SOURCE TRACE	COUNTRY	C/E	ST	PHAGE
846	GCA_000877405.1	2007	WA	N/A	USA	E	36	Vipa10290
3256	GCA_000519405.1	2007	WA	WA	USA	C	36	None
3324	GCA_000877495.1	2007	WA	N/A	USA	C	36	Vipa10290
10290	GCA_000454205.1	1997	WA	WA	USA	C	36	Vipa10290
10296	GCA_000500105.1	1997	WA	WA	USA	C	36	Vipa10290
10329	GCA_001188185.1	1998	WA	WA	USA	C	36	Vipa10290
12310	GCA_000500755.1	2006	WA	WA	USA	C	36	None
12315	GCA_000877535.1	2006	WA	N/A	USA	C	36	Vipa10290-related
48057	GCA_000706825.1	1990	WA	WA	USA	C	36	Vipa10290
48291	GCA_000707525.1	1990	WA	WA	USA	C	36	Vipa10290
04-1290	GCA_000878815.1	2004	Alberta	N/A	Canada	C	36	None
09-3216	GCA_000878785.1	2009	BC	N/A	Canada	C	36	None
09-3216_100	GCA_002144555.1	N/A	N/A	N/A	Canada	N/A	36	None
10-4241	GCA_000878805.1	2006	BC	N/A	Canada	C	36	None
10-4242	GCA_000878755.1	2006	BC	N/A	Canada	C	36	None
10-4245	GCA_000878725.1	2006	BC	N/A	Canada	C	36	None
10-4246	GCA_000878705.1	2006	BC	N/A	Canada	C	36	None
10-4247	GCA_000878665.1	2006	BC	N/A	Canada	C	36	None
10-4248	GCA_000878675.1	2006	BC	N/A	Canada	C	36	None
10-4255	GCA_001006195.1	2006	BC	N/A	Canada	C	36	Other
10-4274	GCA_000878735.1	2005	BC	N/A	Canada	C	36	None
10-4288	GCA_000878645.1	2003	BC	N/A	Canada	C	36	None
10-4293	GCA_000878595.1	2002	BC	N/A	Canada	C	36	Vipa10290
10-4298	GCA_000878565.1	2001	BC	N/A	Canada	C	36	None
10-4303	GCA_000878575.1	2000	BC	N/A	Canada	C	36	Vipa10290
10-7197	GCA_000878585.1	2008	BC	N/A	Canada	C	36	None
A0EZ608	GCA_001608655.1	2000	BC	N/A	Canada	C	36	Vipa10290

A1_9	GCA_006378975.1	2014	N/A	N/A	China	E	36	Other
A1EZ679	GCA_001609075.1	2001	BC	N/A	Canada	C	36	Vipa10290
A1EZ919	GCA_001559895.1	2001	BC	N/A	Canada	C	36	Vipa10290
A2EZ523	GCA_001609095.1	2002	BC	N/A	Canada	C	36	Vipa10290
A2EZ715	GCA_001608675.1	2002	BC	N/A	Canada	C	36	Vipa10290
A5Z652	GCA_001609515.1	2005	BC	N/A	Canada	C	36	None
A5Z878	GCA_001609695.1	2005	BC	N/A	Canada	C	36	None
A5Z905	GCA_001609635.1	2005	BC	N/A	Canada	C	36	None
A5Z924	GCA_001609675.1	2005	BC	N/A	Canada	C	36	None
AM47612	GCA_002198255.1	2011	WA	WA	USA	C	36	None
AM51556	GCA_002198415.1	2012	FL	N/A	USA	C	36	None
AM51816	GCA_002198485.1	2012	CA	N/A	N/A	C	36	None
AM51866	GCA_002198355.1	2012	WA	N/A	USA	C	36	None
AM51867	GCA_002198435.1	2012	WA	WA	USA	C	36	None
C143	GCA_001609395.1	2008	BC	N/A	Canada	C	36	None
C144	GCA_001609415.1	2008	BC	N/A	Canada	C	36	None
C147	GCA_001609495.1	2008	BC	N/A	Canada	C	36	None
CDC_121898	GCA_002198395.1	2012	NJ	N/A	USA	C	36	None
CDC_A8962	GCA_002198515.1	2008	MA	N/A	USA	C	36	Other
CDC_K4639G	GCA_001727895.1	2006	NY	N/A	USA	C	36	None
CDC_K4639W	GCA_001727885.1	2006	NY	N/A	USA	C	36	None
CDC_K5278	GCA_001728405.1	2007	N/A	N/A	USA	C	36	None
CDC_K5278	GCA_001728405.1	2007	WA	N/A	USA	C	36	None
CDC_K5280	GCA_001727625.1	2007	WA	WA	USA	C	36	Other
CDC_K5281	GCA_001727575.1	2007	WA	N/A	USA	C	36	None
CDC_K5308	GCA_001727485.1	2007	AK	N/A	USA	C	36	Other
CDC_K5323G	GCA_001727475.1	N/A	VA	N/A	USA	N/A	36	Other
CDC_K5328	GCA_001727665.1	N/A	N/A	N/A	USA	C	36	None
CDC_K5345G	GCA_001728465.1	2007	IA	N/A	USA	C	36	Other
CDC_K5345W	GCA_001728485.1	2007	IA	N/A	USA	C	36	None
CDC_K53	GCA_001728415.1	2007	PA	N/A	USA	C	36	None

46								
CDC_K5429	GCA_001727585.1	2007	NV	N/A	USA	C	36	None
CDC_K5433	GCA_001727385.1	2007	WA	N/A	USA	C	36	None
CDC_K5437	GCA_001727565.1	2007	WA	N/A	USA	C	36	None
CDC_K5456	GCA_001728545.1	N/A	WA	N/A	USA	C	36	None
CDC_K5457	GCA_001728505.1	2007	WA	N/A	USA	C	36	None
CDC_K5512	GCA_001728585.1	2007	OK	N/A	USA	C	36	Other
CDC_K5629	GCA_001728695.1	2007	GA	N/A	USA	C	36	None
CDC_K5638	GCA_001728745.1	N/A	MD	N/A	USA	C	36	None
CDC_M12_106	GCA_002198375.1	2012	MO	MA	USA	C	36	None
CDC_M12_108	GCA_002198275.1	2012	MO	NY	USA	C	36	None
CFSAN001611	GCA_000707045.2	1997	OR	OR	USA	E	36	Vipa10290
CFSAN001612	GCA_000706825.2	1990	WA	N/A	USA	C	36	Vipa10290
CFSAN001613	GCA_000707625.2	1997	WA	N/A	USA	C	36	Vipa10290
CFSAN001618	GCA_000707525.2	1990	WA	N/A	USA	C	36	Vipa10290
CFSAN001619	GCA_000707545.2	1998	WA	WA	USA	E	36	Vipa10290
CFSAN001620	GCA_000707565.2	1998	NY	NY	USA	E	36	Vipa10290
CFSAN006133	GCA_000707865.2	2013	MD	N/A	USA	C	36	None
CFSAN007461	GCA_000707205.2	2004	N/A	N/A	USA	N/A	36	None
CFSAN007462	GCA_000706905.2	2004	N/A	N/A	USA	N/A	36	None
CFSAN018777	GCA_002198525.1	2012	N/A	N/A	USA	N/A	36	None
CFSAN022336	GCA_002198615.1	2012	Galicia	Galicia	Spain	C	36	None
CTVP2C	In-house	2012	N/A	N/A	USA	C	36	None
CTVP4C	In-house	2012	CT	CT	USA	C	36	None
CTVP5C	In-house	2012	CT	CT or NY	USA	C	36	None
CTVP6C	In-house	2012	CT	NY	USA	C	36	None
CTVP13C	In-house	2012	CT	N/A	N/A	C	36	Vipa36

CTVP16C	In-house	2012	N/A	N/A	USA	C	36	None
CTVP20C	In-house	2013	CT	N/A	N/A	C	36	Vipa26
CTVP22C	In-house	2013	CT	N/A	USA	C	36	None
CTVP23C	In-house	2013	CT	MA or VA	USA	C	36	None
CTVP24C	In-house	2013	CT	N/A	USA	C	36	None
CTVP25C	In-house	2013	CT	CT	USA	C	36	Vipa26
CTVP26C	In-house	2013	CT	N/A	USA	C	36	None
CTVP27C	In-house	2013	CT or VA	N/A	USA	C	36	None
CTVP30C	In-house	2013	CT	CT	USA	C	36	None
CTVP33C	In-house	2013	CT; RI	N/A	USA	C	36	None
CTVP38C	In-house	2013	CT, MA	N/A	USA	C	36	None
CTVP39C	In-house	2013	CT	N/A	USA	C	36	None
CTVP40C	In-house	2013	N/A	N/A	USA	C	36	None
CTVP41C	In-house	2013	CT	MA or PEI	USA, Canada	C	36	Vipa36
CTVP42C	In-house	2013	CT	N/A	USA	C	36	None
CTVP44C	In-house	2013	CT	CT	USA	C	36	Vipa10290
CTVP45C	In-house	2013	CT	N/A	USA	C	36	None
EN2910	GCA_000877685.1	2000	WA	N/A	USA	C	36	Vipa10290
EN9701173	GCA_000877555.1	1997	WA	N/A	USA	C	36	Vipa10290
EN9901310	GCA_000877565.1	1999	WA	N/A	USA	C	36	Vipa10290
F11-3A	GCA_000707545.1	1988	WA	WA	USA	E	36	Vipa10290
F4395	GCA_001608685.1	2006	BC	N/A	Canada	C	36	None
G25	GCA_002198475.1	2012	Galicia	Galicia	Spain	C	36	None
G30	GCA_002198555.1	2012	Galicia	Galicia	Spain	C	36	None
G31	GCA_002198565.1	2012	Galicia	Galicia	Spain	C	36	None
G36	GCA_002198575.1	2012	Galicia	Galicia	Spain	C	36	None
G37	GCA_002198445.1	2012	Galicia	Galicia	Spain	C	36	None
H11523	GCA_001608595.1	2006	BC	N/A	Canada	C	36	None
H18983	GCA_001608895.1	2006	BC	N/A	Canada	C	36	None
H64024	GCA_001608615.1	2006	BC	N/A	Canada	C	36	None
JJ21-1C	GCA_002072885.1	1990	WA	WA	USA	E	38	Vipa10290
K1461	GCA_000958575.1	2004	MA	N/A	USA	C	36	None
M59787	GCA_001608545.1	2006	BC	N/A	Canada	C	36	None
MAVP-1	In-house	2013	MA	VA	USA	C	36	None
MAVP-2	In-house	2013	MA	MA, VA or WA	USA	C	36	Vipa26
MAVP-6	In-house	2013	MA	MA	USA	C	36	Vipa26
MAVP-7	In-house	2013	MA	MA, ME or NB	USA, Canada	C	36	None
MAVP-8	In-house	2013	MA	MA or	USA	C	36	Vipa26

				ME				
MAVP-9	In-house	2013	MA	N/A	N/A	C	36	Vipa26
MAVP-14	In-house	2013	N/A	Multiple	USA; Canada	C	36	None
MAVP-16	In-house	2013	MA	N/A	N/A	C	36	Vipa26
MAVP-19	In-house	2013	MA	CT	USA	C	36	None
MAVP-20	In-house	2013	MA	MA	USA	C	36	None
MAVP-23	In-house	2013	MA	VA	USA	C	36	None
MAVP-24	In-house	2013	MA	MA	USA	C	36	Vipa26
MAVP-26	In-house	2013	MA	MA	USA	C	36	Vipa26
MAVP-29	In-house	2013	MA	Multiple	USA, Canada	C	36	None
MAVP-31	In-house	2013	MA	CT	USA	C	36	Vipa36
MAVP-36	In-house	2013	MA	KB, MA	USA	C	36	Vipa36
MAVP-37	In-house	2013	MA	MA	USA	C	36	Vipa26
MAVP-38	In-house	2013	MA	VA or MA	USA	C	36	Vipa36
MAVP-40	In-house	2013	CT or VA	N/A	USA	C	36	Vipa26
MAVP-41	In-house	2013	MA	N/A	N/A	C	36	Vipa26
MAVP-45	In-house	2013	MA	KB, MA	USA	C	36	Vipa26
MAVP-48	In-house	2013	MA	KB, MA	USA	C	36	Vipa36
MAVP-54	In-house	2013	MA	KB, MA	USA	C	36	Vipa26*
MAVP-60	In-house	2014	MA	KB, MA	USA	C	36	Vipa36
MAVP-61	In-house	2014	MA	N/A	USA	C	36	Vipa36
MAVP-62	In-house	2014	MA	KB, MA	USA	C	36	Vipa36
MAVP-63	In-house	2014	MA	N/A	USA	C	36	Vipa36
MAVP-64	In-house	2014	MA	N/A	USA	C	36	None
MAVP-70	In-house	2014	MA	MA or VA	USA	C	36	Vipa26
MAVP-72	In-house	2014	MA	MA	USA	C	36	None
MAVP-77	In-house	2014	MA	MA or BC	USA, Canada	C	36	None
MAVP-79	In-house	2014	MA	KB, MA	USA	C	36	Vipa36
MAVP-80	In-house	2015	MA	KB, MA	USA	C	36	Vipa36
MAVP-81	In-house	2015	MA	MA or PEI	USA, Canada	C	36	None
MAVP-82	In-house	2015	MA	KB, MA	USA	C	36	Vipa36
MAVP-83	In-house	2015	MA	MA	USA	C	36	Vipa26
MAVP-84	In-house	2015	MA	KB, MA	USA	C	36	Vipa36
MAVP-85	In-house	2015	MA	MA	USA	C	36	Vipa36
MAVP-86	In-house	2015	MA	CA	USA	C	36	None
MAVP-88	In-house	2015	MA	N/A	USA	C	36	Vipa26
MAVP-89	In-house	2015	MA	N/A	N/A	C	36	Vipa36
MAVP-92	In-house	2015	MA	KB, MA	USA	C	36	Vipa36
MAVP-97	In-house	2015	MA	MA	USA	C	36	Vipa26
MAVP-101	In-house	2015	MA	KB MA	USA	C	36	Vipa36
MAVP-	In-house	2015	MA	KB MA	USA	C	36	Vipa36

102								
MAVP-103	In-house	2015	MA	KB MA	USA	C	36	Vipa36
MAVP-104	In-house	2015	MA	MA	USA	C	36	Vipa26
MAVP-105	In-house	2015	MA	MA	USA	C	36	Vipa26
MAVP-111	In-house	2015	MA	MA	USA	C	36	Vipa26
MAVP-114	In-house	2015	MA	N/A	USA	C	36	Vipa26
MAVP-123E	In-house	2015	N/A	MA	USA	E	36	Vipa36
MAVP-127	In-house	2016	MA	MA	USA	C	36	Vipa36
MAVP-V	In-house	N/A	MA	N/A	N/A	C	36	None
MDVP46	GCA_000706905.2	2013	N/A	N/A	USA	C	36	None
MDVP50	GCA_002072875.1	2013	N/A	N/A	USA	C	36	None
MDVP51	GCA_002072835.1	2013	N/A	N/A	USA	C	36	None
MEVP-4	In-house	2013	ME	MA	USA	C	36	Vipa26
MEVP-10	In-house	2014	ME	N/A	N/A	C	36	Other
MEVP-11	In-house	2015	BC	N/A	Canada	C	36	None
NHVP-2	In-house	2013	ME	N/A	USA	C	36	None
NHVP-3	In-house	2013	NH	VA	USA	C	36	Vipa26
NY-3438	GCA_000707565.1	1998	NY	N/A	USA	C	36	Vipa10290
O29-1b	GCA_000707045.1	1997	OR	WA	USA	E	36	Vipa10290
PNUSAV000201	SRR7551165	2018	CT	N/A	USA	C	36	None
PNUSAV000220	SRR7586849	2018	CT	N/A	USA	C	36	None
S037	GCA_000491435.1	1994	N/A	N/A	N/A	N/A	36	Vipa10290
VP12	GCA_000707225.2	2012	MD	NY	USA	C	36	None
VP30	GCA_000706925.2	2013	MD	VA	USA	C	36	None
VP32	GCA_000707325.2	2013	MD	NJ	USA	C	36	None
VP33	GCA_000707845.2	2013	MD	N/A	USA	C	36	None
VP36	GCA_000707865.1	2013	MD	NY	USA	C	36	None
VP38	GCA_000706845.2	2013	MD	CT	USA	C	36	None
VP40	GCA_000706865.2	2013	MD	CT	USA	C	36	None
VP42	GCA_000706965.1	2013	MD	N/A	USA	C	36	None
VP43	GCA_000707205.2	2013	MD	CT	USA	C	36	None
VP43-1A	GCA_002072915.1	1992	N/A	N/A	USA	E	36	Vipa10290
VpG-1	GCA_002198305.1	2006	NY	NY	USA	C	36	None
VpG-3	GCA_002198295.1	2012	MA	NY	USA	C	36	None
VpG-4	GCA_002198335.1	2012	MA	NY	USA	C	36	None

VpG-7	GCA_002198075.1	2012	CA	CA	USA	C	36	None
VpG-8	GCA_002198055.1	2012	CA	CA	USA	C	36	None
VpG-9	GCA_002198175.1	2012	CA	NY	USA	C	36	None
VpG-10	GCA_002198135.1	2012	CT	CT	USA	C	36	None
VpG-11	GCA_002198195.1	2007	GA	WA	USA	C	36	None
VpG-12	GCA_002198185.1	2012	WA	WA	USA	C	36	None

Table S2.3. List of all non-ST36 strains used in this study. All strains used in this study including NCBI assembly ID unless sequenced in-house, year and geographic region collected or reported, clinically or environmentally isolated (C/E respectively) sequence type and phage type. N/A for any category indicates the information was not available.

Strain	Assembly ID	Year	Reporting Location	Country	C/E	ST	Phage
04-2192	GCA_001609555.1	2004	N/A	South Korea	C	629	Other
07-1339	GCA_000972045.1	2007	British Columbia	Canada	C	3	f237
08-0278	GCA_000960645.1	2008	Alberta	Canada	C	216	None
090-96	GCA_000701045.1	1996	N/A	Peru	C	265	None
49	GCA_000877625.1	2007	Washington	USA	E	137	Other
50	GCA_000519385.1	2006	N/A	USA	C	34	None
271	GCA_003408935.1	2015	N/A	China	E	490	Other
605	GCA_000519365.1	2006	Washington	USA	E	3	f237
861	GCA_000524535.1	2007	Washington	USA	E	3	f237
863	GCA_000877485.1	2007	Washington	USA	E	3	f237
930	GCA_000877475.1	2007	Washington	USA	E	3	f237
949	GCA_000454455.1	2006	Washington	USA	C	3	f237
3259	GCA_000454245.1	N/A	N/A	N/A	N/A	479	None
3355	GCA_000877615.1	2007	N/A	USA	C	65	None
3631	GCA_000877595.1	2007	Washington	USA	C	417	Other
3644	GCA_000877755.1	2007	Washington	USA	C	43	Vipa700
3646	GCA_000877765.1	2007	Washington	USA	C	417	None
10292	GCA_000707245.1	1997	Washington	USA	C	50	None
22702	GCA_000958645.1	1998	Georgia	USA	E	NF	Other

04-2549	GCA_000951795.1	2004	Saskatchewan	Canada	C	3	f237
04-2551	GCA_000975195.1	2004	Ontario	Canada	C	3	f237
07-2965	GCA_000960565.1	2007	Saskatchewan	Canada	C	326	None
09-3217	GCA_000960655.1	2009	British Columbia	Canada	C	43	Vipa4291
09-3218	GCA_000974905.1	2009	British Columbia	Canada	C	417	Other
10-4243	GCA_000972105.1	2006	British Columbia	Canada	C	141	Other
10-4244	GCA_001006125.1	2006	British Columbia	Canada	C	141	Other
10-4251	GCA_001006105.1	2006	British Columbia	Canada	C	3	f237
10-4287	GCA_001006185.1	2003	British Columbia	Canada	C	50	None
901128	GCA_000877675.1	1997	N/A	USA	C	135	Other
09-4434	GCA_000972125.1	2009	Alberta	Canada	C	417	Other
09-4435	GCA_000960685.1	2009	British Columbia	Canada	C	3	f237
09-4660	GCA_000972055.1	2009	British Columbia	Canada	C	417	Other
09-4661	GCA_001559885.1	2009	British Columbia	Canada	C	417	Other
09-4663	GCA_001006115.1	2009	British Columbia	Canada	C	417	Other
09-4664	GCA_000972035.1	2009	British Columbia	Canada	C	417	Other
09-4681	GCA_000972025.1	2009	New Brunswick	Canada	C	632	None
10-7205	GCA_001006205.1	2008	British Columbia	Canada	C	417	Other
08-7626	GCA_000960665.1	2008	Alberta	Canada	C	417	Other
13-028-A3	GCA_000737635.1	2013	N/A	Vietnam	E	1166	None
239-09	GCA_001634245.1	2009	Lambayeque	Peru	C	120	VipaP306
281-09	GCA_001634105.1	2009	N/A	Peru	C	120	VipaP306
283-09	GCA_001634165.1	2009	N/A	Peru	C	120	VipaP306
285-09	GCA_001633855.1	2009	N/A	Peru	C	120	VipaP306
98-513-F52	GCA_000707605.1	1998	Louisiana	USA	E	34	None
A1EZ952	GCA_001609185.1	2001	British Columbia	Canada	C	43	Vipa71
A4EZ700	GCA_001559805.1	2004	British Columbia	Canada	C	43	Vipa700

AN-5034	GCA_000182385.1	1998	N/A	Bangla-desh	C	3	f237
AQ3810	GCA_000154045.1	1983	N/A	Japan	C	87	None
AQ4037	GCA_000182365.1	1985	N/A	Japan	C	96	Other
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	Mozam-bique	C	3	f237
BB22OP	GCA_000328405.1	1982	N/A	Bangla-desh	E	88	None
C220-09	GCA_001634105.1	2009	Lambaye que	Peru	C	120	VipaP306
C224-09	GCA_001634125.1	2009	Lambaye que	Peru	C	120	VipaP306
C226-09	GCA_001634185.1	2009	N/A	Peru	C	120	VipaP306
C235	GCA_001634215.1	2009	Cajamarc a	Peru	C	120	VipaP306
C237	GCA_001634015.1	2009	Cajamarc a	Peru	C	120	VipaP306
C244-09	GCA_001634205.1	2009	N/A	Peru	C	120	VipaP306
CT20E	In-house	2013	Connecti cut	USA	E	28	Other
CT24E	In-house	2013	Connecti cut	USA	E	1136	Other
CT4264	In-house	2013	Connecti cut	USA	E	2033	Other
CT4287	In-house	2013	Connecti cut	USA	E	674	Vipa4291
CT4291	In-house	2013	Connecti cut	USA	E	674	Vipa4291
CTVP19C	NKGU00000000	2013	Massach usetts	USA	C	34	None
CTVP29C	In-house	2013	Connecti cut	USA	C	NF	None
CTVP35C	NIXS00000000	2013	Connecti cut	USA	C	194	None
EKP-008	GCA_000510585.1	2007	N/A	Banglad esh	E	479	None
EKP-021	GCA_000571915.1	2008	N/A	Banglad esh	E	3	f237
EKP-026	GCA_000525005.1	2008	N/A	Banglad esh	E	3	f237
EKP028	GCA_000522005.1	2008	N/A	Banglad esh	E	3	f237
EN9701072	GCA_000877715.1	1997	Washingt on	USA	C	43	Vipa71

EN9701121	GCA_000877725.1	1997	Washington	USA	C	50	None
FDA_R31	GCA_000430405.1	2007	Louisiana	USA	N/A	23	Other
FDAARGOS_53	GCA_001188035.2	2004	Alaska	USA	E	59	Other
FIM-S1708+	GCA_000732985.1	2014	N/A	Mexico	E	1167	Other
FORC_004	GCA_001433415.1	N/A	N/A	N/A	N/A	1628	Other
FORC_006	GCA_001304775.1	2014	Gyeongnam	South Korea	N/A	1630	Other
FORC_008	GCA_001244315.1	N/A	N/A	N/A	N/A	984	Other
G8	In-house	2007	New Hampshire	USA	E	NF	Other
G61	In-house	2007	New Hampshire	USA	E	1125	Other
G79	In-house	2007	New Hampshire	USA	E	NF	None
G95	In-house	2007	New Hampshire	USA	E	2017	None
G145	In-house	2007	New Hampshire	USA	E	2018	Other
G149	MPPO00000000	2007	New Hampshire	USA	E	631	None
G151	In-house	2007	New Hampshire	USA	E	83	Other
G227	In-house	2007	New Hampshire	USA	E	1087	Other
G320	In-house	2008	New Hampshire	USA	E	NF	Other
G325	In-house	2008	New Hampshire	USA	E	NF	Vipa26
G360	In-house	2008	New Hampshire	USA	E	NF	None
G363	In-house	2008	New	USA	E	1574	Vipa26

			Hampshire				
G441	In-house	2008	New Hampshire	USA	E	2017	None
G445	In-house	2008	New Hampshire	USA	E	2019	None
G640	In-house	2008	New Hampshire	USA	E	NF	None
G650	In-house	2008	New Hampshire	USA	E	2020	None
G653	In-house	2008	New Hampshire	USA	E	2021	None
G729	In-house	2008	New Hampshire	USA	E	380	Other
G735	In-house	2008	New Hampshire	USA	E	NF	None
G755	In-house	2008	New Hampshire	USA	E	2021	None
G1286	In-house	2009	New Hampshire	USA	E	107	Other
G1334	In-house	2009	New Hampshire	USA	E	2026	None
G1350	In-house	2009	New Hampshire	USA	E	1574	Vipa26
G1355	In-house	2009	New Hampshire	USA	E	1847	Other
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	2031	Vipa26
G1449	In-house	2009	New	USA	E	NF	Vipa26

			Hampshire				
G1463	In-house	2009	New Hampshire	USA	E	NF	None
G1487	In-house	2009	New Hampshire	USA	E	2021	None
G3578	NIYO00000000	2013	New Hampshire	USA	E	674	Vipa4291
G3599	In-house	2013	New Hampshire	USA	E	674	Vipa4291
G3654	In-house	2013	New Hampshire	USA	E	1123	None
G4026	In-house	2013	New Hampshire	USA	E	773	None
G4186	NIYP00000000	2013	New Hampshire	USA	E	34	None
G6494	In-house	2015	New Hampshire	USA	E	1716	Other
G6499	In-house	2015	New Hampshire	USA	E	NF	Vipa26
G6928	In-house	2015	New Hampshire	USA	E	631	None
GCSL_R21	GCA_001726355.1	2007	Texas	USA	E	12	Vipa4291
Guillen-151-Peru	GCA_001633955.1	2009	N/A	Peru	N/A	120	VipaP306
Gxw_7004	GCA_001541615.1	2007	Guangxi	China	C	3	f237
Gxw_9143	GCA_001541625.1	N/A	N/A	N/A	N/A	NF	None
HS-06-05	GCA_001280705.1	N/A	N/A	N/A	N/A	614	Other
HS-13-1	GCA_001270125.1	N/A	N/A	N/A	N/A	676	Other
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	88	None

ISF-25-6	GCA_001267595.1	N/A	N/A	N/A	N/A	NF	Other
ISF-29-3	GCA_001273575.1	N/A	N/A	N/A	N/A	1518	Other
ISF-54-12	GCA_001280635.1	N/A	N/A	N/A	N/A	1631	None
ISF-77-01	GCA_001270285.1	N/A	N/A	N/A	N/A	NF	None
ISF-94-1	GCA_001280645.1	N/A	N/A	N/A	N/A	1632	None
J-C2-34	GCA_000958655.1	N/A	N/A	N/A	N/A	NF	Other
K1198	GCA_001188035.1	2004	Alaska	USA	E	59	Other
K1203	GCA_000707585.1	2004	Alaska	USA	E	59	Other
K23	GCA_001497485.1	N/A	N/A	N/A	N/A	1052	Other
K5030	GCA_000182465.1	2005	N/A	India	C	3	f237
K5073	GCA_001728275.1	2007	Maryland	USA	C	750	Other
M0605	GCA_000523375.1	2013	N/A	Mexico	E	539	Other
MA5	In-house	2013	Massachusetts	USA	E	2118	Other
MA58	In-house	2014	Massachusetts	USA	E	2119	Other
MA60	In-house	2014	Massachusetts	USA	E	2120	Other
MA76	In-house	2014	Massachusetts	USA	E	2120	Other
MA77	In-house	2014	Massachusetts	USA	E	2120	Other
MA78	In-house	2014	Massachusetts	USA	E	2119	Other
MA97	In-house	2014	Massachusetts	USA	E	2119	Other
MA143	In-house	2014	Massachusetts	USA	E	NF	Other
MA145	In-house	2014	Massachusetts	USA	E	1717	Other
MA146	In-house	2014	Massachusetts	USA	E	NF	Other
MA147	In-house	2014	Massachusetts	USA	E	1717	Other
MA157	In-house	2014	Massachusetts	USA	E	771	Other
MA161	In-house	2014	Massachusetts	USA	E	NF	Other
MA175	In-house	2014	Massachusetts	USA	E	1399	Vipa4291
MA239	In-house	2014	Massach	USA	E	2032	Other

			usetts				
MA271	In-house	2015	Massach usetts	USA	E	1185	Vipa4291
MA281	In-house	2015	Massach usetts	USA	E	1556	None
MA303	In-house	2015	Massach usetts	USA	E	NF	Other
MA304	In-house	2015	Massach usetts	USA	E	1556	None
MA371	In-house	2015	Massach usetts	USA	E	1185	Vipa4291
MA398	In-house	2015	Massach usetts	USA	E	2123	None
MA414	In-house	2015	Massach usetts	USA	E	NF	Other
MA432	In-house	2015	Massach usetts	USA	E	1185	Vipa4291
MA448	In-house	2015	Massach usetts	USA	E	1185	Vipa4291
MA459	In-house	2015	Massach usetts	USA	E	1727	None
MA561	MPPM00000000	2015	Massach usetts	USA	E	631	None
MAVP-4	MDWU00000000	2013	Massach usetts	MA	C	NF	None
MAVP-10	In-house	2013	Massach usetts	MA	C	1346	None
MAVP-13	NKGK00000000	2013	Massach usetts	MA	C	1719	Other
MAVP-14	NJAP00000000	2013	Massach usetts	USA	C	324	None
MAVP-21	NIXV00000000	2013	Massach usetts	MA	C	674	Other
MAVP-25	NJAN00000000	2013	N/A	N/A	C	1127	Other
MAVP-30	MDWW00000000	2013	Massach usetts	USA	C	631	Other
MAVP-39	MDWV00000000	2013	Massach usetts	USA	C	631	Other
MAVP-46	NJAO00000000	2013	Massach usetts	USA	C	110	None
MAVP-50	NIXY00000000	2013	N/A	N/A	C	636	None
MAVP-55	NKGI00000000	2013	Massach usetts	USA	C	632	Other
MAVP-56	MDWX00000000	2013	Massach usetts	USA	C	631	Other
MAVP-65	In-house	2014	Massach usetts	USA	C	1727	None
MAVP-66	NKGH00000000	2014	Massach	USA	C	2030	None

			usetts				
MAVP-67	NIXX00000000	2014	Massach usetts	USA	C	308	None
MAVP-69	In-house	2014	Massach usetts	USA	C	43	Vipa71
MAVP-71	NIXZ00000000	2014	Massach usetts	USA	C	43	Vipa71
MAVP-73	NKGG00000000	2014	Massach usetts	USA	C	1722	Other
MAVP-74	MDWY00000000	2014	Massach usetts	USA	C	631	Other
MAVP-75	MDWZ00000000	2014	Massach usetts	USA	C	631	Other
MAVP-76	NKGF00000000	2014	Massach usetts	USA	C	614	Other
MAVP-78	MDXA00000000	2014	Massach usetts	USA	C	631	Other
MAVP-99	NIXW00000000	2015	Massach usetts	USA	C	12	Other
MAVP-108	NIXQ00000000	2015	Massach usetts	USA	C	1716	Other
MAVP-A	MDWP00000000	2010	Massach usetts	N/A	C	631	None
MAVP-E	LBHP00000000	2010	Massach usetts	USA	C	631	None
MAVP-F	NKFY00000000	2011	Massach usetts	USA	C	1726	None
MAVP-G	NIXR00000000	2011	Massach usetts	USA	C	809	Other
MAVP-H	In-house	2011	Massach usetts	USA	C	636	None
MAVP-I	In-house	2011	Massach usetts	USA	C	1126	Vipa3
MAVP-J	NKFX00000000	2011	Massach usetts	USA	C	1727	None
MAVP-K	In-house	2011	Massach usetts	USA	C	8	VipaK
MAVP-L	MDWS00000000	2011	Massach usetts	USA	C	631	None
MAVP-M	In-house	2011	Massach usetts	USA	C	1127	None
MAVP-N	In-house	2011	Massach usetts	USA	C	NF	Other
MAVP-P	MDWQ00000000	2010	Massach usetts	USA	C	631	Other
MAVP-Q	MDWT00000000	2011	Massach usetts	USA	C	631	None
MAVP-R	MPPP00000000	2011	Massach	USA	C	631	None

			usetts				
MAVP-S	In-house	2011	Massach usetts	USA	C	1728	Other
MAVP-T	MDWR00000000	2010	Massach usetts	USA	C	631	None
MAVP-U	NKFV00000000	2011	Massach usetts	USA	C	749	Other
MAVP-W	In-house	N/A	Massach usetts	N/A	C	43	Vipa71
MAVP-X	NKFU00000000	2011	Massach usetts	USA	C	322	Vipa3
MDOH-04-5M732	In-house	2004	Florida	USA	C	3	Vipa3
MEVP-5	In-house	2013	Maine	USA	C	1716	Other
MEVP-6	In-house	2013	Maine	USA	C	1729	None
MEVP-7	NKFQ00000000	2014	Maine	USA	C	1205	Other
NA4	GCA_002880475.1	2017	N/A	Malaysia	E	1911	Other
NA7	GCA_002880415.1	2017	N/A	Malaysia	E	1911	Other
NA9	GCA_002880435.1	2017	N/A	Malaysia	E	1911	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
NIHCB0603	GCA_000454265.1	2006	N/A	Bangla-desh	C	3	f237
NIHCB0757	GCA_000477475.1	2006	N/A	Bangla-desh	C	65	None
NSV_7536	GCA_001471485.1	N/A	N/A	N/A	N/A	50	None
P306	GCA_001633935.1	2009	N/A	Peru	E	120	VipaP306
Peru288	GCA_000522065.1	2001	N/A	Peru	C	3	f237
peru466	GCA_000182345.1	1996	N/A	Peru	C	3	f237
PIURA-17	GCA_001633985.1	2009	Piura	Peru		120	VipaP306
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/	3	None

					A		
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	137	None
RM-14-5	GCA_001273555.1	N/A	N/A	N/A	N/A	1663	None
RM-17-6	GCA_001267655.1	N/A	N/A	N/A	N/A	1346	None
S096	GCA_000493225.1	1999	N/A	South Korea	C	217	Other
S158	GCA_000489315.1	2006	N/A	China	E	419	Other
S161	GCA_000489255.1	2006	N/A	China	E	419	Other
S167	GCA_000489135.1	2007	N/A	China	E	490	Other
S176-10	GCA_001280725.1	N/A	N/A	N/A	N/A	1728	None
S195-7	GCA_001268005.1	N/A	N/A	N/A	N/A	1187	None
S349-10	GCA_001268015.1	2010	N/A	Canada	E	1516	Other
S357-21	GCA_001273635.1	N/A	N/A	N/A	N/A	102	None
S372-5	GCA_001280655.1	N/A	N/A	N/A	N/A	324	None
S383-6	GCA_001267625.1	2011	N/A	Canada	N/A	1134	Other
S439-9	GCA_001270155.1	N/A	N/A	N/A	N/A	1155	None
S440-7	GCA_001270235.1	N/A	N/A	N/A	N/A	34	None
S448-16	GCA_001267635.1	2012	N/A	Canada	N/A	1134	Other
S456-5	GCA_001268045.1	2012	N/A	Canada	N/A	NF	Other
S487-4	GCA_001270215.1	N/A	N/A	N/A	N/A	631	None
S499-7	GCA_001270145.1	2013	N/A	Canada	E	1134	Other
SG176	GCA_000958565.1	2006	Georgia	USA	E	NF	Other
SNUVpS-1	GCA_000315135.1	2012	N/A	Korea	E	917	Other
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	GCA_000591495.1	N/A	N/A	N/A	N/	413	None

_D06_S3					A		
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	Other
UCM-V493	GCA_000568495.1	2002	N/A	Spain	E	471	None
v110	GCA_000388025.1	2010	Hong Kong	China	E	809	Other
V14-01	GCA_000558885.1	2001	N/A	Chile	C	3	f237
V223-04	GCA_000558905.2	2004	N/A	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	Other
VIP4-0444	GCA_000500485.1	N/A	N/A	N/A	N/A	2165	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	Other
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	34	None
VP8	GCA_000707425.1	2012	Maryland	USA	C	631	None
VP9	GCA_000707385.1	2012	Maryland	USA	C	631	None
VP10	GCA_000707125.2	2012	Maryland	USA	C	43	Vipa71
VP11	GCA_000707105.1	2012	Maryland	USA	C	1116	None

VP13	GCA_000707685.1	2012	Maryland	USA	C	678	Other
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	VipaK
VP20	GCA_000707825.1	2010	Maryland	USA	C	8	VipaK
VP21	GCA_000707645.1	2010	Maryland	USA	E	8	VipaK
VP22	GCA_000707905.1	2010	Maryland	USA	E	676	None
VP23	GCA_000707665.1	2010	Maryland	USA	E	8	VipaK
VP24	GCA_000707265.1	2010	Maryland	USA	E	8	VipaK
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	VipaK
VP31	GCA_000707445.1	2013	Maryland	USA	C	631	Other
VP34	GCA_000707005.1	2012	Maryland	USA	C	653	None
VP35	GCA_000707465.1	2013	Maryland	USA	C	631	None
VP39	GCA_000706985.1	2013	Maryland	USA	C	896	Other
VP41	GCA_000707485.1	2013	Maryland	USA	C	631	None
VP44	GCA_000707505.1	2013	Maryland	USA	C	631	None
VP45	GCA_000706885.1	2013	Maryland	USA	C	631	None
Vp47	GCA_002153925.1	2012	N/A	China	E	1772	Other
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
Vp294	GCA_002154015.1	2013	N/A	China	E	NF	Other
VP766	GCA_000877605.1	2007	Washington	USA	E	133	None
VP551	GCA_000877415.1	2007	Washington	USA	E	3	f237
VP2007-007	GCA_000558925.1	2007	Mississippi	USA	E	306	Other
VP2007-095	GCA_000454165.1	2007	Florida	USA	C	631	None
VP-48	GCA_000593285.1	1996	N/A	India	C	152	f237
VP43-1A	GCA_002072915.1	1992	N/A	USA	E	39	Vipa10290
VPCR-2009	GCA_000593305.1	N/A	N/A	N/A	N/A	1567	Other
VPCR-2010	GCA_000454475.1	2010	N/A	USA	E	308	Other
VP-NY4	GCA_000454145.1	1997	N/A	India	C	3	f237
VPTS-	GCA_000593325.1	N/A	N/A	N/A	N/	1013	None

2009					A		
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
W90A	GCA_002072845.1	1986	Washington	USA	E	21	Other

CHAPTER TWO

The influence of the filamentous phage Vipa26 on the ecological fitness of an invasive *Vibrio parahaemolyticus* ST36 clinical isolate originating in the Gulf of Maine

Jillian Means[2], Wayne Faegerberg[2], Cheryl 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

Vibrio parahaemolyticus, a common inhabitant of coastal waters, is the leading cause of bacterial seafood-borne illnesses in the United States of America. The incursion of the pathogenic lineage sequence type (ST) 36 into US North Atlantic Coastal waters from the Pacific Northwest has contributed greatly to disease burden in the region. A clonal clade of ST36 with high clinical prevalence in the Gulf of Maine (GOM) basally acquired a filamentous bacteriophage of the family *Inoviridae* (inoviruses) and held the prophage as its population clonally expanded into the GOM. In this work we investigate the potential impact of this Inovirus, Vipa26, on the fitness of a ST36 clinical isolate. ST36 strains harboring Vipa26 integrated into its chromosome I *dif* site actively produce progeny, capable of infecting new susceptible hosts. Curiously, infections are reversible, and spontaneous phage loss occurred under some culture conditions. Though sustained infection by Vipa26 were benign and did not discernably impair growth, new infections were immediately detrimental and abruptly impaired growth observed as both turbid plaques and an early plateau during exponential growth. Sustained phage integration protected ST36 strains from detrimental effects of superinfections by related phages. Furthermore phage-harboring strains were more fit in direct competitions with an isogenic susceptible host likely due to phage predation of the susceptible competitor. Although natural phase variation

between non-isogenic phage-harboring and phage deficient ST36 strains used to examine environmental fitness effects of phage carriage impacted strain fitness confounded our analyses of potential advantages of phage in natural seawater microcosms, this model for environmental fitness is a potential avenue for further investigation. The protection of the host from superinfection by similar phages, and increased fitness in direct competition against a susceptible strain argue that the acquisition of Vipa26 early in the establishment of the GOM clonal population may have contributed to its successful expansion. Further insight into the mechanism of infection of Vipa26 and impact on the host cell will expand our understanding of the complex interplay between bacteriophages and their hosts.

Introduction

V. parahaemolyticus, a ubiquitous estuarine bacterium, is a major cause of food-borne illness in the United States, accounting for about 45,000 reported illnesses a year (10). Historically, *V. parahaemolyticus* was of little concern in the Northeast US; however, compared to the early 2000s the number of cases traced to Northeast product has increased, with a major outbreak in 2013 (33, 75). This was in part due to the introduction of an invasive virulent sequence type (ST) endemic to the Pacific Northwest, ST36, a lineage that caused the 2013 outbreak and has persisted in several harvest areas causing recurrent infections (75, 77, 80).

ST36 was likely introduced into the Atlantic Ocean multiple times (76), with several clades established in various regions along the US East Coast, including in the Gulf of Maine (GOM) and the Long Island Sound (LIS) with a clonal subpopulation in Katama Bay (KB) (Chapter 2). Whereas ST36 was traced to several other Atlantic coastal locations including Spain and the US mid-Atlantic including Virginia and New Jersey, disease incidence from these other locations remains low suggesting ST36 may not have successfully established robust, persistent populations everywhere it invaded (76) (Chapter 2). However, the two most northern populations of ST36 in the GOM and KB continue to cause illnesses.

Prior to clonal expansion and establishment of the GOM and KB sub-populations, the related but genetically unique progenitors independently acquired and held distinct related filamentous phages (Vipa26 in the GOM population and Vipa36 in the KB population) of the family Inoviridae (Chapter 2 A similar but distinctive phage (Vipa10290) consistently associates with clinical isolates sourced to the Pacific in prior decades, known as the old Pacific Northwest (PNW) population (76); however, ST36 isolates traced to other Atlantic coastal locations, and recent clinical isolates originating in the Pacific largely lack Inoviruses (76) (Chapter 2). It is notable that the clonal complex of pandemic strains originating in Asia, ST3, that spread globally and still causes the most illnesses worldwide, commonly identified as serotype O3:K6, also carries an Inovirus related to those in *V. parahaemolyticus* ST36 (69, 74, 149) (see Chapter 2). The ST3-associated phage, called f237, is nearly universally associated with this highly successful invasive pathogen and is diagnostic for this lineage (69, 74, 149). Whereas intra-oceanic movement of *V. parahaemolyticus* can occur naturally, ST36 is remarkable in that it not only underwent inter-oceanic incursion but established populations in several new locations, similarly to the pandemic ST3 (76, 78). ST3 acquired f237 basally prior to its pandemic spread, in contrast, ST36 phage acquisition occurred independently by only some members of the population and these phage-harboring lineages have been particularly successful in their new locations. The presence of these phages in these highly successful pathogenic lineages suggests that the phages could influence the ecological fitness or virulence of their hosts.

Filamentous phages of the family Inoviridae have diverse impacts on their bacterial hosts (92). They do not typically kill the host cell during infection, but instead produce a chronic infection, reproducing and extruding progeny phage particles from a living cell (181). Thus, phage infected-cells likely incur a fitness cost when producing phage including membrane stress from phage secretion (182, 183). This growth cost may be balanced by potential benefits to the host. For example, a persistently infected cell may also infect susceptible competitors who would incur the fitness cost thereby leveling

the playing field; although, based on distribution of phage in natural communities, it is unlikely that all *V. parahaemolyticus* are susceptible (see Chapter 2). Furthermore, persistent infection could protect hosts from other related, and more-costly phage, including lytic dsDNA phage (162) and integration of the phage into the chromosome often down-regulates production of phage particles, minimizing the impact on the host cell (184, 185). These phages are also instruments of horizontal gene transfer, sometimes containing accessory content that confers new and useful traits upon their hosts, such as access to novel environments (92). For instance, in addition to toxins encoded as part of the core genome of the *Vibrio cholerae* filamentous Inoviridae phage, CTXϕ, which include accessory cholera enterotoxin (*ace*) and zonula occludens toxin (*zot*), CTXϕ encodes cholera toxin (*ctx*) as part of its accessory genome (12, 143). The phages in *V. parahaemolyticus* that are related to CTXϕ including f237, Vipa26, Vipa36, Vipa10290 and others, also encode *ace* and *zot* and vary in accessory content from each other (Chapter 2). Accessory toxins produced by pathogenic variants of *V. parahaemolyticus* not only confer increased virulence but also may deter predation by bacterivorous protists and amoebae (66, 186), suggesting phage encoded toxins may confer ecologically relevant benefits. Filamentous phages are also important in the evolution of pathogens; the phage RS1 can cause CTXϕ to excise from the chromosome, skewing the population towards nontoxigenic derivatives and potentially attenuating epidemics (145). Despite the well-established precedent for phage influencing virulence and population dynamics of other human pathogenic bacteria (187), studies to investigate the potential impact of f237 on ST3 are lacking, although one report speculated that ORF8 may encode an adhesive protein involved in gut colonization, and another suggested a potential role of UV induction of f237 on decline of ST3 in South America (74, 150, 188).

The presence of Vipa26 and Vipa36 phages among clinical samples derived from two successful environmental populations is likely not serendipity. The potential costs and benefits associated with these phages have not been quantified and the contribution of phage-encoded attributes that could

promote environmental fitness or virulence have not been examined. In this study, we focus on Vipa26, the phage associated with the successful GOM sub-population of ST36, to determine the dynamics and impact of phage infection on its host including competitive fitness in culture and microcosms, and for resistance to predation. These studies aim to provide additional understanding and insight into kinetics of infection by an active filamentous phage and the potential phenotypic differences conferred by this phage.

Methods

Bacterial strains, plasmids and culture conditions

Bacterial strains and plasmids used in this study are listed in Table 3.1. *V. parahaemolyticus* was routinely cultured at 37°C with aeration or at room temperature in static culture either in Heart Infusion broth (HI) for isolation of phage or in Luria Bertani (LB) amended with 10 mM $MgCl_2$ for phage infections, in LB amended with 3% NaCl and buffered with 50mM TrisHCl pH 7.4 (LBS) for routine culturing, and on LBS plates containing 1.5% agar for colony isolation. *Escherichia coli* was cultured in LB broth and agar. Media was supplemented with antibiotics for maintenance and selection of plasmids at the following concentrations: for *E. coli* - kanamycin (Kan) at 50 µg/mL, and chloramphenicol (Chl) at 25 µg/mL; for *V. parahaemolyticus* - Kan at 50 µg/mL, Chl at 2.5 µg/mL, and erythromycin (Erm) at 10 µg/mL. Solid media was supplemented with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) at a final concentration of 20µg/mL for evaluating presence or absence of betagalactosidase activity from a cloned *lacZ* gene for blue/white screening. *Cafeteria roenbergensis* ATCC 50301 was cultivated as recommended by ATCC at 25°C in 1525 seawater 802 medium which is 1x artificial seawater supplemented with 0.125% rye grass cerophyll (Wards Scientific) that was bacterized with *Enterobacter aerogenes* ATCC 13048. *Acanthamoeba castellanii* ATCC 30234 was maintained in axenic culture in medium 712: PYG with additives, at 25°C according to ATCC guidelines (189, 190).

Table 3.1 List of strains and plasmids used in this study.

Strain and plasmids	Notable characteristics	Citation
V. parahaemolyticus		
F11-3A	Clinical, ST36, opaque; Vipa10290	GCA_000707545.1
G149	Environmental, ST631, opaque	This study
G320	Environmental, STNF, opaque, other phage	This study
G1445	Environmental, STNF, opaque	This study
MAVP-3	Clinical, ST3, opaque; f237	This study
MAVP-20	Clinical, ST36, translucent	This study
MAVP-20(26)	Clinical, ST36, opaque; Vipa26	This study
MAVP-20(26)-2	Clinical, ST36, translucent; Vipa26	This study
MAVP-26	Clinical, ST36, opaque; Vipa26	This study
MAVP-26::Erm	Clinical, ST36, opaque; Vipa26::Erm^R	This study
MAVP-26PD	Clinical, ST36, opaque	This study
MAVP-C	Clinical ST3, opaque, f237	This study
MAVP-V	Clinical, ST36, opaque	This study
MAVP-V(26)	Clinical, ST36, opaque; Vipa26	This study
E. coli		
Top10	F- *mcrA Δ*(*mrr-hsdRMS-mcrBC*) ϕ80 *lacZΔ M15 Δ lacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL* (Sm^R) *endA1 nupG*	Invitrogen, Carlsbad, CA
DH5αλpir	*supE44 ΔlacU169 (*ϕ*lacZ*ΔM15) *recA1 endA1 hsdR17* *thi-1 gyrA96 relA1*; λpir phage lysogen	(191)
NEB10β	*Δ(ara-leu)7697 araD139 fhuA ΔlacX74 galK16 galE15 e14*-ϕ80 *dlacZ ΔM15 recA1relA1 endA1 nupG rpsL* (Sm^R) *rph spoT1 Δ(mrr-hsdRMS-mcrBC)*	New England Biolabs, Ipswich, MA
Plasmids		
pEVS104	R6Kγ oriV, Kan^R RP4-derived conjugative plasmid	(192)
pEVS79	Vibrio suicide cloning vector, Chl^R, Tc^R	(193)
pSEElostfoX	pEVS79 containing the *tfoX* gene from *V. parahaemolyticus* MAVP-26 under the control of the pBAD promoter and induced by growth on arabinose through the araC regulator; Chl^R	Courtesy of S.E. Eggert
pVSV103	Blue on X-gal, Kan^R, *lac-Z*	(194)
pCAW7B1	White on X-gal, pVSV103 containing *lacZ* Δ147–1080 bp; Kan^R	(195)

pCR2.1-TOPO	PCR product cloning vector, Kan^R, Amp^R	Invitrogen, Carlsbad, CA
pJM3F1	pCR2.1-TOPO containing *tlh* qPCR target amplicon, Kan^R	This study
pJM3F3	pCR2.1-TOPO containing 5' Vipa26 prophage junction qPCR target amplicon, Kan^R	This study
pJM3F4	pCR2.1-TOPO containing 3' Vipa26 prophage junction qPCR target amplicon, Kan^R	This study
pJM3F6	pCR2.1-TOPO containing no phage *dif* site qPCR target amplicon, Kan^R	This study
pJM3G1	pCR2.1-TOPO containing Vipa26 replicative form junction qPCR target amplicon, Kan^R	This study
pVCW18	pCR2.1-TOPO containing Erm^R cassette flanked by homologous regions of phage genome, Kan^R	This study

Phage isolation, imaging, and infection

To produce Vipa26 for visualization, 200mL cultures of MAVP-26 were grown in HI broth at 37°C with aeration to stationary phase (~18 hours), cells were pelleted by centrifugation at 4500g for 10 minutes and the supernatant further cleared with a second centrifugation at the same conditions. Phage particles in the resultant supernatant were subsequently purified using a modified PEG-8000 precipitation method (196). Briefly, the top 160mL of supernatant was transferred to a new tube and 40mL 2.5M NaCl/20%(w/v) PEG-8000 solution was added and the mixture incubated overnight at 4°C. Subsequently phage were pelleted at 12000g for 15 minutes and the pellet resuspended in 10mL SB buffer. The precipitation was repeated with 2mL of NaCl/PEG-8000 solution added to the resuspended pellet, incubated on ice for 1hr and pelleted again at 12000g for 15 minutes. The final pellet was resuspended in 1mL SM buffer (100 mM NaCl, 25 mM Tris-HCl pH 7.5, 8 mM MgSO4, 0.01% (w/v) gelatin) and stored at 4°C. Four parallel phage precipitation preps were concentrated using an Amicon Ultra-4 10K filter (MilliporeSigma, Burlington, MA, USA) then washed and resuspended in 1mL 0.1 M ammonium acetate (pH 7.0) before being stained for 45-60 seconds with 5% uranyl acetate and imaged on carbon coated Formvar grids using a Zeiss/LEO 922 Omega TEM.

Phage presence and infection of susceptible strains was determined by a double overlay plaque assay as previously described (74, 149) or by comparing the growth kinetics, as estimated by the optical densities (OD_{600}) of cultures grown in phage-conditioned media. Briefly, bacterial cells were embedded in agar by combining 120µL of bacterial culture grown in HI medium to 0.5 OD_{600} (approximately 1.5 hr) with 7mL LB broth amended with 10mM $MgCl_2$ and 0.6% agar held at 50°C and poured onto an LB agar plate. Bacterial-free phage suspensions were produced from phage-harboring strains grown overnight in HI as described above, and cell-free supernatant was generated by pelleting cells, and then filtering the supernatant with a 0.45 µm PES filter. The absence of contaminating bacteria was confirmed by lack of turbidity after overnight incubation, and by plating filtrates on LBS agar. A total of 20µl were subsequently spotted onto the soft agar overlay and the plates incubated at 37ºC until turbid plaques were observed (~ 3hrs).

Kinetics of growth of isolates either newly- and stably-phage infected

To determine impact of phage on growth rates, strains with or without phage were grown in HI, LBS and Hepes Minimal Media (HMM) in triplicate at 37°C with shaking and OD_{600} was determined in a Tecan Infinite plate reader for 12 hours. Concurrent OD_{600} readings and CFU/mL determined by serially diluting in Instant Ocean and plating on LBS were compared to determine the strain specific correlation of OD_{600} to CFU. Specific growth rate K was calculated using OD data from the log phase batch culture and compared between strains and conditions. Specific growth rates were assessed for significance using Tukey HSD test to compare means.

For detection of phage by their impact on growth kinetics, specifically the early plateau during exponential growth, LB supplemented with 10mM $MgCl_2$ was conditioned with 2-10% phage conditioned media produced as described above (depending on the experiment and predicted yield based on plaque assays) or similarly conditioned with broth from phage free-supernatant. 1mL of LB was

inoculated with 18µL of bacterial culture grown in HI medium to OD_{600} 0.5, and three to five 200 µL replicates aliquoted in each well of a 96-well polystyrene microplate. The growth of bacterial cultures was estimated by kinetic measurements of OD_{600} at 28°C, every 20 minutes for 18-36 hours in a Tecan M200 plate reader. Presence of phage was determined by the characteristic early plateau of growth by susceptible cultures (e.g. MAVP-V) after ~7-8 hours (OD_{600} ~0.6) in cultures grown in broth similarly conditioned with phage-free supernatants. Samples were stored at -20°C for later qPCR analysis.

The relative abundance of phage produced by infected cells following different treatments was also estimated by comparison of growth kinetics of susceptible cells in broth conditioned with serial dilutions of phage-containing filtrates. Briefly, following experimental treatment, phage-harboring cells were grown in HI medium over-night and the final cell densities the cultures normalized to an OD_{600} of 3.0 in HI broth. Following cell removal and filtration, 10µl of normalized phage suspensions was added to 230 µl of LB with 10mM $MgCl_2$ that was pre-inoculated with a phage susceptible strain (MAVP-26PD) as described above, and 1:5 serial dilutions of the phage performed by sequential transfer of 40µl into 200 µl of the neighboring well. Growth was monitored as described above and the change in growth kinetics compared for each dilution allowing the identification of dilutions exhibiting similar changes in growth kinetics.

Spontaneous loss of Vipa26

To evaluate the presence of phage in archived stocks, these stocks were directly plated onto LBS agar and lysates prepared from the resulting 20-30 individual colonies. Lysates were produced by suspending the cells in diH_20 and boiling. A total of 0.7 µl of the lysate was screened by PCR-amplification. The oligonucleotide primers ST36PhageF2 and ST36PhageR2 were used to specifically amplify ORF3-ORF5 on the phage genome and multiplexed with an internal control, the primers tlh-F2 and tlh-R specific to *tlh* (see Table 3.2 and Chapter 2) with published cycling parameters (175). Presence

of phage was confirmed by the production of an amplicon of the correct size as visualized following separation on a 0.7% agarose gel containing 1x Gel Red. Subsequent analysis was conducted by multiplex PCR with three primers: two flanking the phage-integration *dif* site (dif_siteF and dif_siteR) which only produce an amplicon from phage-free chromosomes, and a third outward primer in Hypothetical Protein A that only produces an amplicon with the dif_SiteF primer when phage are integrated (NEHypR; Table 3.2). PCR parameters were as follows: an initial denaturation of 3min 94°C followed by 25 cycles of 1min 94°C, 30sec 55°C, and 1min 30sec at 72°C followed by a final elongation of 6min at 72°C. Amplicons were differentiated by size by separation on a 0.7% agarose gel containing 1x Gel Red, as compared to controls (MAVP-20 which is phage free, and MAVP-26 which contains an integrated phage).

In order to investigate the dynamics of phage loss in greater detail, lysates from 90 isolated colonies grown on agar directly from a single freezer stock were subsequently screened using the multiplex PCR specific for the Vipa26-integrated or phage-free *dif* site (NEHypR, dif_siteF and dif_siteR) as described above. Colonies with mixed phage/no phage genotype were re-streaked on LBS plates and screened over successive days (Figure 3.6). Phage deficient isolates were cryopreserved for future analysis. In parallel, phage content in liquid culture with passaging was examined. Specifically, 3mL of LBS broth was inoculated with five Vipa26 positive (confirmed by PCR) colonies (MAVP-26) in triplicate and grown for 6 hours at 37°C with shaking. Subsequently, a sterile loop was used to streak for isolated colonies on LBS agar and 30µl of the 6hr culture was subcultured into fresh LBS broth and grown overnight. Sampling and subculturing were repeated on day two and overnight to day three for a total of 5 time points. Individual colonies (~30 per replicate at each time point) were screened using the multiplex specific for integrated/excised prophage as above.

To determine the effects of prolonged growth on Vipa26 excision, colonies on a single plate were sampled over time and tested for phage content using the integrated/excised prophage multiplex.

MAVP-26 cultured on LBS agar incubated at 37°C for four days sealed with parafilm to prevent drying. Every day, lysates of five whole individual colonies were collected and frozen at -20°C prior to being tested with qPCR to quantitatively determine the abundance of Vipa26 in integrated and replicative form (see Fig. 3.1; Fig. 3.2 and Table 3.3; Table 3.4). Change of abundance of phage-free *dif* site over time was assessed for significance using Tukey Kramer's HSD.

Table 3.2. List of end point PCR primers used in this study. Reactions were regularly carried out in AccuStart II PCR Supermix (QuantaBio, Beverly, MA) under recommended conditions. PCR-SOE cloning reactions were performed with Phusion High Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA) under recommended conditions.

Name	Sequence	Source
dif_siteR	5' CCGAATTGATATGATCTTGATGG 3'	This paper
dif_siteF	5' CCATTGGATTGATAGGAACTGTG 3'	This paper
NEHypR	5' GATTACTGAGCCTCTAAAGCCGTC 3'	(175)
ST36PhageF2	5' AGCAACGAAAACGCCTGT 3'	(175)
ST36PhageR2	5' ACCGTATCACCAATGGACTGT 3'	(175)
PhToxSoAF	5' GGTTGAGTTCGTTTGCTATC 3'	This paper
PhToxSoAR2	5' GTTTCCGCCATTCTTTGGTCCTCACTTGCTCCGC 3'	This paper
PhToxSoBF	5'CGGAGCAAGTGAGGACCAAAGAATGGCGGAAAC 3'	This paper
PhToxSoBR	5' CAATCGTCCTAGCCCGTTTACAAAAGCGACTCATAGA 3'	This paper
PhToxSoCF	5' TCTATGAGTCGCTTTTGTAAACGGGCTAGGACGATTG 3'	This paper
PhToxSoCR2	5' AGGTTCAGGGGTTGGCA 3'	This paper
Tlh-F2	5' AGAACTTCATCTTGATGACACTGC 3'	(174)
Tlh-R	5' GCTACTTTCTAGCATTTTCTCTGC 3'	(157)

QPCR design, validation and quantification of phage content

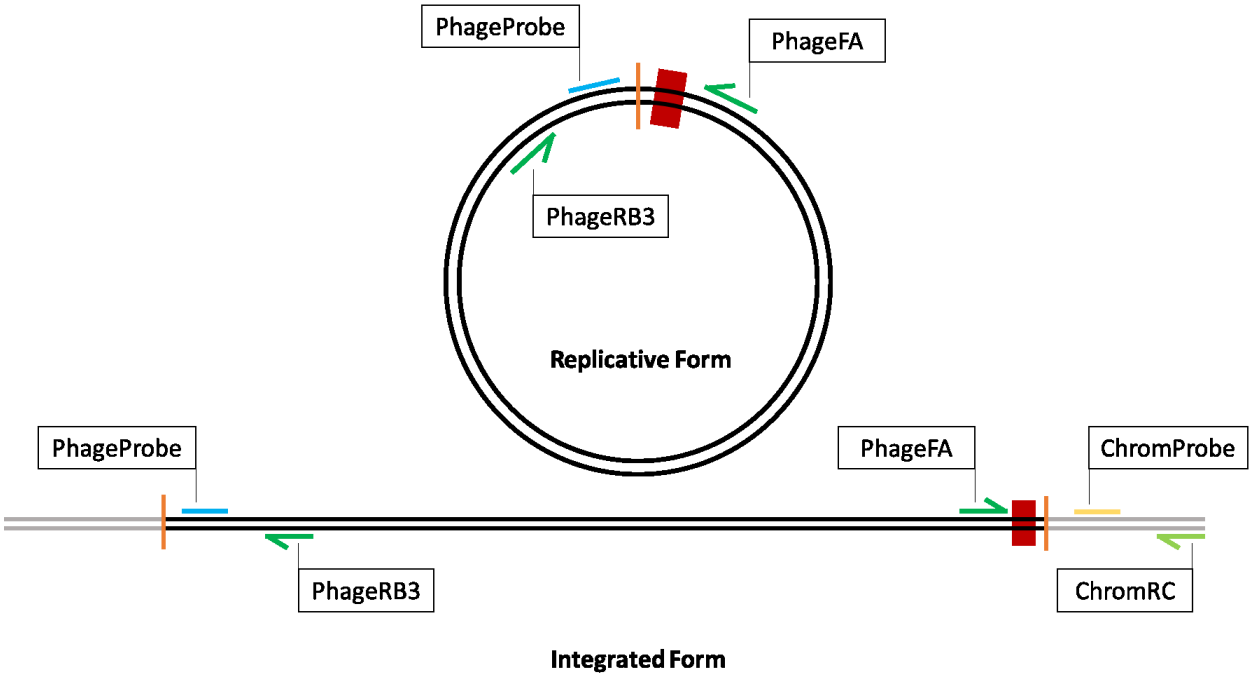

Figure 3.1. Diagram of probe-based qPCR multiplex assessing replicative phage and integrated phage abundance. Primers are in green, probes are in yellow or blue; the red box shows the *dif* site on the phage genome. The multiplex also includes primers and probes specific to *tlh.* The diagram is not to scale.

Probe-based multiplex PCR assays were designed to quantify the number of integrated prophage (IP) and replicative form (RF) of Vipa26. Probes were synthesized by Applied Biosystems, Thermo Fisher Scientific (Waltham, MA, USA). Two primers on the phage that flank the circularization junction, PhageFA and PhageRB3, degrade the PhageProbe during DNA synthesis only from the RF (Fig. 3.1). When PhageFA is combined with the ChromRC primer, that primes synthesis from the chromosome downstream of the *dif* prophage integration site, synthesis from the PhageFA primer degrades the ChromProbe only from an IP template. QtlhF and QtlhR combined with tlhProbe amplify a *V. parahaemolyticus* species-specific marker *tlh* (197).

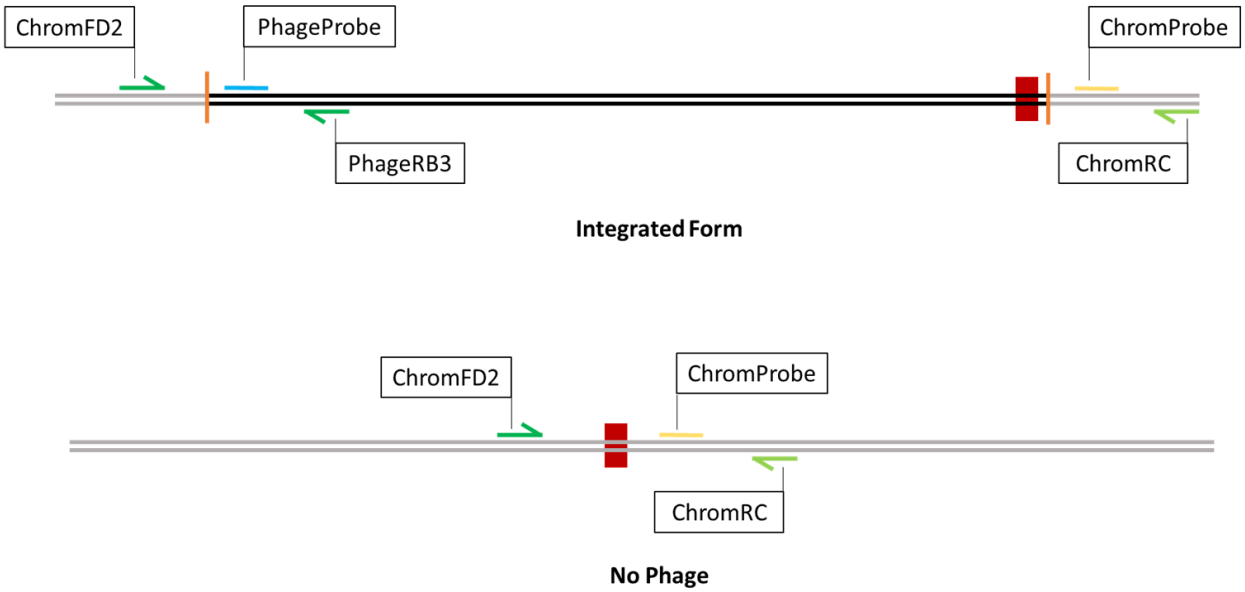

Figure 3.2 Diagram of probe-based qPCR multiplex assessing quantity of integrated phage and no phage. Primers are in green and probes in blue and yellow, the red box indicates the *dif* site on the phage genome and the bacterial genome. Primers and probe specific to *tlh* not pictured are also included in the multiplex. The diagram is not to scale.

A complementary assay was designed to assess abundance of the IP form phage and the *dif* site lacking a prophage, Integrated prophage deficient (IPD), to quantify phage-harboring and phage-deficient cells. The ChromDF2 primer, degrades the ChromProbe during amplification with the ChromRC primer only in the absence of a prophage in the chromosome I *dif* site. Similarly, the ChromDF2 primer only degrades the PhageProbe by amplification with the PhageRB3 primer. As before, QtlhF and QtlhR with tlhProbe are specific to *tlh.*

Table 3.3. Primers and probes for qPCR assessing replicative form, integrated phage, no phage and tlh. Cycling conditions are hot start for 1:00min at 95°C, then two-step amplification of 15sec at 95°C and 45 sec at 59°C for 40 cycles. Reactions were carried out in PerfeCTa FastMix II (QuantaBio, Beverly, MA).

Name	Sequence	Reference
ChromDF2	5’ ACACTTATGAAAGGCTTAATCAAAGAG 3’	This paper
PhageFA	5’ CAAGTCCACAGGAACCACTATATCAGT 3’	This paper
PhageRB3	5’ GTTAAACCACATTCAAATTCACGAA 3’	This paper
PhageProbe	5’ FAM[a]-CAATGAAGTATCACGAAATGA-MGBNFQ[b] 3’	This paper

ChromRC	5 'TCTGAAAGCGATAGGAGAGCAAAG 3'	This paper
ChromProbe	5' TET[c]-AGTGTAACCATACGTCAGAT-MGBNFQ[b] 3'	This paper
tlh forward	5' ACTCAACACAAGAAGAGATCGACAA 3'	(197)
tlh reverse	5' GATGAGCGGTTGATGTCCAA 3'	(197)
tlh probe	5' TxRED[d]-CGCTCGCGTTCACGAAACCGT-BHQ2[e] 3'	(197)

[a] - FAM, 6-carboxyfluorescein; [b] - MGBNFQ, minor groove binding nonfluorescent quencher; [c] - TET, tetrachlorofluorescein; [d] - TxRED, Texas Red; [e] - BHQ2, black hole quencher 2.

Table 3.4. Amplicon primers and probes, target and length.

Primer and Probe	Target	Amplicon Length
PhageFA – PhageRB3 PhageProbe	Replicative Form	198bp
PhageFA – ChromRC ChromProbe	3' Integrated Form	106bp
ChromFD2 – ChromRC ChromProbe	No Phage	104bp
ChromFD2 – PhageRB3 PhageProbe	5' Integrated Form	194bp
QtlhF – QtlhR tlhProbe	*tlh*	208bp

For use in assay optimization and for positive controls and references, each amplicon was cloned into the plasmid pCR2.1-TOPO and transformed into *E. coli* TOP10. Plasmids were purified from cryopreserved stocks following manufacturer specification and EconoSpin All-In-One Silica Membrane Mini Spin Columns (Epoch Life Science, Missouri City, TX) and sequenced by Sanger-sequencing (Genewiz, Cambridge, MA) to ensure amplicon accuracy. Copy number was calculated from the concentration and known sizes of the plasmid stocks, and 10-fold serial dilutions of 10^7 through 1 copies/µL of the plasmid were prepared, including mixed stocks containing all three target plasmids for each assay. Standard curves were constructed and evaluated based on slope and fit to determine the

linear range of detection (Fig 3.1A-C; Fig. 3.2A-C). The multiplex curve was compared to the single-amplicon assay to ensure accuracy. Because detection was variable below 10 copies, we considered this to be the limit of detection.

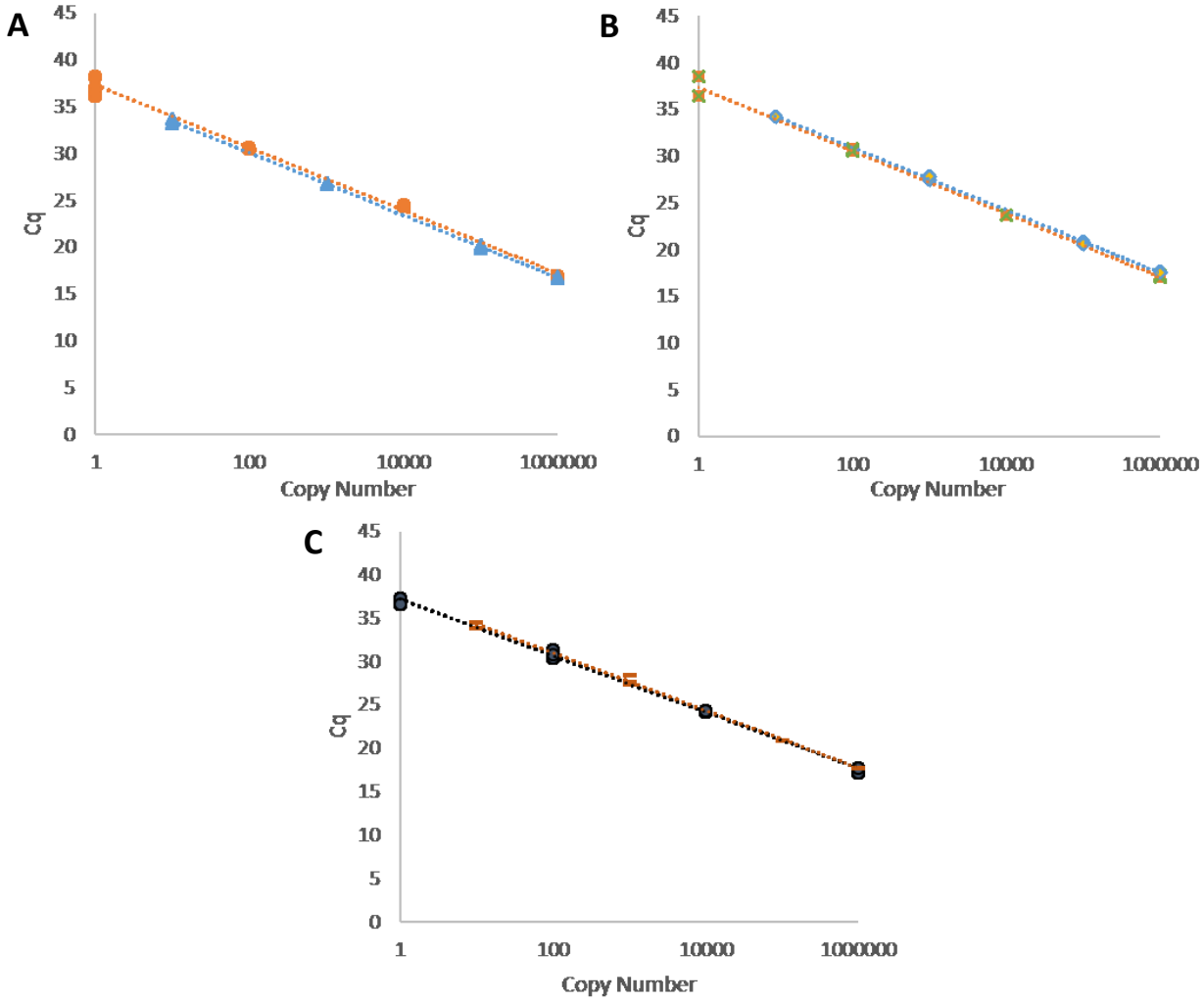

Figure 3.3. **Standard Curves for the integrated and no phage qPCR multiplex. A.** Integrated form amplicon in single- (orange) and multiplex (blue). **B.** No phage amplicon in single- (green) and multiplex (light blue). **C**. *tlh* amplicon in single- (black) and multiplex (red). All curves had efficiencies of 100% +/- 4 and R^2>0.99.

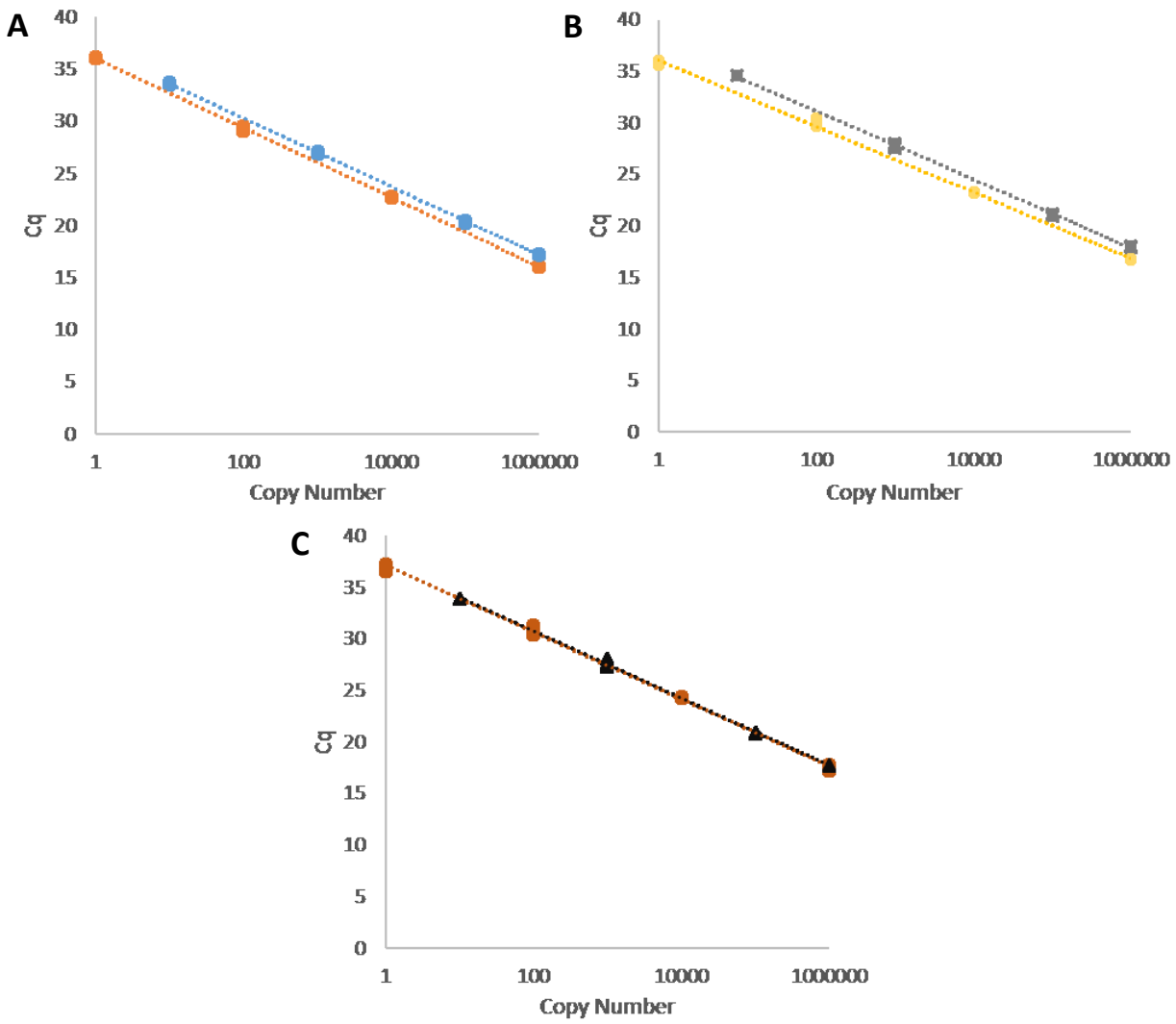

Figure 3.4. Standard curves for replicative form and integrated form multiplex. A. Replicative form amplicon in single- (orange) and multiplex (blue). **B.** Integrated form amplicon in single- (yellow) and multiplex (grey). **C.** *tlh* amplicon in single- (red) and multiplex (black). All curves had efficiencies of 100% +/- 4 and R^2>0.99.

Relative abundance of phage replicative form, UV-C phage induction and resulting cell viability

Relative copy number of the circular replicative form of the phage was assessed by comparing copies of circularized phage to integrated phage using multiplex qPCR as described above. Data was log transformed prior to analysis, and all pairs Tukey-Kramer HSD test to compare means was used in JMP13 to assess statistical significance. To determine ratios of integrated prophage, phage deficient and replicative form, samples from the growth kinetic assay were evaluated using the above multiplex assay.

Data were log transformed and divided by *tlh* log copy number as an approximation number per chromosomes in each sample.

Phage-harboring and phage-deficient strains were grown in LBS at room temperature to late log phase (OD_{600} of ~1.0-1.5) and resuspended in 40% artificial seawater (ASW). 50µL of the culture was exposed to 10mJoules UV-C light in a GS Genelinker UV Chamber (Biorad, Hercules, CA, USA) on an open sterile petri plate, while another was kept as a control. Both were recovered by immediate suspension in LBS and cells incubated at 37°C for 1 hour. Samples were then resuspended in 40% ASW prior to storing at -20°C for later qPCR analysis.

To assess cell viability, cultures were grown in LBS at room temperature to late log phase and exposed to 10mJ of UV-C light as described above, allowed to recover for 1 hour in LBS, then ten-fold serially diluted into Instant Ocean (Instant Ocean, Blacksburg, VA, USA), a formulated sea salt solution commonly used in salt water aquariums to mimic natural conditions. 5µL of the dilution series were spotted onto an LBS plate and relative growth was assessed compared to an untreated control. The cell counts for the lowest concentration with robust colonies were acquired and the fraction surviving after UV treatment compared using an all pairs Tukey-Kramer HSD test in JMP13. All assays were performed in triplicate.

Environmental fitness in microcosms

To assess the impact of natural conditions on ecological fitness in competition, phage harboring and phage deficient strains were transformed with plasmids conferring resistance to Kan and containing either an intact *lacZ* gene which produces blue colonies when grown in the presence of 20µg/mL X-gal colorometric substrate or a *lacZ* gene containing an internal deletion resulting in a frame shift mutation that produces white colonies on X-gal substrate, to allow for recovery and differentiation of strains during competitions. Natural seawater was collected from Adam's Point (Great Bay, NH) within an hour

of low tide. The water was then distributed into flasks in triplicate for each competition performed and blue/white-tagged stationary phase *V. parahaemolyticus* strains (see Table 3.1) were washed with 40% ASW then inoculated in competition at a final concentration of ~10^6 CFU/mL. Samples were immediately taken for confirmation of initial counts and either plated to obtain live cell counts or frozen at -80°C for later qPCR analysis. Competition of isogenic strains differing only in plasmid content were completed to confirm that the forms of the plasmid conferred no difference inf fitness. . Microcosms experiments were conducted in triplicate with a 12 hour light/dark cycle at room temperature, ~22-24°C. Flasks were sampled every 24 hours, with serial dilutions into Instant Ocean, then plating on LBS plates with 50µg/mL kanamycin and 20µg/mL X-Gal to obtain cell counts. In parallel, samples were collected and stored at -80°C for future qPCR analysis. Data were log transformed prior to analysis, percent survival calculated and compared on each day using Tukey-Kramer HSD in JMP13 to assess significance.

Microcosms were treated to remove potential predators/competitors and elucidate the mechanism of fitness differences. Treated microcosms were set up as detailed above; however prior to inoculation by *V. parahaemolyticus* the seawater would be treated as specified. For antibiotic treatment to remove native bacteria, kanamycin was added to the seawater at a final concentration of 50ug/mL and allowed to rest 1 hour, then the *V. parahaemolyticus* was added and water samples collected and frozen at -80C for future qPCR analysis. Added *V. parahaemolyticus* was kanamycin resistant due to plasmid content. Filtered microcosms were sterilized using a 0.22um pore vacuum filtration system to remove native microorganisms, leaving viral content prior to addition of *V. parahaemolyticus* strains, then sampled and analyzed as noted above.

Quantification of biofilm

Biofilm formation was assessed using a crystal violet assay as previously described (198). Briefly, strains were grown in LBS for 3 hours in a polystyrene, non-coated 96-well plate (CellTreat; Pepperell, MA, USA)

and OD_{600} measured on a Tecan Infinite M200 Plate Reader to determine cell density. Without pelleting, the excess cells and media were expelled by inverting the plate leaving only adherent biofilm . The 96-well plate was then rinsed three times in water and dried to ensure no media or cell debris other than biofilm remained. Then 0.1% crystal violet was added and the 96-well plate was incubated for 20 minutes to stain remaining biofilm attached to the wells. The excess crystal violet was then rinsed three times in water and the stained biofilm allowed to completely dry prior to dissolving in 30% acetic acid for 15 minutes and taking the OD_{550} to quantify crystal violet remaining from the dissolved biofilms. Assays were performed by aliquoting a single colony resuspended in media for each strain into a 96-well plate in replicates of 6 to 8. Absorbance of crystal violet (OD_{550}) was standardized against OD_{600} to account for differences in cell density and analyzed using Tukey-Kramers HSD to assess significance in JMP13.

Protist predation and virulence

Cafeteria roenbergensis ATCC 50301 and *Acanthamoeba castellanii* ATCC 30234 were used to assess interaction of *V. parahaemolyticus* with bacterivorous eukaryotic predators. *V. parahaemolyticus* survival in coculture with *C. roenbergensis* was determined as previously described (66). In brief, stationary phase *V. parahaemolyticus* cultures diluted into 40% ASW at a final concentration of ~10^6CFU/mL in a 24-well tissue culture plate (Cell Treat, Massachusetts) and incubated at room temperature. *C. roenbergensis* cultures were grown in media containing 50µg/mL kanamycin bacterized with heat-killed *V. parahaemolyticus* to remove most of the *E. aerogenes* from the xenic culture. Five-day old cultures were washed in 40% ASW and subsequently added to the 24-well plate containing washed *V. parahaemolyticus* to a final concentration of 10^3cells/mL. Cell counts for *C. roenbergensis* were determined using a standard hemocytometer on the red blood cell grid. Live cell counts of *V. parahaemolyticus* were obtained every 24 hours and compared to the initial concentration to obtain percent survival. Assays were performed in triplicate and compared to *V. parahaemolyticus* survival in

40% ASW with no *C. roenbergensis*. Significance of percent survival at the end time point (4 days) was assessed using a Student's T-Test and Tukey Kramers HSD to assess significance of means.

Amoeba seeded plate assays were used as a model of *V. parahaemolyticus* virulence as described (66, 199) and to determine their resistance to predation. A five-day old culture of *A. castellanii* was adjusted to a concentration of ~10^6cells/mL after enumerating using a hemocytometer and resuspended in 40% ASW prior to plating 1.5mL of suspension onto LB plates. The plates were dried for 1-2hrs in a biosafety cabinet, and then incubated at room temperature (~22-24°C) overnight. Stationary phase *V. parahaemolyticus* cultures grown in LBS at 37°C with shaking were then pelleted and resuspended in 40% ASW, adjusted to OD_{600} of 5.0, serially diluted in Instant Ocean and spotted onto the amoeba seeded plates in triplicate. Virulence was scored by the lowest concentration of cells to produce robust colonies in the presence of the amoebas compared to replicates on LB with no amoeba challenge. Two strains from pathogenic linages were used as references: MAVP-C, an ST3 isolate harboring f237 and G149, an ST631 environmental isolate with no inovirus or pathogenicity island.

Results

Vipa26 is an active filamentous phage capable of forming turbid plaques

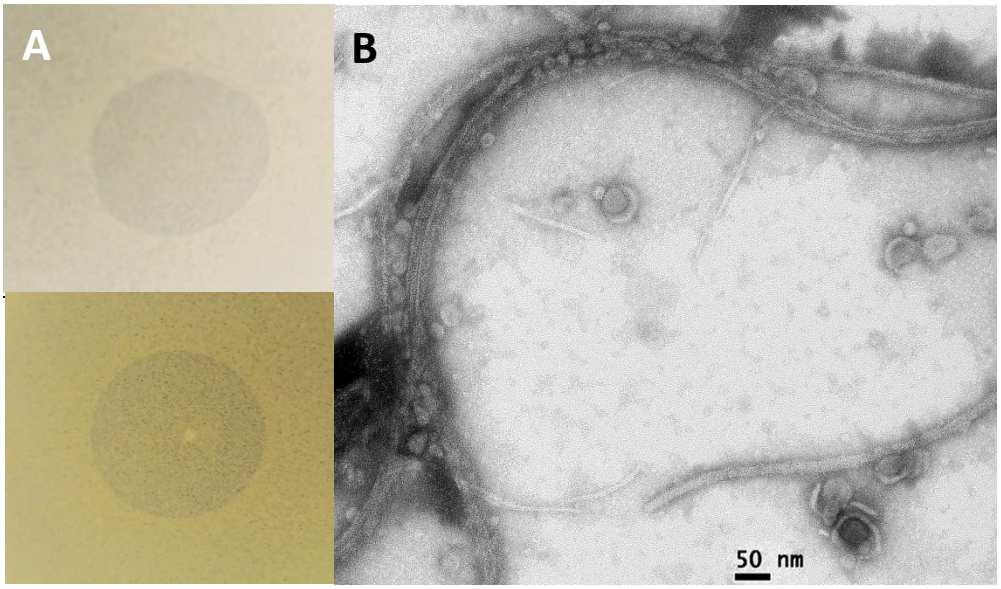

Figure 3.5. Evidence of phage activity. A. Cell-free supernatants of MAVP-26 ("26" with Vipa26) formed a turbid plaque on two ST36 isolates that do not harbor inoviruses. **B**. Visualization Vipa26 progeny by transmission electron microscopy as long bundles with some branching individual phage. Vipa26 is approximately 1000-2000 nm long and 8-12 nm wide.

The Vipa26 phage have been consistently maintained in the ST36 lineage resident in the GOM environmental population as evidenced by its continued association with clinical samples collected between 2013-2017 (See Chapter 2) despite potential fitness costs of persistent infection and expected production of progeny phage. However, if the integrated phage was cryptic and had lost the ability to produce progeny phage, much of the cost of sustained infection would be mitigated, whereas potential beneficial traits could still be conferred. Therefore, our first line of inquiry was to determine whether cells infected with Vipa26 produced progeny. The isolate in which the Vipa26 prophage was first described, MAVP-26, was tested for production of virions. Filtered supernatant confirmed free of bacterial contamination formed the turbid plaques characteristic of inoviruses on phage-deficient ST36 strains, including MAVP-20 and MAVP-V (Fig. 3.5A). PCR analysis of bacteria isolated from the center of the turbid plaque revealed they had acquired Vipa26. We designated these newly infected derivatives as MAVP-20(Vipa26) and MAVP-V(Vipa26). The appearance of turbid plaques on lawns of susceptible cells

following phage exposure and acquisition reveals that at least during the early stages of new infections, Vipa26 exacts a fitness cost in the form of impaired growth.

TEM imaging confirmed the production of virions displaying the expected characteristics of a filamentous phage. The phages purified from supernatants formed large bundles with some individual branching particles (Fig. 3.5B) with estimated dimensions of 1000-2000nm long and 8-12nm wide, as is typical of Inoviruses.

Vipa26 is spontaneously lost during laboratory culture

Curiously, on several occasions, cultures derived from individual colonies of cryopreserved MAVP-26 did not produce plaques on the phage-deficient ST36 strain MAVP-V. This inconsistency suggested the potential that some individuals no longer produced phage progeny, which could have resulted from phage inactivation or phage loss, also known as curing. To evaluate whether phage were cured, individual colonies derived from multiple laboratory stocks were evaluated using primers that amplify all Vipa26 (ST36PhageF2 and ST36PhageR2), including both integrated prophage (IP) and replicative form (RF). PCR produced no phage amplicon in up to 10% of individual colonies suggesting they no longer harbored detectable phage. Subsequent screening of several of individual colonies using primers specific to the IP and intact form of the chromosomal *dif* site, where Vipa26 is integrated, hereafter referred to as integrated prophage deficient (IPD), revealed that Vipa26 had excised from the chromosome in these isolates.

To quantify the dynamic of excision, we subsequently screened populations of bacteria founded by individual colonies and grown by serial passage in broth culture and on agar plates. All 30 individuals derived from each of the 3 parallel liquid cultures over three days (a total of 90 per five time points) harbored phage, both IP and RF (Data not shown); however, as many as 23 out of 90 individuals propagated by serial passage on agar plates had spontaneously lost the prophage, and the proportion of

phage deficient (PD) derivatives lacking even RF of Vipa26 increased significantly ($p<0.05$) with successive passages (Fig. 3.6). Though Vipa26 excision was easily detected from colonies passaged on plates, the relative proportion of the total population undergoing a change in genotype is not directly comparable between liquid cultures and plates because liquid culture populations were randomly sampled and passaged, whereas colonies undergoing a change in genotype (mixed genotype) on agar were selectively passaged to increase resolution of the dynamic of phage loss overtime (Fig. 3.6). Even so, that phage loss was detected in multiple colonies propagated on plates, a change of genotype distribution, demonstrates a realized fitness cost. A phage deficient isolate (MAVP-26PD) was eventually attained by successive passage and archived for further analysis.

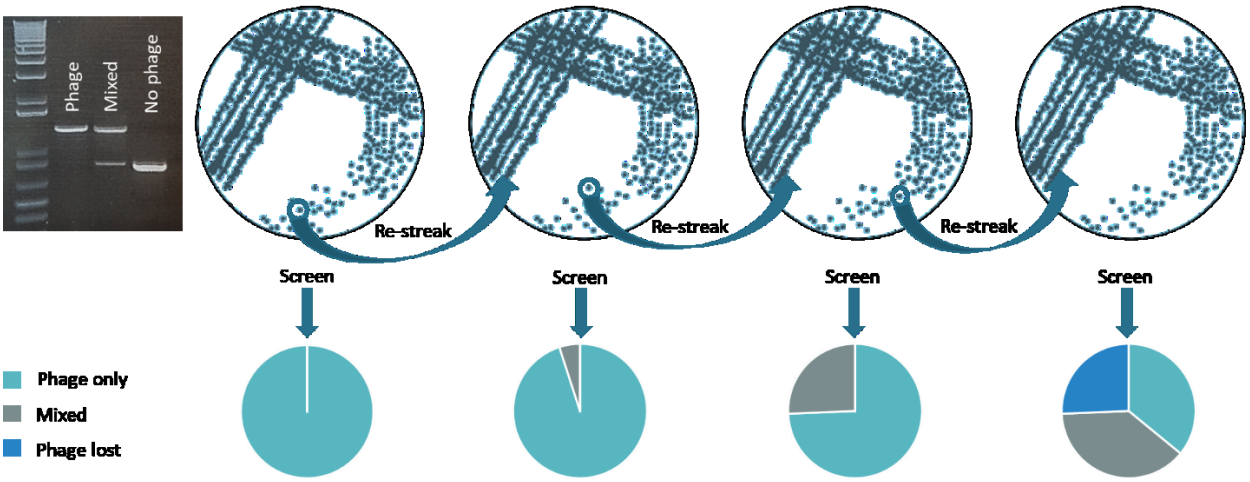

Figure 3.6. Dynamic of phage loss from *dif* chromosomal insertion site in serially passaged colonies. Phage were lost from a GOM clinical isolate following passaging under laboratory conditions as confirmed by PCR screening of progeny colonies.

The dynamic change in phage genotype within parallel clonally derived colony populations was next investigated using the more sensitive method of qPCR. Each day for four days, the frequency of integrated Vipa26 in MAVP-26 was evaluated by destructive sampling of five different aging colonies, and the abundance of Integrated prophage deficient (IPD) *dif* sites and integrated prophage (IP) was

measured over time in each colony population. As MAVP-26 colonies aged, the abundance of IPD *dif* sites increased (Fig. 3.7, orange).

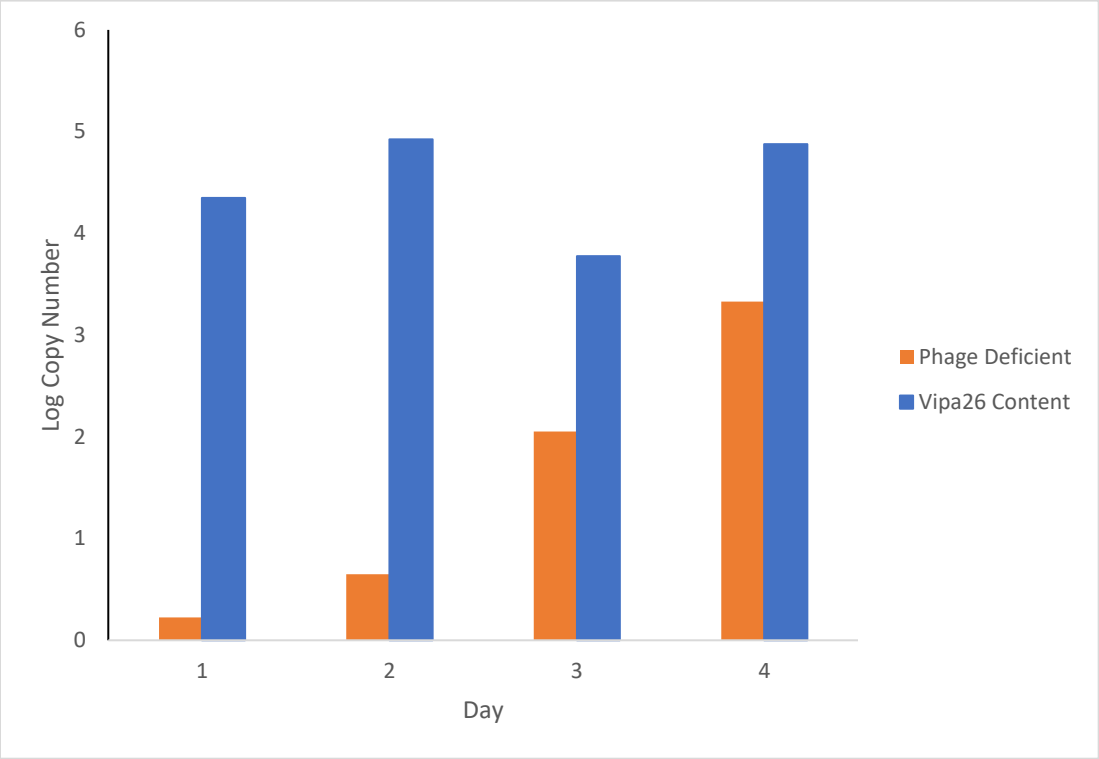

Figure 3.7. Frequency of phage harboring vs phage deficient genotypes over time. Log copy number integrated phage as determined by qPCR is represented by blue, phage deficient is orange. Five whole colonies per day were sampled and the data pooled to illustrate change in frequency over time. Significance over time was calculated using Tukey Kramer's HSD test of means.

Vipa26 prophage inhibited plaque formation by Inoviruses

One potential benefit for harboring Vipa26 is the protection it could provide against new Inovirus infections (162, 200). To evaluate the ability of phage integration to block infection, we investigated plaque-forming ability of Vipa26 and another ST36 Inovirus (Vipa10290) on strains either harboring or lacking Inovirus prophage (Table 3.5). Whereas both Vipa26 and Vipa10290 caused turbid plaques indicative of infection and growth impairment on three different ST36 strains that lack Inoviruses, including MAVP-26PD, they did not cause plaques on MAVP-26 or any other strain known to

harbor an Inovirus (Table 3.5). This suggests at least the potential that phage integration is protective of growth impairment that new infections apparently cause.

Table 3.5. Host range of and cross-protection by phage. The ability of filtrate from MAVP-26PD (no phage), MAVP-26 (harboring Vipa26) and F11-3A (harboring Vipa10290), to cause plaques on soft-agar embedded bacteria. Lack of plaque formation designated by (-) and turbid plaques designated by (+).

	ST36					ST3	STNF	STNF
	No Phage			Vipa26	Vipa10290	f237	Vipa26	Other
Strain:	MAVP-26PD	MAVP-20	MAVP-V	MAVP-26	F11-3A	MAVP-C	G1445	G320
Filtrate:								
MAVP-26PD	-	-	-	-	-	-	-	-
MAVP-26	+	+	+	-	-	-	-	-
F11-3A	+	+	+	-	-	-	-	-

Persistent Vipa26 infections are benign whereas new infections are detrimental

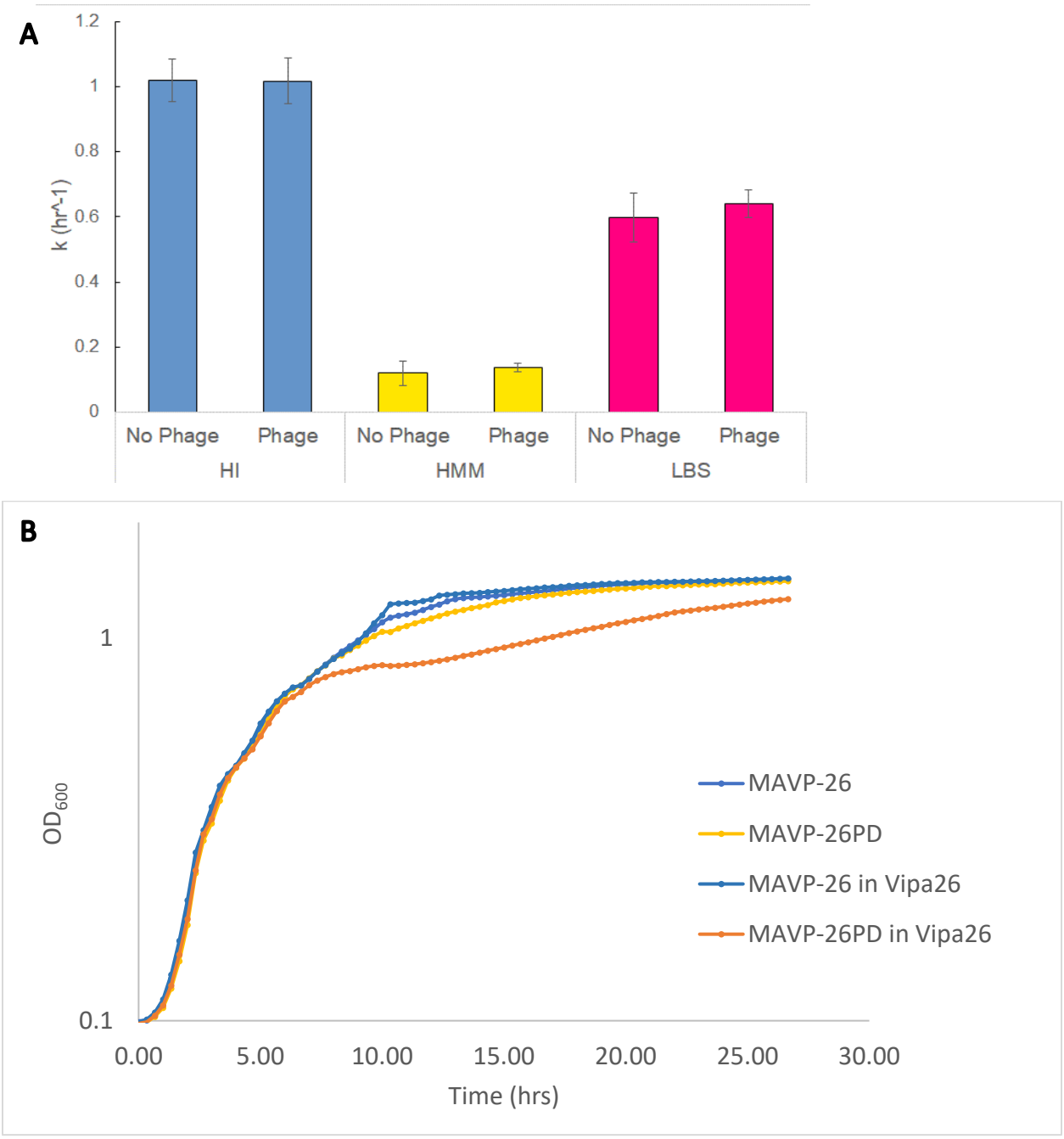

Figure 3.8. Growth of MAVP-26 and susceptible MAVP-26PD with and without Vipa26. A. Specific growth rate in HI (blue), Minimal Media (HMM) (yellow), and Luria-Bertani broth with 3% NaCl (LBS) (pink). MAVP-26 naturally harbors Vipa26 whereas MAVP-26PD is a laboratory derived strain lacking Vipa26. Error bars represent standard deviation. **B.** Growth curve of MAVP-26 and MAVP-26PD (Dark blue and yellow, respectively) grown in culture with phage-free filtered supernatant, MAVP-26 and MAVP-26PD grown in culture conditioned with 5% filtered supernatant containing Vipa26 (blue and orange, respectively).

Vipa26 excision and loss by MAVP-26 occurred frequently in aging colonies (Fig. 3.7) but not in broth cultured cells. Though this potential difference could be explained by any of several or combination of factors including different rates of excision or patterns of phage diffusion and re-infection, it could indicate that persistent phage infections are not detrimental in mass-action cultured cells. To further assess the impact of Vipa26, we examined the growth of isogenic strains either with (MAVP-26) or without (MAVP-26PD) Vipa26 in complex and defined media. The specific growth rates did not differ from each other (Fig. 3.8A) nor did the cultures achieve different total viable cell counts (data not shown). This suggests persistent infections do not detectably impair fitness as determined by growth rate and yield.

Conversely, plaque turbidity suggests newly-initiated Vipa26 infection may slow bacterial cell growth, whereas the unaffected growth rate and absence of plaques by persistently-infected cells indicates sustained infection likely protects cells from the burden of new infections. If this interpretation is correct, Vipa26-growth impairment may be apparent in mass action liquid cultures provided a sufficient proportion of the population was newly infected. To quantify the impact of new infections on growth kinetics, we seeded MAVP-26 and MAVP-26PD at a low starting inoculum into liquid media conditioned with Vipa26 or with phage-free supernatant and monitored the growth kinetics for 18-14 hours. As anticipated, MAVP-26 grew equally well in either phage-conditioned or phage-free broth and with a growth kinetic similar to that of MAVP-26PD in phage-free conditioned media (Fig. 3.7B). In contrast, when grown in media conditioned with Vipa26, MAVP-26PD exhibited a striking growth defect ($p<0.0001$) (Fig. 3.8B; dark blue). Specifically, after ~8 hours of growth in Vipa26-condition broth, MAVP-26PD abruptly and prematurely ceased growth as determined both by optical density (Fig. 3.8B) and cell counts (data not shown) and cultures entered a sustained plateau similar to the stationary phase. In contrast, MAVP-26PD grown in phage-free conditioned media increased in density for several more hours and cell counts increased by an additional order of magnitude (data not shown) before entering

stationary phase. MAVP-26PD did eventually achieve a cell density that was similar to the other treatments after 16-24 hours of growth (Fig. 3.8B). Screening of 25 individual colonies derived from MAVP-26PD grown in Vipa26-condition media for 18 hours using primers that detect all forms of Vipa26 revealed all survivors contained Vipa26 suggesting high efficiency of infection and potentially strong selection for phage retention under the assay conditions.

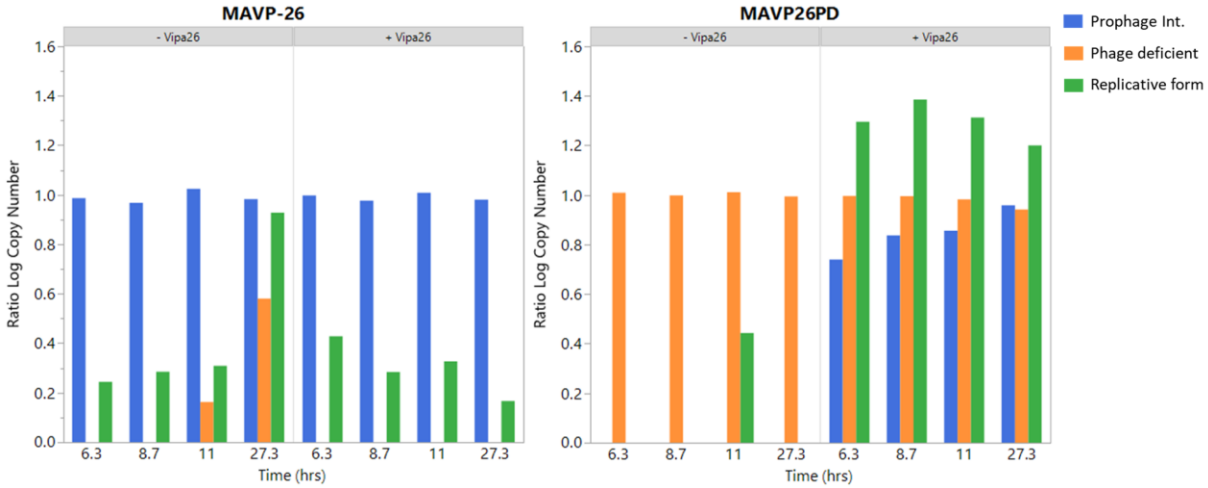

Figure 3.9. Effect of Vipa26 integration on replicative form abundance in new and stable infections. Ratio of PD (orange), RF (green) and PI (blue) forms standardized to a *tlh* control as measured with qPCR in **A.** MAVP-26 and **B.** MAVP-26PD in media either conditioned with phage-less cell-free supernatants or with Vipa26.

Because screening individual colonies to determine distribution of phage genotype limits resolution of infection dynamics, we employed qPCR analysis of phage genotype on populations of cultured cells. Specifically, we quantified absolute and relative number of replicative (RF) and integrated prophage (IP) form abundance and chromosomes without integrated prophage (IPD) in parallel cultured populations of MAVP-26 and MAVP-26PD grown in media conditioned with phage-free supernatants or with Vipa26 and used the species genetic marker *tlh* to standardize for chromosomal copy number. The amount of IP Vipa26 remained high and constant in persistently infected MAVP-26 populations but interestingly, a subpopulation of cells did undergo prophage excision which also coincided with a marked increase in RF suggesting excision enhances phage replication (Fig. 3.9A). Vipa26 rapidly infected

the MAVP-26PD population and the number of chromosomes with IP form increased over time while the number of IPD chromosomes concurrently declined. Newly infected MAVP-26PD produced significantly ($p<0.0001$) more RF, based on total copy number than persistently infected MAVP-26 at every time point. Though, by 27 hrs post Vipa26 exposure, RF decreased significantly in the newly infected population concurrently with the increase integration as gleaned from IP form. Considering end-point PCR for Vipa26 does not discriminate between IP and RF, these data combined with the prior isolate analysis that 100% of survivors tested harbored Vipa26 post-exposure under the same culture conditions suggests that not every new phage infection leads to immediate chromosomal integration.

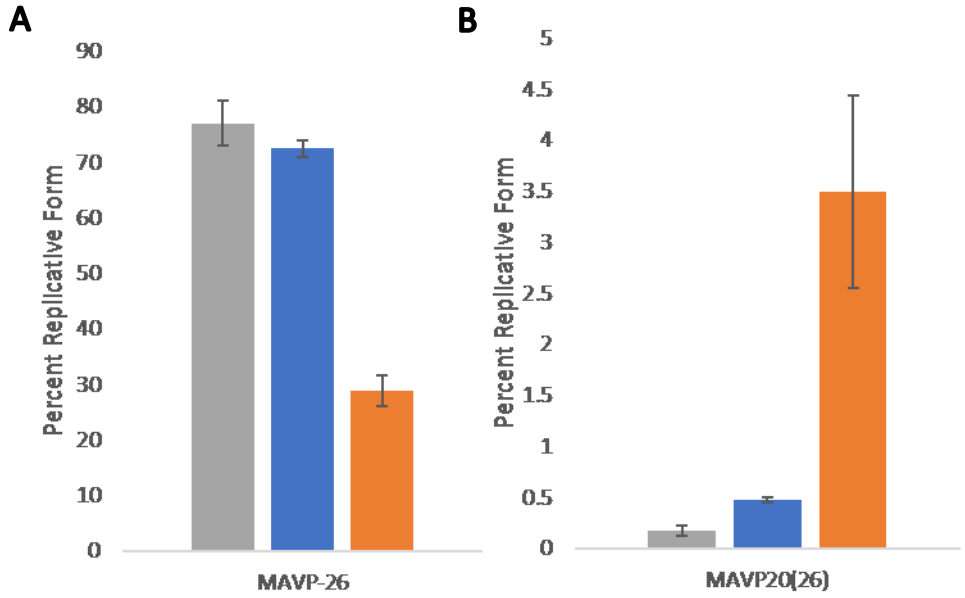

Figure 3.10. Relative copy number of Vipa26 replicative form in non-isogenic strains. A. Relative copy number of the Vipa26 replicative form compared to the copy number of prophage in MAVP-26 **(A)** and MAVP-20(26) (**B**). Blue, orange and grey bars each represent a separate run of the assay. Error bars represent standard deviation. Note the different scales of the y-axis.

New infections are costly and characterized by high abundance of replicative form and growth retardation; however, persistent infections do not cause growth defects that we could discern. One notable difference between the stable and newly infected strains is the higher abundance of replicative form copy number in newly initiated infections. To elucidate the range of replicative form abundance in stable infections, we quantified copy number of Vipa26 in two non-isogenic ST36 strains.. This revealed

that replicative form is maintained at a low copy number (less than one copy per cell) in stationary phase cultures of MAVP-26 and was nearly undetectable in MAVP-20(26), but this varied greatly from experiment to experiment (Fig. 3.10A and B). MAVP-26 maintained between 28 and 76% RF relative to IP form , and MAVP-20(26) maintained between 0.1% and 3.5% RF relative to IP. The high variability between not only the strains but also each run complicated the attempt to use replicative form as a measure of phage activity to investigate environmental triggers that may cause an increase in Vipa26 production. Beyond the first explosion of replicative form production at the initiation of infection, it appears that the phage becomes largely quiescent (Fig. 3.10).

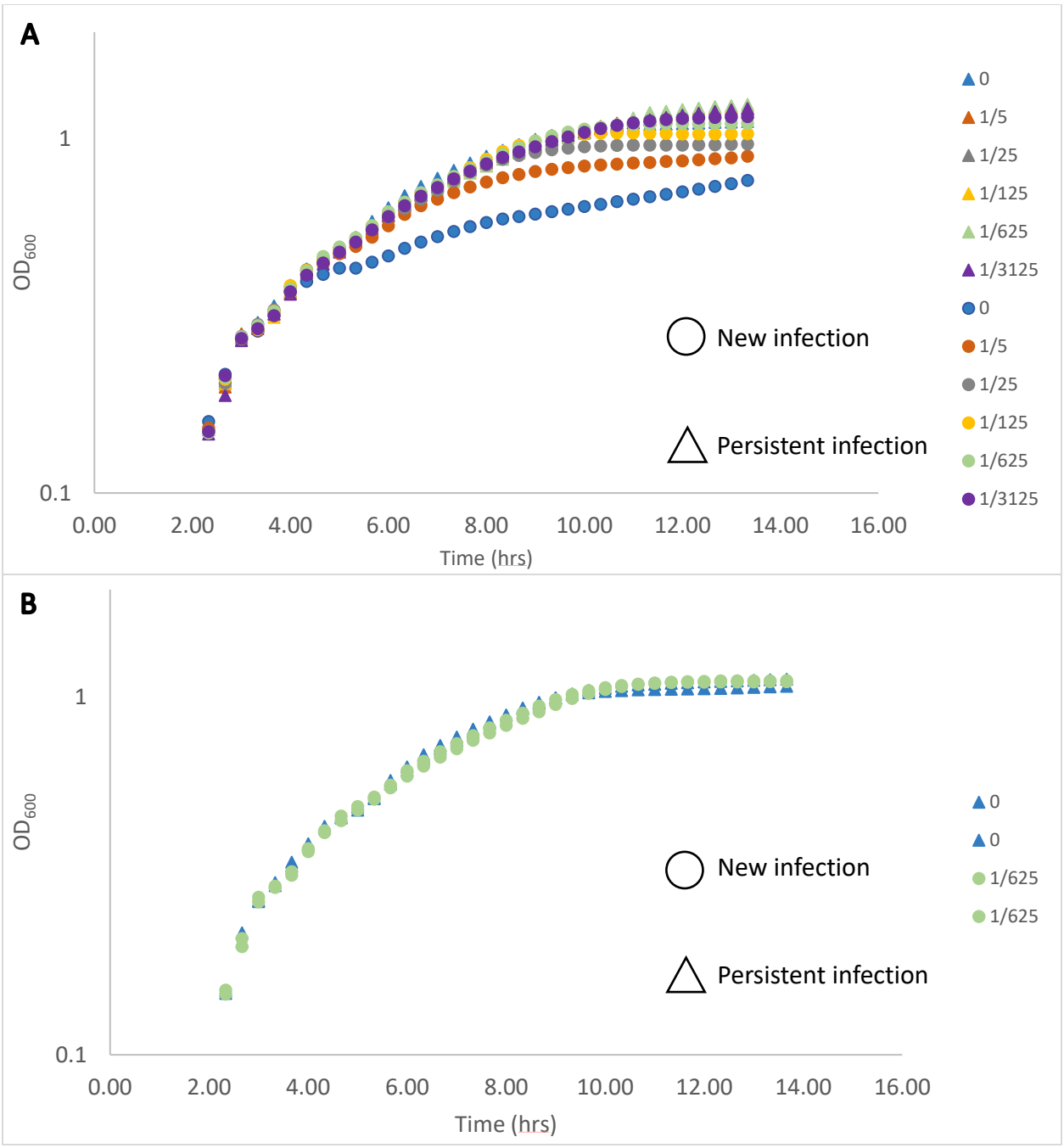

Figure 3.11. Quantification of viable phage progeny. A. MAVP-26PD and MAVP-26 recovered from growth in Vipa26 suspension was resuspended in fresh media and grown to over-night at room temperature to late log phase. Susceptible MAVP-26PD growth in 5-fold serial dilutions of the resulting supernatant was measured. The OD_{600} of MAVP-26PD in dilutions of phage produced by MAVP-26 are triangles, circles represent OD_{600} in MAVP-26PD produced phage dilutions. Color represents dilution factor as specified in the legends; blue for the original media containing 5% phage filtrate (0 dilution factor); subsequent dilutions are in order: orange (1/5); grey (1/25); yellow (1/125); light green (1/625); and purple (1/3125). **B.** Overlapping growth curves: growth in the second lowest concentration (1/625; light green) of Vipa26 suspension from newly infected MAVP-26PD (circle) is equal to that in the highest concentration (0, blue) of Vipa26 produced by MAVP-26 (triangle).

These data indicate newly infected cells harbor a higher relative abundance of RF suggesting that these cells could produce more progeny virions, which could help explain the greater fitness cost, as suggested by the plateau in the growth of newly infected cells compared to persistently infected cells. To ascertain whether newly infected cells produce more virions we utilized a semi-quantitative conditioned broth assay. Filtrate from newly-infected cells and persistently infected cells were serially diluted, the resulting broth was inoculated with MAVP-26PD and the population growth kinetics subsequently determined and compared to identify phage dilutions with a similar ability to impair growth, and by extrapolation that had similar numbers of phage (Fig. 3.11A). When directly compared, filtrate from newly-infected cells suppressed the growth of MAVP-26PD better than filtrate from persistently-infected cells consistent with a greater production of phage (Fig. 3.11A). Identification of the dilution of filtrate from newly infected cells that conferred an identical modest kinetic of growth impairment conferred by undiluted Vipa26 from persistently infected cells indicates that newly infected cells produce ~625-times more infective progeny (Fig. 3.11B).

Evidence of ecological benefits conferred by persistent phage infection

The early growth plateau of newly infected strains in broth seeded with Vipa26 and the extremely high titer of phage produced by these strains, suggest a fitness cost that may translate to a competitive deficit when phage deficient cells are in direct competition with phage harboring cells. Conversely, strains harboring a persistent infection do not experience this detriment, a potential benefit in mixed communities. To directly measure competitive fitness, strains were cocultured either in broth conditioned with Vipa26 or in phage free broth. Ratios of viable cells were determined at the first and last time points and presence of Vipa26 in survivors was analyzed with end point PCR using primers specific to the Vipa26 core region (see methods), and qPCR assessing the abundance of integrated prophage compared to integrated prophage deficient *dif* site.

MAVP-26PD labeled with a plasmid conferring *lacZ* (blue) was co-cultured with MAVP-26 labeled with a frame-shift mutant copy of *lacZ* (white) at a ratio of 3:1 (75%) in phage-free broth, and this proportion of PD community members decreased to ~1:32 (3%) following 27 hours of co-culture, although ~29% of the population were newly infected and derived from the original MAVP-26PD population (Fig. 3.12). In broth seeded with Vipa26, the shift in population was similar, with Vipa26-free MAVP-26PD isolates declining from 50% of the population to approximately 1% by 27 hrs (Fig. 3.12). Potentially newly Vipa26-infected MAVP-26PD accounted for 17% of the population. Thus, even though seeded phage greatly decreased the competitiveness of a phage deficient variant (Fig. 3.12C), the presence of the stably infected MAVP-26 was sufficient to ensure the vast majority of the population contained Vipa26.

To provide greater resolution into the dynamics of the competition without the limits of end point PCR, a qPCR assay to assess levels of Vipa26 integration revealed that the ratio of prophage to integrated prophage deficient *dif* sites (IP to IPD) increased significantly ($p<0.05$) during the 27 hr competition (PD broth – 0.33 to 0.63 respectively; Vipa26 broth – 0.43 to 0.87; Fig. 3.12A and B). These trends were similar but less dramatic than the near total disappearance of prophage deficient (PD) individuals from end point PCR screening.

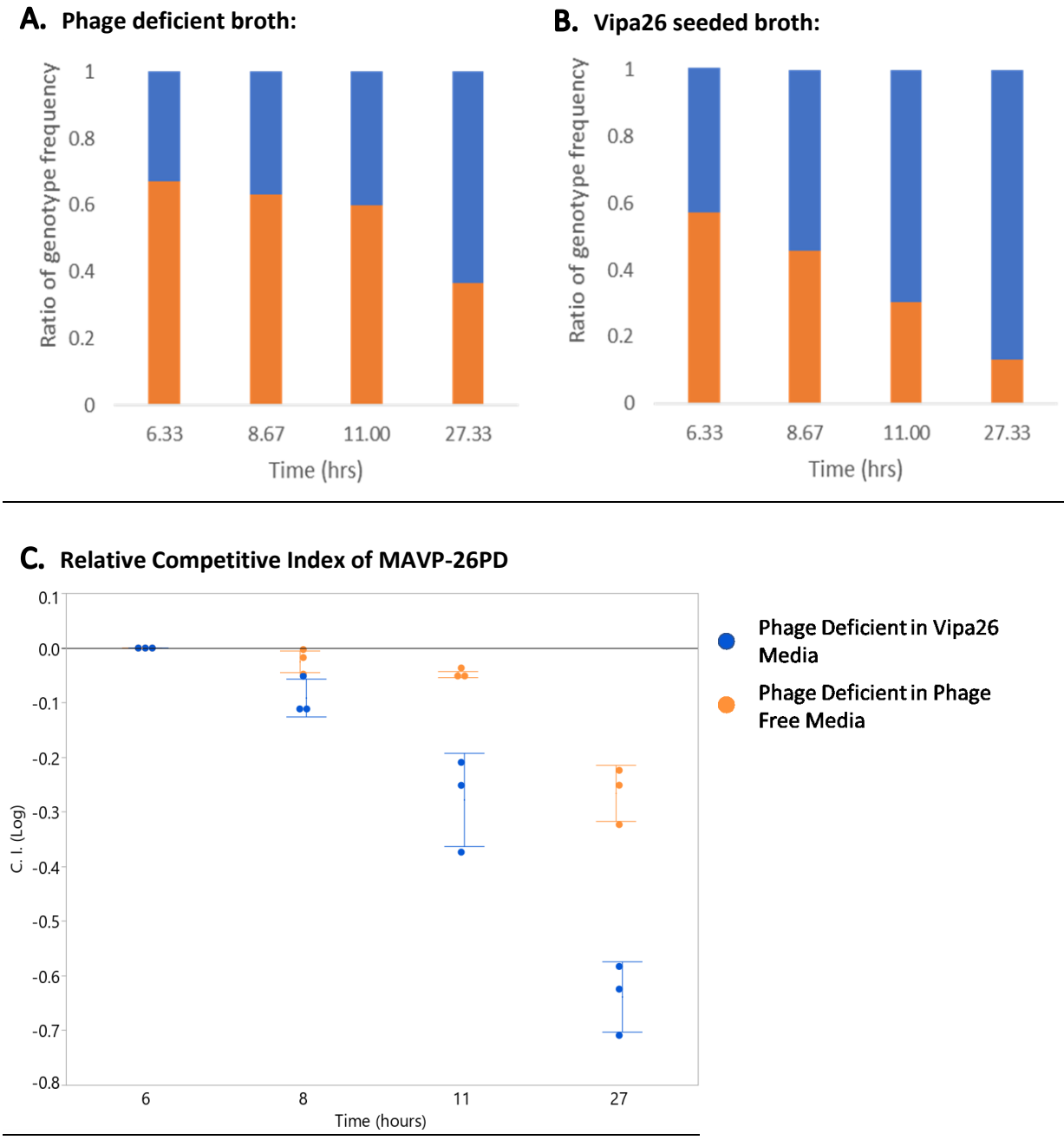

Figure 3.12. Shift of genotypic frequency of integrated prophage in direct competition. A. Ratio of IP (blue) compared to IPD *dif* site (orange) in a competition in phage deficient broth as determined with qPCR. **B.** Ratio of integrated prophage in competition in broth conditioned with Vipa26 determined by qPCR. Colors are the same as part A. **C.** Relative competitive index of MAVP-26PD in coculture with MAVP-26 in Vipa26 conditioned or phage-free broth.

Table 3.6. Shift in frequency of Vipa26 in viable cells in competition. Presence or absence of Vipa26 in viable cells by colony end point PCR specific to core phage ORFs, in direct competition in phage-free broth and Vipa26 seeded broth; new lysogens are noted in the parentheses. All forms (RF and prophage) of Vipa26 are indistinguishable with these primers.

Phage deficient broth end point PCR:

Time(hrs):	Vipa26-Harboring	Vipa26-Deficient	Total Colonies
6.33	9	36	45
8.67	16	29	45
11	17	28	45
27.33	105 (32)	2	107

Vipa26 seeded broth end point PCR:

Time(hrs):	Vipa26-Harboring	Vipa-26-Deficient	Total Colonies
6.33	23	22	45
8.67	35	10	45
11	28	14	42
27.33	89 (15)	1	90

Vipa26 selected against susceptible strains in competition, significantly decreasing their competitive fitness, a clear benefit for strains with integrated prophage. Conversely, the inovirus excises at a certain rate under laboratory conditions, alluding to a realized fitness cost. Vipa26 may have broader influences on the host cell not yet elucidated. Since Vipa26 is actively produced even in persistent infection, it is possible that the phage is impacting the host cell in other ways after infection is established; a process known as lysogenic conversion for temperate phage (132). One potential fitness impact could be in response to UV-C light exposure, as UV light is frequently correlated with phage induction. To investigate this, we compared viability of phage harboring and phage deficient isolates after exposure to 10mJ UV-C light and subsequent recovery. Vipa26 harboring isolates recovered better after exposure, compared to isogenic phage deficient isolates (Fig. 3.13). This difference was especially apparent for MAVP-26 and its phage deficient counterpart MAVP-26PD. The recently infected MAVP20(26) was more viable compared to the natively phage deficient MAVP-20; however, phage infection did not rescue it to the same level as MAVP-26. MAVP-20(26) had almost identical viability to the phage deficient isolate MAVP-26PD. Despite this difference, UV light did not appear to induce the phage (data not shown). Further analysis of the phage genome using BLAST and BPROM (201) identified no LexA binding sites further supporting that Vipa26 is not induced by UV damage.

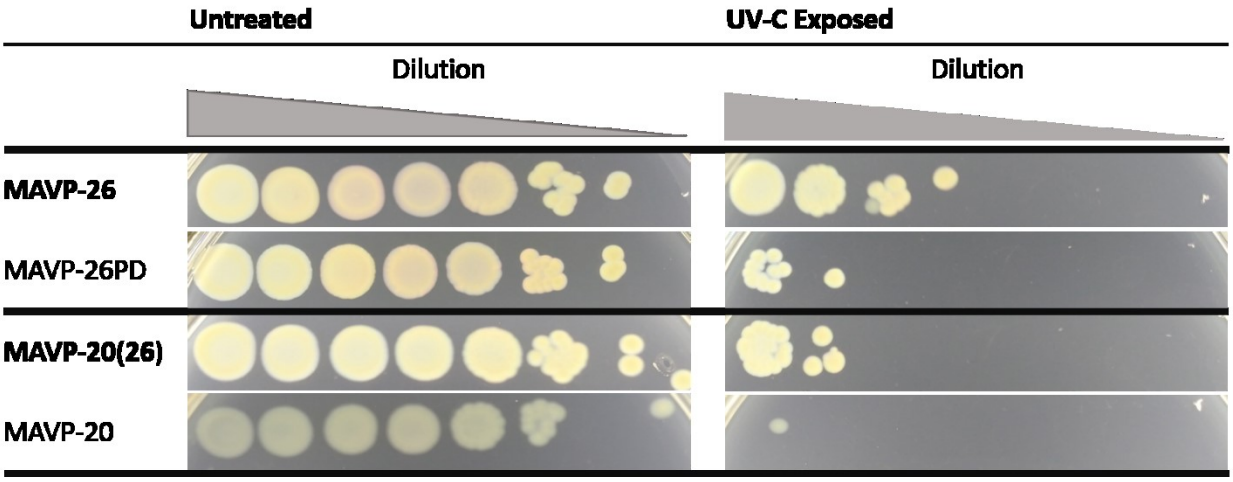

Figure 3.13. Viability after exposure to UV-C. Viability of MAVP-26 (phage; bold) compared to MAVP-26PD (no phage) and MAVP-20(26) (phage; bold) compared to MAVP-20 (no phage) after exposure to 10mJ UV-C light. First spot is undiluted and subsequent spots are 10-fold dilutions. Relative decrease in viability of phage deficient strains compared to phage harboring strains is significant based on cell counts compared using a Student's T-test ($p<0.05$).

Another potential mechanism by which the Vipa26 may impact ecological fitness of their hosts is by enhancing biofilm formation. Previous studies have linked filamentous phages to changes in this (125, 139) through increasing the cell death and eDNA release critical for biofilm formation and directly stabilizing the biofilm through interactions with the negatively charged filamentous phage particles. Crystal violet staining of biofilm of both phage harboring and phage deficient isolates elucidated the potential role of Vipa26 on biofilm. Phage content appeared to decrease biofilm in MAVP-26 (phage) compared to MAVP-26PD (no phage), but the change in absorbance was not significant (Fig. 3.14). MAVP-20(26) produced much more biofilm ($p<0.05$) than MAVP-20, likely due to the opaque and translucent respective phenotypes. A plaque assay produced a secondary MAVP-20(26) isolate (MAVP-20(26)-2) that was also translucent and had very similar biofilm formation to MAVP-20.

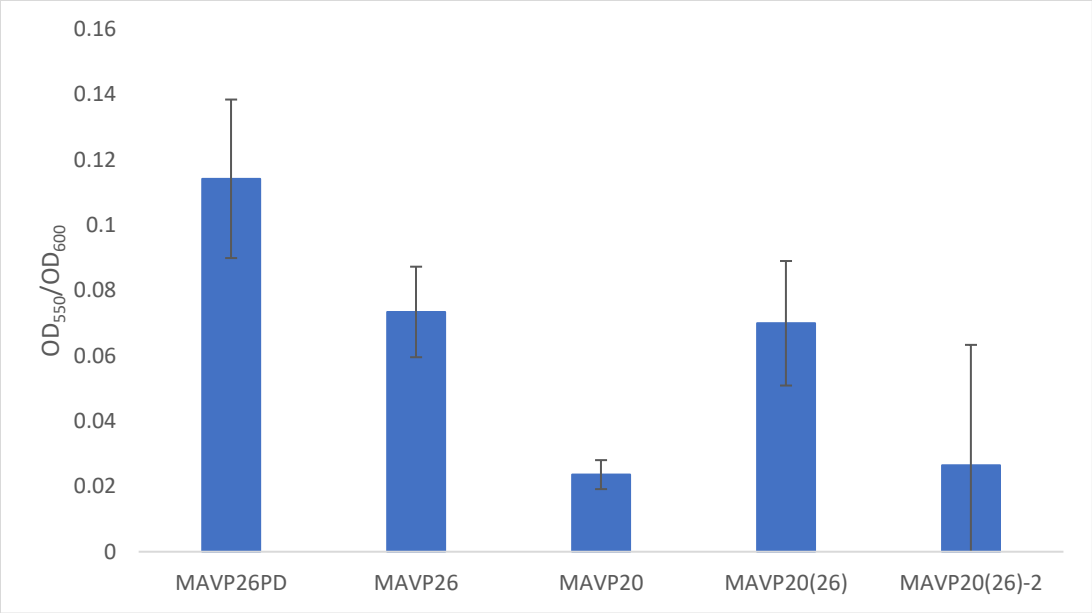

Figure 3.14. Biofilm formation of phage harboring and phage deficient isolates. Phage deficient isolates MAVP-20 and MAVP-26PD compared to Vipa26 harboring MAVP-20(26), MAVP-20(26)-2 and MAVP-26. MAVP-20 and MAVP-20(26)-2 are translucent phase variants, the others are opaque.

Survival in microcosms as a model for ecological fitness

Although biofilm formation did not differ, and phase variation confounded the potential analysis of UV survival, the protection against the burden of new infection points to potential fitness benefits of harboring the Vipa26 in the environment. However, testing of specific traits under laboratory conditions fails to capture the actual conditions found in the natural environment where selection would occur. To address this shortcoming, natural seawater microcosms were employed as a model to test overall environmental fitness. Phage harboring strain (MAVP-20(26)) that had undergone phase-variation to an opaque phenotype, survived significantly ($p<0.01$) better than the parental strain MAVP-20 which exhibits a translucent colony type on days two and three of the microcosm as determined by cell counts (Fig. 3.15A) and qPCR of prophage compared to integrated prophage deficient *dif* sites (Fig. 3.15B). qPCR of MAVP-20(26) reciprocal self-competition indicated the chromosomally integrated prophage did not excise over the course of the microcosm, validating the use of prophage content to differentiate

phage harboring and phage deficient strains. Future microcosms were analyzed solely with qPCR (Fig. 3.15B). Copy number determined by qPCR has a linear relationship to CFU/mL (Fig. S3.1).

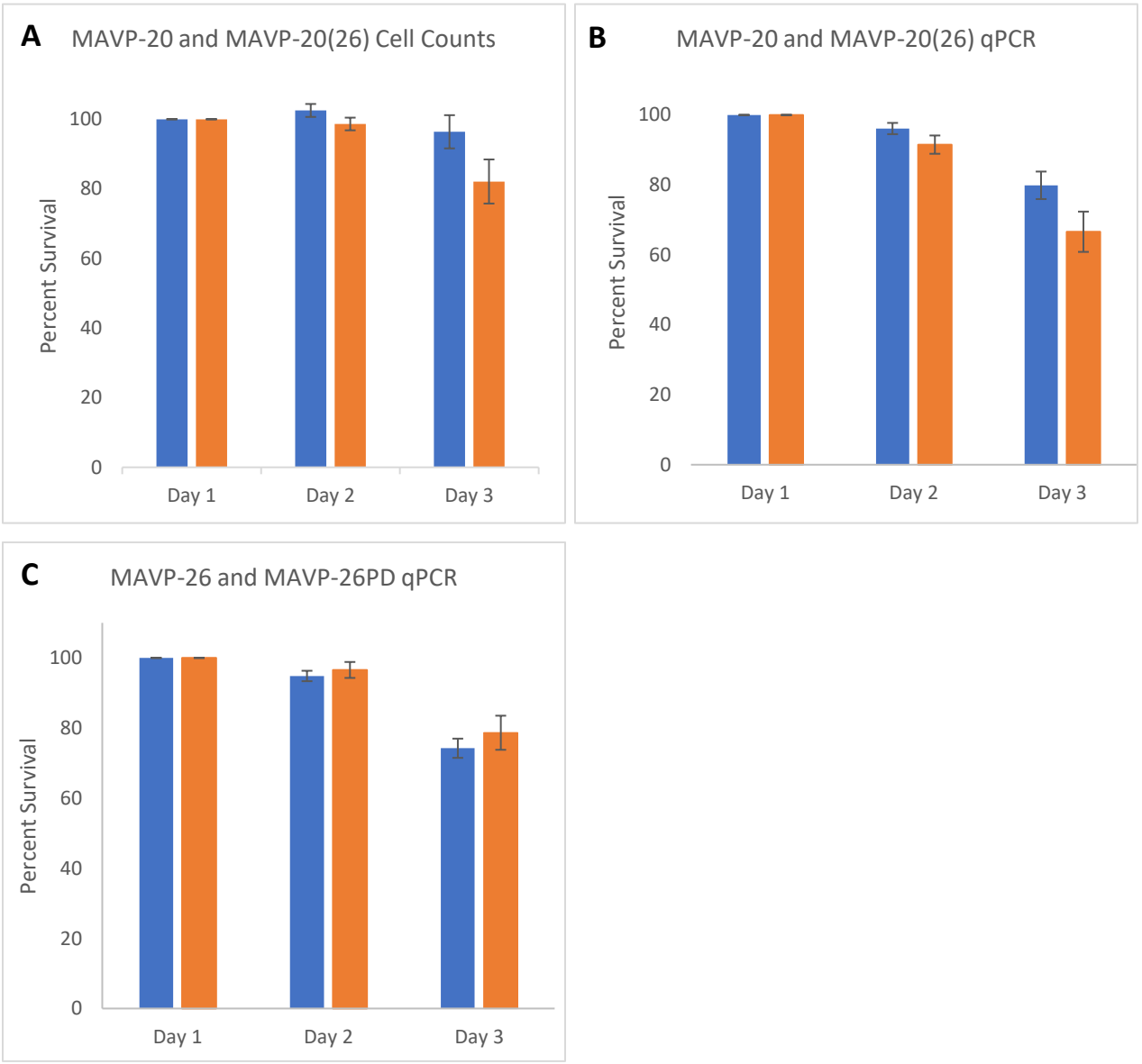

Figure 3.15. Survival of MAVP-20 and MAVP-20(26) in natural seawater microcosms in competition. A. Differentially tagged strains were competed in seawater in reciprocal coculture for one week. Percent survival determined by viable cell counts of Vipa26 harboring strain MAVP-20(26) is blue and the phage deficient strain MAVP-20 is orange. **B.** Percent survival of MAVP-20(26) (phage; blue) and MAVP-20 (no phage, orange) as measured with qPCR. **C.** MAVP-26 (phage; blue) competed with MAVP-26PD (no phage; orange), measured with qPCR, both strains are opaque. All error bars represent standard deviation.

MAVP-26 (phage harboring) and MAVP-26PD (phage deficient) competed in untreated microcosms did not differ in percent survival on any day of the microcosm (Fig 3.15C). Percent survival across different microcosms of MAVP-20(26), MAVP-26, and MAVP-26PD did not differ significantly, with MAVP-20 alone exhibiting decreased survival ($p<0.001$) (Fig. 3.15). MAVP-20 has a translucent phenotype due to the phase variation observed in *V. parahaemolyticus* (202), whereas the MAVP-20(26) strain has an opaque phenotype (Fig 3.14). Microcosms were incubated for a week; however, levels of both Vipa26-harboring and Vipa26-deficient strains dropped below the level of detection by day four for both plate counts and qPCR. *V. parahaemolyticus* and Vipa26 content were undetected in natural seawater with no added bacteria at any time point of the microcosm; however, the natural community was not investigated beyond this.

One of the major factors influencing ecological fitness is competition and predation amongst community members. By treating the seawater collected to remove specific parts of the community, we narrowed down the potential source of the observed difference in survival in the untreated microcosms. Filter sterilizing removes all native prokaryotic and eukaryotic species which might interact with *V. parahaemolyticus* leaving only the viral load, whereas kanamycin treatment kills most of the bacterial species present while leaving the eukaryotes. Percent survival of MAVP-20 and MAVP-20(26) competed in either of the treated microcosms did not differ significantly (Fig. 3.16A and B). Both MAVP-20 and MAVP-20(26) survived longer in filtered seawater in comparison to the untreated seawater, with a measurable population still present on day four (Fig. 3.16A). Kanamycin treatment lowered overall percent survival on days two and three compared to the untreated set, with greater variability in survival on day three.

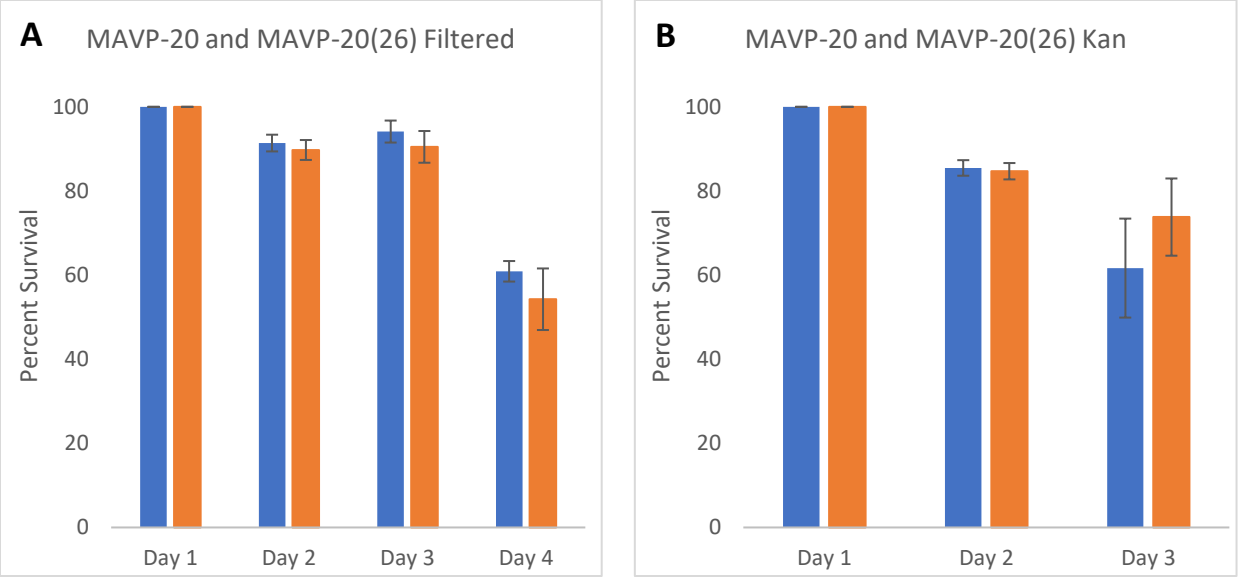

Figure 3.16. Percent survival in treated microcosms. A. Survival of phage harboring (MAVP-20(26)) and phage deficient (MAVP-20) strains in competition in a filter sterilized microcosm. Colors correspond to Fig 3.15: MAVP-20(26) (blue) and MAVP-20 (orange). **B.** Percent survival in a seawater microcosm treated with 50µg/mL kanamycin (same colors as A). Error bars represent standard deviation.

Phage does not impact protist predation or preliminary models of virulence

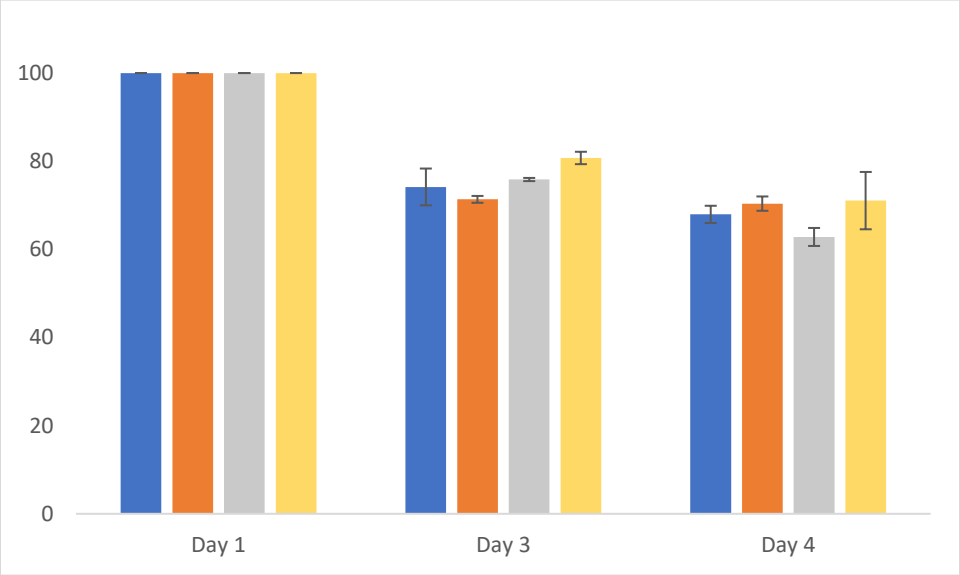

Figure 3.17. Percent survival after incubation with predatory protist challenge. Survival of MAVP-26 (Vipa26; blue) and MAVP-26PD (no phage; orange) after incubation in 40% ASW with the bacterivore *C. roenbergensis.* MAVP-C (ST3; f237; yellow) and G149 (environmental; no phage; grey) were included as controls to recapitulate previously reported trends (66). Error bars represent standard deviation.

Harboring a phage may confer a fitness advantage illustrated by greater survival in untreated microcosms and mitigated by the kanamycin treatment to remove potential bacterial competition and predation, although attributing this advantage to Vipa26 is complicated by the translucent phenotype of MAVP-20. Although this differential survival suggests that bacterial influences are the source of the advantage, pressure by eukaryotic predators is another component of environmental challenges facing *V. parahaemolyticus*. To investigate eukaryotic predation, resistance to *C. roenbergensis* grazing was quantified. Incubation in coculture with *C. roenbergensis* depressed the survival of *V. parahaemolyticus* over several days with no difference in survival of Vipa26 harboring MAVP-20(26) and phage deficient MAVP-20 (Fig. 3.17). Growth in media amended with 0.04% bile salts to stimulate virulence factors (203) and in competition also did not affect ability to survive in the presence of the protist (data not shown). Contrary to previous reports, the population size of the ST3 isolate MAVP-C did not increase in culture with *C. roenbergensis,* demonstrating killing of the protist, and the environmental isolate G149 performed similarly to the clinical MAVP-26 (Fig. 3.17)(66).

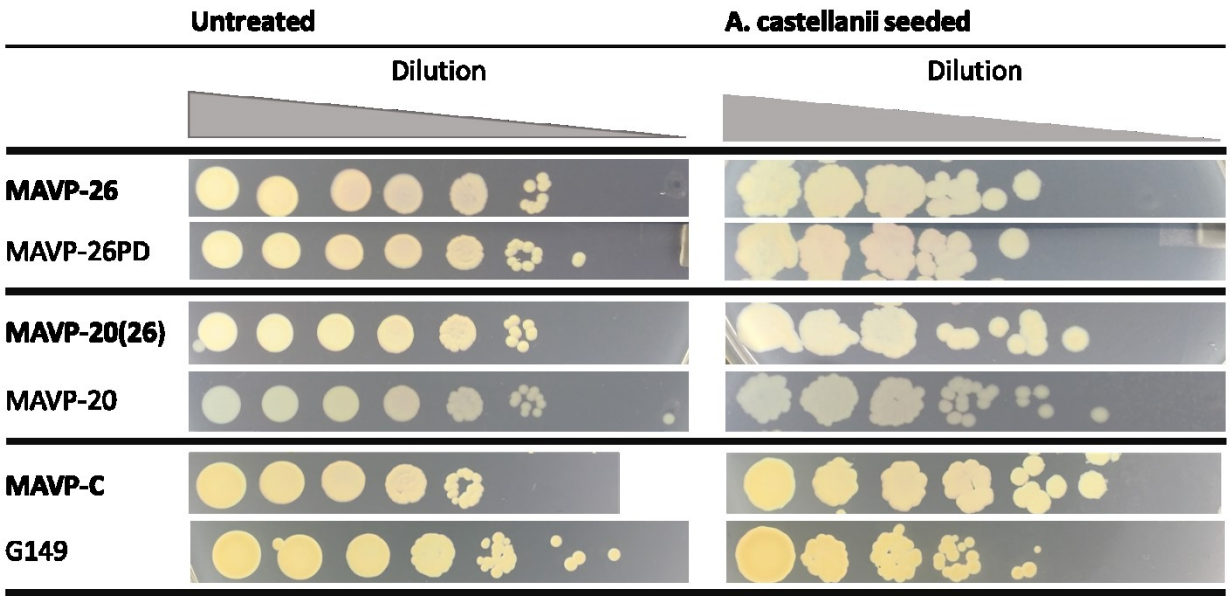

Figure 3.18. Virulotyping of isolates. MAVP-26 and MAVP-20 with Vipa26 (bold) and without Vipa26 grown on plates seeded with *A. castellanii* to determine relative virulence against eukaryotes. MAVP-C (ST3 containing f237) and G149 (environmental isolate with no f237-like phage) are phage harboring and phage deficient references respectively.

V. parahaemolyticus is an environmental pathogen, therefore the first and arguably most important step in the infective pathway is ecological fitness. If a strain is outcompeted and unsuccessful it is unlikely to accumulate in an oyster to the level necessary to cause illness. Measuring ecological fitness is critical to understanding the impact of the phage on the success of ST36; however direct effects of virulence could also explain the high clinical prevalence of this population, perhaps due to selection for Vipa26 in the gut. To investigate this, virulotyping of phage harboring and phage deficient isolates was completed using amoeba seeded plate assays with *A. castellanii* as previously described (66, 199). *V. parahaemolyticus* viability was impaired compared to the control; however, the minimum surviving concentration between any of the isolates tested did not change (Fig. 3.18). An environmental isolate from a pathogenic lineage, G149, which lacks a pathogenicity island and is phage deficient, survived on amoebas at the same concentration as the clinical ST36 isolates, calling into question the validity of testing virulence using this assay.

Discussion

Inoviruses are known for their diverse impacts on host cell phenotypes through many mechanisms, including lysogenic conversion, effecting host transcription with phage encoded regulators and influencing biofilm formation (136, 140, 143). The discovery of these phages associated closely with successful clonal populations of the invasive *V. parahaemolyticus* ST36 clade in the Northeast United States as well as the previous reports of a similar phage in the only pandemic clonal complex, ST3, raises the question of how these phages impact the fitness or virulence of their hosts. In this study we begin to describe the complex relationship of Vipa26, the phage stably maintained in the Gulf of Maine ST36 population, and its host.

Vipa26, an active filamentous phage of the class Inovirus, is closely associated with the highly successful GOM ST36 population, despite the potential burden on the host cell due to production of

progeny phage. Spontaneous phage loss observed under laboratory conditions suggests that there may be a realized fitness cost to the burden of an active, chronic phage infection, yet phage harboring isolates remain clinically prevalent year after year. Naïve strains exposed to Vipa26 also demonstrate significant growth defects during the process of infection and integration of the phage into the host chromosome (Fig. 3.8A). The early plateau in exponential growth corresponds to a significantly higher titer of phage progeny as well as a high load of the replicative form (Fig. 3.9; Fig 3.11A and B). Other species of *Inovirus* slow host growth upon infection through the overexpression of gI (homologous to *zot* through size and location, see Chapter 2), together with membrane stress of secreting virions (182, 204). This can incite the phage shock protein response, a system conserved in many Gram-negative bacterial species that mitigates membrane stress and can be harmful if excessively upregulated (92, 205). The direct burden of diverting valuable resources to production of higher levels of Vipa26 progeny is another likely culprit behind the early plateau of the growth in newly initiated infections. However, persistently infected phage harboring derivatives exhibited no measurable growth defects (Fig. 3.8B) compared to phage deficient isogenic strains even though Vipa26 was actively produced by the chronically infected isolates when grown in rich medium (e.g. HI broth cultures, Fig. 3.8B). The benign nature of stable Vipa26 prophage infection is somewhat unexpected as during chronic infection, some inoviruses do slow the growth of their hosts (183). If even a low level of spontaneous excision Vipa26 such as that observed on agar plates regularly occurred in liquid culture even in a small portion of the population (Fig. 3.6 and 3.7) this could initiate new or highly productive infections generating large bursts of phage by a few individuals without perceivably changing the growth kinetic of the entire population. Vipa26 virions would be produced and only a small minority of the population would be impacted in this benign infection model. However, since total excision of the prophage from the population would mitigate any cost associated with Vipa26 infection, then the persistence of Vipa26 points to a benefit of infection that outweighs any cost.

Maintenance of the replicative form of Vipa26 in MAVP-26 is less than one copy per chromosome (Fig. 3.10), whereas it is maintained in much greater abundance in newly infected strains, correlating to a 625 fold increase in infectivity (Fig. 3.9 and 3.11B). Thus, the fitness cost in stably infected strains could be mitigated by the low proportion of the replicative form and therefore less strain on the host cell to produce progeny. In *Pseudomonas aeruginosa,* superinfection by the inovirus Pf4 leads to increased phage production linked to small colony variants or cell death of the host (136, 137). The down regulation of the replicative form and subsequently phage production in *V. parahaemolyticus* upon integration of the prophage likely rescues the host from the potentially deadly or harmful effects of high virion production. Taken together, these data suggest that once chromosomally integrated, Vipa26 is relatively benign to ST36 hosts, exacting little cost upon achieving a stable, chronic infection.

Certainly, the inability of Vipa26 to produce a plaque on cells harboring similar phage or slow their growth taken with its ability to produce pronounced plaques on the phage-deficient ST36 derivatives (Table 3.5), suggests persistent infection provides protection from growth impairment during the early stages of new infections (Fig. 3.8A). This benefit could be strong enough to outweigh any burden associated with the benign persistent infection as demonstrated by prophage excision and loss. In the dynamic of phages and bacterial hosts, protection against superinfection can push the population structure towards maintaining phage infection (119). This is beneficial to the phage as well as the host, encouraging the vertical propagation of the phage and protection of the bacterium (178, 200). This protection; however, only partially explains the success of a phage harboring strain in competition with a susceptible strain. Phages can also be used as a weapon against naïve competitors, suppressing susceptible strains in mixed culture with hosts with established infections that are protected from superinfection (102). The appearance of newly infected isolates and drastic genotypic shift towards integrated Vipa26 in direct competition supports this sword and shield hypothesis (Fig 3.12). Even in

broth that is not pre-conditioned to contain active virions, the native production of phage by MAVP-26 is sufficient to create new infections in its susceptible competitor and lead to a competitive advantage that depresses the growth of the susceptible strain (Fig. 3.12).

Vipa26 infection was nearly total in the viable cells recovered, and the burden on cells with newly initiated infections was very high likely due to the stresses associated with producing 625 times more Vipa26 progeny than chronically infected strains. Even without the growth suppression compromising cell reproduction and the competitiveness of the susceptible strain, Vipa26 shifted the genotype of the population to nearly entirely phage-harboring. As the initial acute infection progresses, phage chromosomal integration apparently leads to down-regulation of RF production, settling into a stable infection, a process characterized in other types of phage but not well understood in inoviruses other than CTXϕ (92, 206, 207). CTXϕ allows superinfection of tandem copies even after integration; therefore, it is not a good model of the dynamics of integration of protective filamentous phages such as Vipa26 (184, 206). In Vipa26, the production of a large quantity of progeny based on the availability of susceptible hosts that were newly infected prior to this down regulation is advantageous and ensures propagation of the phage through populations. Maintenance of the phage due to some conferred fitness advantage to the host, such as protection from new infections or other traits associated with phage integration, then benefits both the phage and the host which managed to survive the rough initial stages of infection.

Interestingly, the recently infected strain MAVP-20(26) produces a very low amount of the replicative form relative to prophage abundance compared to MAVP-26 (28-76% for MAVP-26 versus 0.1-3.5% for MAVP-20(26)) (Fig. 3.10). This may be related to the opaque/translucent phase variation expressed by this isolate or other strain-specific genetic variation. The low amount of the replicative form in MAVP-20(26) may be a due to the regulatory differences in this strain. Other recently infected strains produce significantly more virions than MAVP-26 (Fig. 3.11), so this down regulation may be

specific to MAVP-20. While there is a gap in understanding in the regulation of integration and phage production of inoviruses, it is likely that host regulators are involved (146, 206, 207). Further study into the regulatory underpinnings of phase variation, and potential relation to phage regulation could broaden our understanding of the toleration of chronic infection and ecological impact of phase variation.

Beyond the direct effects of infection associated with the burden of producing virions, filamentous phages are transducing agents that often introduce novel and potentially useful genetic material to the host cell (92). Although more replicates are required, the potential increase in viability of Vipa26 harboring strains after UV challenge could be due to the phage encoding a gene related to DNA repair and the SOS response (Fig. 3.13). Similar filamentous phages have regulatory systems that are intertwined with host cell encoded proteins such as UvrD. UvrD is involved in mismatch DNA repair and is upregulated during the SOS response, including in response to UV light. CTXϕ and Pf4 in *P. aeruginosa* are UvrD dependent for replication, and close homology of their initiation protein domains compared to that of f237 (208, 209) suggests that Vipa26 could also be UvrD dependent. UvrD is a member of a partially LexA regulated operon in CTXϕ, where phage replication is under control of both LexA and the phage-encoded RstR, a potential homolog to ORF9 based on size and location as well as a low amino acid identity (Chapter 2) (210). In contrast to CTXϕ, Vipa26 lacks LexA binding sites, and therefore it is unlikely that the prophage is triggered to excise in the same way that CTXϕ prophage are. Indeed, strains did not demonstrate more phage loss or progeny phage production after UV-C challenge. Although the phage may influence the SOS response regulators, Vipa26 does not have the LexA binding sites characterized in better studies phages, warranting further research into the molecular basis of the phage-host interaction of Vipa26. Resistance to UV and possibly other forms of stress are a crucial benefit to the bacteria, which are exposed to sunlight and other adverse conditions in their natural environment.

The dynamics of the mixed community *V. parahaemolyticus* is a member of in natural seawater are important influences on ecological fitness. Vipa26-harboring and phage-deficient MAVP-20 variants competing in natural seawater microcosms revealed a potential increase in survival of the phage-harboring isolate, but not in MAVP-26 competing with MAVP-26PD (Fig. 3.15). Complicating the potential success of MAVP-20(26) is the phenotypic switch observed in the phage deficient strain, leading to a translucent phenotype compared to the opaque Vipa26 harboring strain, adding a confounding variable. MAVP-20(26) does not produce high levels of the replicative form compared to MAVP-26 (Fig. 3.9), suggesting there is differential regulation of the phage, potentially influencing the impact of Vipa26 on each isolate. Vipa26 infection alone did not to lead to the switch between opaque and translucent, as several translucent phage-harboring derivatives of MAVP-20 were subsequently recovered (data not shown). The fitness impact of the translucent colony type in *V. parahaemolyticus* is not well understood beyond a study comparing killing by oyster hemocytes (211); however these morphology variants exhibit biofilm deficiencies (Fig. 3.14) (27, 202). As phage deficient isolates derived from MAVP-26 survived as well as both Vipa-26 harboring isolates, MAVP-26 and MAVP20(26), the fitness defect may relate to the translucent phenotype rather than phage content (Fig. 3.15).

Treating microcosms by filter-sterilization and kanamycin treatment microcosms eliminated the fitness benefit of the opaque strains, therefore bacterial competition and/or predation is likely where MAVP-20 is less fit (Fig. 3.16). Deficiency in biofilm, typical of translucent variants and demonstrated by MAVP-20 and the translucent MAVP-20(26)-2 (Fig 3.14), may leave the bacteria more vulnerable to predation by bacteria such as *Halobacteriovorax,* a known predator of *V. parahaemolyticus* (212). The overall lower performance of the strains in kanamycin could be due to the stress of growing in antibiotic even when carrying a resistance cassette, whereas similar survival is due to the removal of the potential competitor/predator. In summary, although phage-harboring MAVP-20(26) was more fit than phage-deficient MAVP-20, this difference could be explained by the additional phenotypic variation that

derivation of this strain produced (opacity) that is not phage-dependent since phage-harboring derivatives were recovered that remained translucent. Examining fitness of the translucent phage-harboring variants in microcosms would address this limitation.

Although Vipa26 did not impact the virulence of MAVP-26 as determined by the amoeba seeded plate assay, the lack of significant differences between the highly virulent ST36 isolates and the environmental G149 calls into question the validity of this method for measuring virulence in *V. parahaemolyticus.* Although, this method has been used previously to virulotype *V. parahaemolyticus,* we were unable to reproduce prior results although notably we did not use the same ST3 strain (Fig. 3.18) (66). More robust methods such as cytotoxicity to human colon cells or *in vivo* pathogenicity studies may reveal differential toxicity. Vipa26 encodes two known putative toxins, *ace* and *zot,* as well as a number of hypothetical proteins located in the same region as *ctxAB* in CTXϕ (12). These toxins and uncharacterized content could contribute the virulence and therefore high clinical prevalence of ST36.

The balance of cost and benefit of mutualistic phage infection drives environmental fitness and evolution (180, 200). Mutualistic association between phage and host occurs when the benefits outweigh the defects for both the phage and its host (119). The relatively benign nature of established infection, protection against superinfection by similar phages and a potentially more robust SOS response may offset the burden of holding Vipa26 exemplified by spontaneous loss, as well as the deficit in growth early in the infective cycle. Further study into the mechanism contributing to potential increased viability after exposure to UV could address the gap in knowledge of the interaction of phage regulation and the SOS response as well as shed light on the potential fitness benefit linked to acquisition of these phage. UV sterilization of circulating seawater during the depuration process in oysters is a common practice (213), greater resistance to killing by UV light could contribute to the survival of ST36 in harvested oysters, directly leading to higher disease incidence. Further application of microcosms as a model for ecological fitness using the translucent phage harboring and phage deficient

strains will confirm if the benefit is due to phenotypic variation or phage content. This study suggests that Vipa26 provides a concrete fitness benefit through protecting against the infection of similar phages suppressing the growth of susceptible strains, along with potentially a better survival of the SOS response; explaining the continuing occurrence of Vipa26 in the population despite a realized fitness cost.

Supplementary Figures

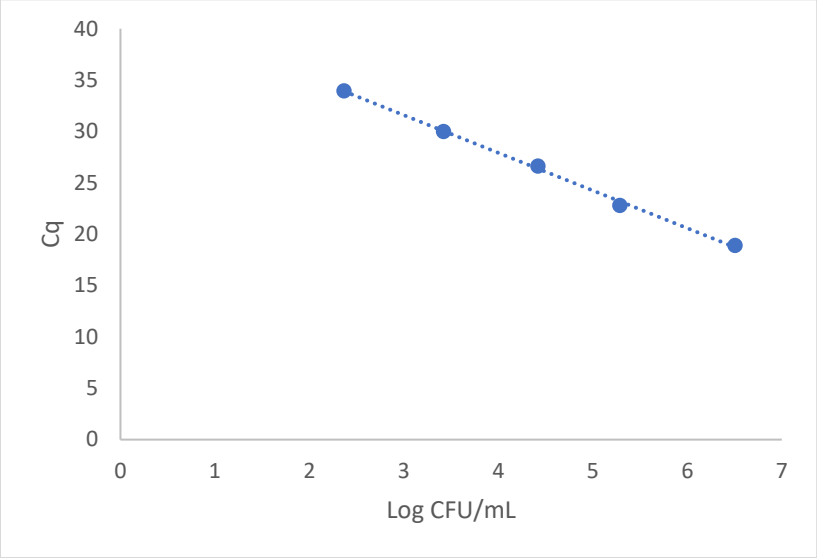

Figure S3.1. Validation of relationship between Cq and cell densities. Crude lysates were analyzed in 40%ASW, contributing to a reduction of primer efficiency to 87%; however, fit of the line is good with R^2=0.998.

LITERATURE CITED

1. Scallan E, Hoekstra RM, Angulo FJ, Tauxe R V., Widdowson MA, Roy SL, Jones JL, Griffin PM. 2011. Foodborne illness acquired in the United States-Major pathogens. Emerg Infect Dis.

2. Scallan E, Griffin PM, Angulo FJ, Tauxe R V., Hoekstra RM. 2011. Foodborne illness acquired in the United states-Unspecified agents. Emerg Infect Dis.

3. Scharff RL. 2012. Economic Burden from Health Losses Due to Foodborne Illness in the United States. J Food Prot.

4. Havelaar AH, Kirk MD, Torgerson PR, Gibb HJ, Hald T, Lake RJ, Praet N, Bellinger DC, de Silva NR, Gargouri N, Speybroeck N, Cawthorne A, Mathers C, Stein C, Angulo FJ, Devleesschauwer B, Adegoke GO, Afshari R, Alasfoor D, Baines J, Balakrishnan K, Hamza WM Bin, Black RE, Bolger PM, Chaicumpa W, Cravioto A, Döpfer D, Ehiri JE, Fazil A, Ferreccio C, Fèvre EM, Hall G, Kasuga F, Keddy KH, Lanata CF, Lei H, Liu X, Manyindo B, Nasinyama G, Ongolo-Zogo P, Pitt JI, Rokni MB, Sripa B, van Leeuwen R, Verger P, Willingham AL, Zhou XN, Aspinall W, Buchanan R, Budke C, Caipo ML, Carabin H, Cole D, Cooke RM, Crump JA, El-Jardali F, Fischer-Walker C, Fürst T, Haagsma JA, Hall AJ, Henao O, Hoffmann S, Jensen H, Jessani N, Koopmans MPG, Levine MM, de Noordhout CM, Majowicz S, McDonald SA, Pires S, Scallan E, Sripa B, Thomas MK, Verhoef L, Wu F, Zeilmaker M. 2015. World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. PLoS Med.

5. Romalde JL, Diéguez AL, Lasa A, Balboa S. 2014. New Vibrio species associated to molluscan microbiota: A review. Front Microbiol.

6. Thompson FL, Iida T, Swings J. 2004. Biodiversity of Vibrios. Microbiol Mol Biol Rev.

7. Okada K, Iida T, Kita-Tsukamoto K, Honda T. 2005. Vibrios commonly possess two chromosomes. J Bacteriol.

8. Trucksis M, Michalski J, Deng YK, Kaper JB. 1998. The Vibrio cholerae genome contains two unique circular chromosomes. Proc Natl Acad Sci.

9. Egidius E, Wiik R, Andersen K, Hoff KA, Hjeltnes B. 1986. *Vibrio salmonicida* sp. nov., a New Fish Pathogen. Int J Syst Bacteriol.

10. Centers for Disease Control and Prevention. 2018. Questions and Answers | Vibrio Illness (Vibriosis). Quest Answers What.

11. Ali M, Nelson AR, Lopez AL, Sack DA. 2015. Updated global burden of cholera in endemic countries. PLoS Negl Trop Dis.

12. Davis BM, Waldor MK. 2003. Filamentous phages linked to virulence of Vibrio cholerae. Curr Opin Microbiol.

13. Harris JB, LaRocque RC, Qadri F, Ryan ET, Calderwood SB. 2012. CholeraThe Lancet.

14. Shamim Hasan Zahid M, Nashir Udden SM, Faruque ASG, Calderwood SB, Mekalanos JJ, Faruque SM. 2008. Effect of phage on the infectivity of Vibrio cholerae and emergence of genetic variants. Infect Immun.

15. De Magny GC, Hasan NA, Roche B. 2014. How community ecology can improve our understanding of cholera dynamics. Front Microbiol.

16. Watson JT, Gayer M, Connolly MA. 2007. Epidemics after natural disasters. Emerg Infect Dis.

17. Shapiro RL, Altekruse S, Hutwagner L, Bishop R, Hammond R, Wilson S, Ray B, Thompson S, Tauxe R V., Griffin PM, Group the VW. 1998. The Role of Gulf Coast Oysters Harvested in Warmer Months in Vibrio vulnificus Infections in the United States, 1988–1996. J Infect Dis.

18. Bross MH, Soch K, Morales R, Mitchell RB. 2007. Vibrio vulnificus infection: Diagnosis and treatment. Am Fam Physician.

19. Jacobs Slifka KM, Newton AE, Mahon BE. 2017. Vibrio alginolyticus infections in the USA, 1988-2012. Epidemiol Infect.

20. Gómez-León J, Villamil L, Lemos ML, Novoa B, Figueras A. 2005. Isolation of Vibrio alginolyticus and Vibrio splendidus from aquacultured carpet shell clam (Ruditapes decussatus) larvae associated with mass mortalities. Appl Environ Microbiol.

21. Dechet AM, Yu PA, Koram N, Painter J. 2008. Nonfoodborne *Vibrio* Infections: An Important Cause of Morbidity and Mortality in the United States, 1997–2006. Clin Infect Dis.

22. Fujino T, Okuno D, Nakada A, Aoyama K, Fukai T, Mukai, Ueho. 1953. On the bacteriological examination of Shirasu-food poisoning. Med J Osaka Univ.

23. Jahangir Alam M, Tomochika KI, Miyoshi SI, Shinoda S. 2002. Environmental investigation of potentially pathogenic Vibrio parahaemolyticus in the Seto-Inland Sea, Japan. FEMS Microbiol Lett.

24. Joseph SW, Colwell RR, Kaper JB. 1982. Vibrio parahaemolyticus and related halophilic vibrios. Crit Rev Microbiol.

25. Liu B, Liu H, Pan Y, Xie J, Zhao Y. 2016. Comparison of the effects of environmental parameters on the growth variability of Vibrio parahaemolyticus coupled with strain sources and genotypes analyses. Front Microbiol.

26. Ellis CN, Schuster BM, Striplin MJ, Jones SH, Whistler CA, Cooper VS. 2012. Influence of seasonality on the genetic diversity of Vibrio parahaemolyticus in new hampshire shellfish waters as determined by multilocus sequence analysis. Appl Environ Microbiol.

27. Hsieh YC, Liang SM, Tsai WL, Chen YH, Liu TY, Liang CM. 2003. Study of capsular polysaccharide from Vibrio parahaemolyticus. Infect Immun.

28. Kaneko T, Colwell RR. 1973. Ecology of Vibrio parahaemolyticus in Chesapeake Bay. J Bacteriol.

29. Kaneko T, Colwell RR. 1975. Adsorption of Vibrio parahaemolyticus onto Chitin and Copepods. Appl Envir Microbiol.

30. Kaysner CA, Abeyta C, Stott RF, Lilja JL, Wekell MM. 1990. Incidence of urea-hydrolyzing Vibrio parahaemolyticus in Willapa Bay, Washington. Appl Environ Microbiol.

31. Zimmerman AM, DePaola A, Bowers JC, Krantz JA, Nordstrom JL, Johnson CN, Grimes DJ. 2007. Variability of total and pathogenic vibrio parahaemolyticus densities in Northern Gulf of Mexico water and oysters. Appl Environ Microbiol.

32. Burnham VE, Janes ME, Jakus LA, Supan J, Depaola A, Bell J. 2009. Growth and survival differences of vibrio vulnificus and vibrio parahaemolyticus strains during cold storage. J Food Sci.

33. Urquhart EA, Jones SH, Yu JW, Schuster BM, Marcinkiewicz AL, Whistler CA, Cooper VS. 2016. Environmental Conditions Associated with Elevated Vibrio parahaemolyticus Concentrations in Great Bay Estuary, New Hampshire. PLoS One.

34. Vezzulli L, Brettar I, Pezzati E, Reid PC, Colwell RR, Höfle MG, Pruzzo C. 2012. Long-term effects of ocean warming on the prokaryotic community: Evidence from the vibrios. ISME J.

35. Sudheesh PS, Xu HS. 2001. Pathogenicity of Vibrio parahaemolyticus in tiger prawn Penaeus monodon Fabricius: Possible role of extracellular proteases. Aquaculture.

36. Lee KK, Liu PC, Huang CY. 2003. Vibrio parahaemolyticus infectious for both humans and edible mollusk abalone. Microbes Infect.

37. Alcaide E, Amaro C, Todolí R, Oltra R. 1999. Isolation and characterization of Vibrio parahaemolyticus causing infection in Iberian toothcarp Aphanius iberus. Dis Aquat Organ.

38. Frischer ME, Thurmond JM, Paul JH. 1990. Natural plasmid transformation in a high-frequency-of-transformation marine Vibrio strain. Appl Environ Microbiol.

39. Chimalapati S, Santos M de S, Servage K, De Nisco NJ, Dalia AB, Orth K. 2018. Natural transformation in Vibrio parahaemolyticus: A rapid method to create genetic deletions. J Bacteriol.

40. Chen Y, Dai J, Morris JG, Johnson JA. 2010. Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic Vibrio parahaemolyticus O3:K6. BMC Microbiol.

41. Gode-Potratz CJ, McCarter LL. 2011. Quorum sensing and silencing in Vibrio parahaemolyticus. J Bacteriol.

42. Xu F, Gonzalez-Escalon N, Drees KP, Sebra RP, Cooper VS, Jones SH, Whistler CA. 2017. Parallel evolution of two clades of an Atlantic-endemic pathogenic lineage of Vibrio parahaemolyticus by independent acquisition of related pathogenicity islands. Appl Environ Microbiol.

43. Newton A, 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.

44. Centers for Disease Control and Prevention. 2018. Cholera and Other Vibrio Illness Surveillance (COVIS).

45. Miliotis M, Dennis S, Buchanan R, Potter M. 2008. Role of epidemiology in microbial risk assessment. Food Addit Contam Part A 25:1052–1057.

46. Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe R V. 1999. Food-related illness and death in the United States. Emerg Infect Dis 5:607–625.

47. Lovell CR. 2017. Ecological fitness and virulence features of Vibrio parahaemolyticus in estuarine environments. Appl Microbiol Biotechnol.

48. Nishibuchi M, Fasano A, Russell RG, Kaper JB. 1992. Enterotoxigenicity of Vibrio parahaemolyticus with and without genes encoding thermostable direct hemolysin. Infect

Immun.

49. Letchumanan V, Chan KG, Lee LH. 2014. Vibrio parahaemolyticus: A review on the pathogenesis, prevalence, and advance molecular identification techniques. Front Microbiol.

50. Bhoopong P, Palittapongarnpim P, Pomwised R, Kiatkittipong A, Kamruzzaman M, Nakaguchi Y, Nishibuchi M, Ishibashi M, Vuddhakul V. 2007. Variability of properties of Vibrio parahaemolyticus strains isolated from individual patients. J Clin Microbiol.

51. Hazen TH, Lafon PC, Garrett NM, Lowe TM, Silberger DJ, Rowe LA, Frace M, Parsons MB, Bopp CA, Rasko DA, Sobecky PA. 2015. Insights into the environmental reservoir of pathogenic Vibrio parahaemolyticus using comparative genomics. Front Microbiol.

52. Gutierrez West CK, Klein SL, Lovell harles R. 2013. High frequency of virulence factor genes tdh, trh, and tlh in Vibrio parahaemolyticus strains isolated from a pristine estuary. Appl Environ Microbiol.

53. Matsuda S, Kodama T, Okada N, Okayama K, Honda T, Iida T. 2010. Association of Vibrio parahaemolyticus thermostable direct hemolysin with lipid rafts is essential for cytotoxicity but not hemolytic activity. Infect Immun.

54. Honda T, Ni Y, Miwatani T, Adachi T, Kim J. 1992. The thermostable direct hemolysin of *Vibrio parahaemolyticus* is a pore-forming toxin. Can J Microbiol.

55. Kishishita M, Matsuoka N, Kumagai K, Yamasaki S, Takeda Y, Nishibuchi M. 1992. Sequence variation in the thermostable direct hemolysin-related hemolysin (trh) gene of Vibrio parahaemolyticus. Appl Environ Microbiol.

56. Takahashi A, Kenjyo N, Imura K, Myonsun Y, Honda T. 2000. Cl- secretion in colonic epithelial cells induced by the Vibrio parahaemolyticus hemolytic toxin related to thermostable direct hemolysin. Infect Immun.

57. Park KS, Ono T, Rokuda M, Jang MH, Iida T, Honda T. 2004. Cytotoxicity and Enterotoxicity of the Thermostable Direct Hemolysin-Deletion Mutants of Vibrio parahaemolyticus. Microbiol Immunol.

58. Ming X, Yamamoto K, Honda T. 1994. Construction and characterization of an isogenic mutant of Vibrio parahaemolyticus having a deletion in the thermostable direct hemolysin- related hemolysin gene (trh). J Bacteriol.

59. Boyd EF, Cohen AL V., Naughton LM, Ussery DW, Binnewies TT, Stine OC, Parent MA. 2008. Molecular analysis of the emergence of pandemic Vibrio parahaemolyticus. BMC Microbiol.

60. Hazen TH, Pan L, Gu JD, Sobecky PA. 2010. The contribution of mobile genetic elements to the evolution and ecology of Vibrios. FEMS Microbiol Ecol.

61. Hurley CC, Quirke AM, Reen FJ, Boyd EF. 2006. Four genomic islands that mark post-1995 pandemic Vibrio parahaemolyticus isolates. BMC Genomics.

62. Okada N, Iida T, Park KS, Goto N, Yasunaga T, Hiyoshi H, Matsuda S, Kodama T, Honda T. 2009. Identification and characterization of a novel type III secretion system in trh-positive vibrio parahaemolyticus strain TH3996 reveal genetic lineage and diversity of pathogenic machinery beyond the species level. Infect Immun.

63. Zhou X, Gewurz BE, Ritchie JM, Takasaki K, Greenfeld H, Kieff E, Davis BM, Waldor MK. 2013. A Vibrio parahaemolyticus T3SS Effector Mediates Pathogenesis by Independently Enabling Intestinal Colonization and Inhibiting TAK1 Activation. Cell Rep.

64. Hiyoshi H, Kodama T, Saito K, Gotoh K, Matsuda S, Akeda Y, Honda T, Iida T. 2011. VopV, an F-actin-binding type III secretion effector, is required for vibrio parahaemolyticus-induced enterotoxicity. Cell Host Microbe.

65. Makino K, Oshima K, Kurokawa K, Yokoyama K, Uda T, Tagomori K, Iijima Y, Najima M, Nakano M, Yamashita A, Kubota Y, Kimura S, Yasunaga T, Honda T, Shinagawa H, Hattori M, Iida T. 2003. Genome sequence of Vibrio parahaemolyticus: A pathogenic mechanism distinct from that of V cholerae. Lancet.

66. Matz C, Nouri B, McCarter L, Martinez-Urtaza J. 2011. Acquired type III secretion system determines environmental fitness of epidemic vibrio parahaemolyticus in the interaction with bacterivorous protists. PLoS One.

67. Chatterjee BD, Sen T. 1974. Vibrio parahaemolyticus serotypes in Calcutta, India. Bull World Health Organ.

68. Abbott SL, Powers C, Kaysner CA, Takeda Y, Ishibashi M, Joseph SW, Janda JM. 1989. Emergence of a restricted bioserovar of Vibrio parahaemolyticus as the predominant cause of Vibrio-associated gastroenteritis on the West Coast of the United States and Mexico. J Clin Microbiol.

69. 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.

70. Martinez-Urtaza J, Blanco-Abad V, Rodriguez-Castro A, Ansede-Bermejo J, Miranda A, Rodriguez-Alvarez MX. 2012. Ecological determinants of the occurrence and dynamics of Vibrio parahaemolyticus in offshore areas. ISME J.

71. Martinez-Urtaza J, Huapaya B, Gavilan RG, Blanco-Abad V, Ansede-Bermejo J, Cadarso-Suarez C, Figueiras A, Trinanes J. 2008. Emergence of asiatic vibrio diseases in south america in phase with El Niño. Epidemiology.

72. Ruiz GM, Rawlings TK, Dobbs FC, Drake L a, Mullady T, Huq A, Colwell RR. 2000. Global spread of microorganisms by ships. Nature.

73. Depaola A, Capers GM, Motes ML, Olsvik O, Fields PI, Wells J, Kaye Wachsmuth I, Cebula TA, Koch WH, Khambaty F, Kothary MH, Les Payne W, Wentz BA. 1992. Isolation of Latin American epidemic strain of Vibrio cholerae 01 from US Gulf Coast. Lancet.

74. 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.

75. Newton AE, Garret N, Stroika SG, Haplin JL, Turnsek M, Mody RK. 2014. Increase in *Vibrio parahaemolyticus* infections associated with consumption of Atlantic Coast Shellfish - 2013. Morb Mortal Wkly Rep.

76. Martinez-Urtaza J, Van Aerle R, Abanto M, Haendiges J, Myers RA, Trinanes J, Baker-Austin C, Gonzalez-Escalona N. 2017. Genomic variation and evolution of vibrio parahaemolyticus ST36 over the course of a transcontinental epidemic expansion. MBio.

77. 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.

78. 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.

79. González-Escalona N, Martinez-Urtaza J, Romero J, Espejo RT, Jaykus LA, DePaola A. 2008. Determination of molecular phylogenetics of Vibrio parahaemolyticus strains by multilocus sequence typing. J Bacteriol.

80. 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.

81. Daniels N a, Ray B, Easton a, Marano N, Kahn E, McShan a L, Del Rosario L, Baldwin T, Kingsley M a, Puhr ND, Wells JG, Angulo FJ. 2000. Emergence of a new Vibrio parahaemolyticus serotype in raw oysters: A prevention quandary. JAMA.

82. 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. Nat Clim Chang.

83. Vezzulli L, Colwell RR, Pruzzo C. 2013. Ocean Warming and Spread of Pathogenic Vibrios in the Aquatic Environment. Microb Ecol.

84. d'Hérelle F. 1921. Le Bactériophage: Son Rôle dans l'Immunité.

85. Smith J. 1924. The bacteriophage in the treatment of typhoid fever. Br Med J.

86. Bruynoghe R, Maisin J. 1921. Essais de thérapeutique au moyen du bactériophage. C R Biol 85:1120–1121.

87. Kutter E, De Vos D, Gvasalia G, Alavidze Z, Gogokhia L, Kuhl S, Abedon S. 2010. Phage Therapy in Clinical Practice: Treatment of Human Infections. Curr Pharm Biotechnol.

88. Pelfrene E, Willebrand E, Cavaleiro Sanches A, Sebris Z, Cavaleri M. 2016. Bacteriophage therapy: A regulatory perspective. J Antimicrob Chemother.

89. Hanlon GW. 2007. Bacteriophages: an appraisal of their role in the treatment of bacterial infections. Int J Antimicrob Agents.

90. Demerec M, Fano U. 1945. Bacteriophage-Resistant Mutants in Escherichia Coli. Genetics.

91. Karam JD, Miller ES. 2010. Bacteriophage T4 and its relatives. Virol J 7:293.

92. Mai-Prochnow A, Hui JGK, Kjelleberg S, Rakonjac J, McDougald D, Rice SA. 2015. "Big things in small packages: The genetics of filamentous phage and effects on fitness of their host." FEMS Microbiol Rev.

93. Breitbart M, Rohwer F. 2005. Here a virus, there a virus, everywhere the same virus? Trends Microbiol 13:278–284.

94. Clokie MRJ, Millard AD, Letarov A V, Heaphy S. 2011. Phages in nature. Bacteriophage 1:31–45.

95. TAYLOR MJ, THORNE CB. 1963. TRANSDUCTION OF BACILLUS LICHENIFORMIS AND BACILLUS SUBTILIS BY EACH OF TWO PHAGES. J Bacteriol.

96. ROMIG WR, BRODETSKY AM. 1961. Isolation and preliminary characterization of bacteriophages for Bacillus subtilis. J Bacteriol.

97. Buckling A, Rainey PB. 2002. Antagonistic coevolution between a bacterium and a bacteriophage. Proc R Soc B Biol Sci.

98. Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pašić L, Thingstad TF, Rohwer F, Mira A. 2009. Explaining microbial population genomics through phage predation. Nat Rev Microbiol.

99. Petrov VM, Ratnayaka S, Nolan JM, Miller ES, Karam JD. 2010. Genomes of the T4-related bacteriophages as windows on microbial genome evolution. Virol J 7:292.

100. Casjens S. 2003. Prophages and bacterial genomics: What have we learned so far? Mol Microbiol.

101. Bohannan BJM, Lenski RE. 2000. The Relative Importance of Competition and Predation Varies with Productivity in a Model Community. Am Nat.

102. Joo J, Gunny M, Cases M, Hudson P, Albert R, Harvill E. 2006. Bacteriophage-mediated competition in Bordetella bacteria. Proc R Soc B Biol Sci.

103. Furuse K, Osawa S, Kawashiro J, Tanaka R, Ozawa A, Sawamura S, Yanagawa Y, Nagao T, Watanabe I. 1983. Bacteriophage distribution in human faeces: Continuous survey of healthy subjects and patients with internal and leukaemic diseases. J Gen Virol.

104. Furuse K. 1987. Distribution of coliphages in the environment: general considerations. John Wiley & Sons, New York.

105. Ashelford KE, Day MJ, Fry JC. 2003. Elevated abundance of bacteriophage infecting bacteria in soil. Appl Environ Microbiol.

106. Danovaro R, Manini E, Dell'Anno A. 2002. Higher abundance of bacteria than of viruses in deep Mediterranean sediments. Appl Environ Microbiol.

107. Bergh Ø, Børsheim KY, Bratbak G, Heldal M. 1989. High abundance of viruses found in aquatic environments. Nature.

108. Ritchie JM, Wagner PL, Acheson DWK, Waldor MK. 2003. Comparison of Shiga toxin production by hemolytic-uremic syndrome-associated and bovine-associated Shiga toxin-producing Escherichia coli isolates. Appl Environ Microbiol.

109. Middelboe M, Hagström A, Blackburn N, Sinn B, Fischer U, Borch NH, Pinhassi J, Simu K, Lorenz MG. 2001. Effects of bacteriophages on the population dynamics of four strains of pelagic marine bacteria. Microb Ecol.

110. Faruque SM, Naser I Bin, Islam MJ, Faruque ASG, Ghosh AN, Nair GB, Sack DA, Mekalanos JJ. 2005. Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. Proc Natl Acad Sci U S A 102:1702 LP – 1707.

111. Jensen MA, Faruque SM, Mekalanos JJ, Levin BR. 2006. Modeling the role of bacteriophage in the control of cholera outbreaks. Proc Natl Acad Sci.

112. Faruque SM, Islam MJ, Ahmad QS, Faruque ASG, Sack DA, Nair GB, Mekalanos JJ. 2005. Self-

limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. Proc Natl Acad Sci.

113. Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. Nat Rev Microbiol 8:317.

114. Jansen R, Van Embden JDA, Gaastra W, Schouls LM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol.

115. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (80-).

116. Sorek R, Kunin V, Hugenholtz P. 2008. CRISPR - A widespread system that provides acquired resistance against phages in bacteria and archaea. Nat Rev Microbiol.

117. Dupuis MÈ, Villion M, Magadán AH, Moineau S. 2013. CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. Nat Commun.

118. Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. Nat Rev Microbiol.

119. Chibani-Chennoufi S, Bruttin A, Dillmann M-L, Brüssow H. 2004. Phage-Host Interaction: an Ecological Perspective. J Bacteriol 186:3677 LP – 3686.

120. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GPC. 2009. The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair. Proc Natl Acad Sci 106:894 LP – 899.

121. Forde A, Fitzgerald GF. 1999. Bacteriophage defence systems in lactic acid bacteriaAntonie van Leeuwenhoek, International Journal of General and Molecular Microbiology.

122. Wang X, Kim Y, Wood TK. 2009. Control and benefits of CP4-57 prophage excision in Escherichia coli biofilms. Isme J 3:1164.

123. Lopez CA, Winter SE, Rivera-Chávez F, Xavier MN, Poon V, Nuccio S-P, Tsolis RM, Bäumler AJ. 2012. Phage-Mediated Acquisition of a Type III Secreted Effector Protein Boosts Growth of Salmonella by Nitrate Respiration. MBio 3:e00143-12.

124. Obeng N, Pratama AA, Elsas JD van. 2016. The Significance of Mutualistic Phages for Bacterial Ecology and Evolution. Trends Microbiol 24:440–449.

125. Carrolo M, Frias MJ, Pinto FR, Melo-Cristino J, Ramirez M. 2010. Prophage spontaneous activation promotes DNA release enhancing biofilm formation in Streptococcus pneumoniae. PLoS One.

126. Gödeke J, Paul K, Lassak J, Thormann KM. 2010. Phage-induced lysis enhances biofilm formation in Shewanella oneidensis MR-1. Isme J 5:613.

127. Nedialkova LP, Sidstedt M, Koeppel MB, Spriewald S, Ring D, Gerlach RG, Bossi L, Stecher B. 2016. Temperate phages promote colicin-dependent fitness of Salmonella enterica serovar Typhimurium. Environ Microbiol.

128. Bondy-Denomy J, Davidson AR. 2014. When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. J Microbiol 52:235–242.

129. Keen EC, Dantas G. 2018. Close Encounters of Three Kinds: Bacteriophages, Commensal Bacteria,

and Host Immunity. Trends Microbiol.

130. Eklund MW, Poysky FT, Reed SM, Smith CA. 1971. Bacteriophage and the toxigenicity of Clostridium botulinum type C. Science (80-) 172:480–482.

131. FREEMAN VJ. 1951. Studies on the virulence of bacteriophage-infected strains of Corynebacterium diphtheriae. J Bacteriol.

132. Fortier LC, Sekulovic O. 2013. Importance of prophages to evolution and virulence of bacterial pathogens. Virulence.

133. Brussow H, Canchaya C, Hardt W-D. 2004. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. Microbiol Mol Biol Rev.

134. Rakonjac J, Bennett NJ, Spagnuolo J, Gagic D, Russel M. 2009. Filamentous Bacteriophage: Biology, Phage Display and Nanotechnology Applications. Curr Issues Mol Biol.

135. Frost LS. 1993. Conjugative Pili and Pilus-Specific PhagesBacterial Conjugation.

136. Webb JS, Lau M, Kjelleberg S. 2004. Bacteriophage and phenotypic variation in Pseudomonas aeruginosa biofilm development. J Bacteriol.

137. Webb JS, Thompson LS, James S, Charlton T, Tolker-Nielsen T, Koch B, Givskov M, Kjelleberg S. 2003. Cell death in Pseudomonas aeruginosa biofilm development. J Bacteriol.

138. Barken KB, Pamp SJ, Yang L, Gjermansen M, Bertrand JJ, Klausen M, Givskov M, Whitchurch CB, Engel JN, Tolker-Nielsen T. 2008. Roles of type IV pili, flagellum-mediated motility and extracellular DNA in the formation of mature multicellular structures in Pseudomonas aeruginosa biofilms. Environ Microbiol.

139. Gödeke J, Paul K, Lassak J, Thormann KM. 2011. Phage-induced lysis enhances biofilm formation in Shewanella oneidensis MR-1. ISME J.

140. Jian H, Xiao X, Wanga F. 2013. Role of filamentous phage SW1 in regulating the lateral flagella of shewanella piezotolerans strain WP3 at low temperatures. Appl Environ Microbiol.

141. Wang F, Wang F, Li Q, Xiao X. 2007. A novel filamentous phage from the deep-sea bacterium Shewanella piezotolerans WP3 is induced at low temperature. J Bacteriol.

142. Karlsson F, Malmborg-Hager A-C, Albrekt A-S, Borrebaeck CA. 2005. Genome-wide comparison of phage M13-infected vs. uninfected *Escherichia coli*. Can J Microbiol.

143. Waldor MK, Mekalanos JJ. 1996. Lysogenic Conversion by a Filamentous Phage Encoding Cholera Toxin. Science (80-).

144. Choi S, Dunams D, Jiang SC. 2010. Transfer of cholera toxin genes from O1 to non-O1/O139 strains by vibriophages from California coastal waters. J Appl Microbiol 108:1015–1022.

145. Kamruzzaman M, Robins WP, Nayeemul Bari SM, Nahar S, Mekalanos JJ, Faruque SM. 2014. RS1 satellite phage promotes diversity of toxigenic vibrio cholerae by driving CTX prophage loss and elimination of lysogenic immunity. Infect Immun.

146. Kawasaki T, Nagata S, Fujiwara A, Satsuma H, Fujie M, Usami S, Yamada T. 2007. Genomic characterization of the filamentous integrative bacteriophages ΦRSS1 and ΦRSM1, which infect Ralstonia solanacearum. J Bacteriol.

147. Addy HS, Askora A, Kawasaki T, Fujie M, Yamada T. 2012. The Filamentous Phage φRSS1 Enhances Virulence of Phytopathogenic *Ralstonia solanacearum* on Tomato. Phytopathology.

148. Addy HS, Askora A, Kawasaki T, Fujie M, Yamada T. 2012. Loss of Virulence of the Phytopathogen *Ralstonia solanacearum* Through Infection by ϕRSM Filamentous Phages. Phytopathology.

149. Iida T, Hattori A, Tagomori K, Nasu H, Naim R, Honda T. 2001. Filamentous phage associated with recent pandemic strains of Vibrio parahaemolyticus. Emerg Infect Dis.

150. 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.

151. Bhuiyan NA, Ansaruzzaman M, Kamruzzaman M, Alam K, Chowdhury NR, Nishibuchi M, Faruque SM, Sack DA, Takeda Y, Nair GB. 2002. Prevalence of the pandemic genotype of Vibrio parahaemolyticus in Dhaka, Bangladesh, and significance of its distribution across different serotypes. J Clin Microbiol.

152. Jensen MA, Faruque SM, Mekalanos JJ, Levin BR. 2006. Modeling the role of bacteriophage in the control of cholera outbreaks. Proc Natl Acad Sci U S A.

153. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med.

154. Jones SH, Striplin M, Mahoney J, Cooper VS, Whistler CA. 2010. Incidence and abundance of pathogenic Vibrio species in the Great Bay Estuary, New Hampshire, p. 127–134. *In* Proceedings of the Seventh International Conference on Molluscan Shellfish Safety. Quae Publishing, Nantes, France.

155. 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.

156. 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 Environ Microbiol.

157. 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.

158. Tritt A, Eisen JA, Facciotti MT, Darling AE. 2012. An Integrated Pipeline for de Novo Assembly of Microbial Genomes. PLoS One.

159. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol.

160. Darling AE, Mau B, Perna NT. 2010. Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS One.

161. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: A genome comparison visualizer. Bioinformatics.

162. Roux S, Krupovic M, Daly RA, Borges AL, Nayfach S, Schulz F, Sharrar A, Matheus Carnevali PB, Cheng JF, Ivanova NN, Bondy-Denomy J, Wrighton KC, Woyke T, Visel A, Kyrpides NC, Eloe-Fadrosh EA. 2019. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's

biomes. Nat Microbiol.

163. Krupovic M, Dutilh BE, Adriaenssens EM, Wittmann J, Vogensen FK, Sullivan MB, Rumnieks J, Prangishvili D, Lavigne R, Kropinski AM, Klumpp J, Gillis A, Enault F, Edwards RA, Duffy S, Clokie MRC, Barylski J, Ackermann HW, Kuhn JH. 2016. Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. Arch Virol.

164. Adriaenssens EM, Rodney Brister J. 2017. How to name and classify your phage: An informal guide. Viruses.

165. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol.

166. Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426.

167. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol.

168. Gardner SN, Slezak T, Hall BG. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics.

169. Letunic I, Bork P. 2011. Interactive Tree of Life v2: Online annotation and display of phylogenetic trees made easy. Nucleic Acids Res.

170. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics.

171. Meier-Kolthoff JP, Göker M. 2017. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics.

172. Farris JS. 1972. Estimating Phylogenetic Trees from Distance Matrices. Am Nat.

173. Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics.

174. Whistler CA, Hall JA, Xu F, Ilyas S, Siwakoti P, Cooper VS, Jones SH. 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.

175. Marcinkiewicz A. 2016. Bacterial and phage interactions influencing Vibrio parahaemolyticus ecology. University of New Hampshire.

176. Haendiges J, Jones J, Myers RA, Mitchell CS, Butler E, Toro M, Gonzalez-Escalona N. 2016. A nonautochthonous U.S. strain of Vibrio parahaemolyticus isolated from Chesapeake Bay oysters caused the outbreak in Maryland in 2010. Appl Environ Microbiol.

177. Chang B, Taniguchi H, Miyamoto H, Yoshida SI. 1998. Filamentous bacteriophages of vibrio parahaemolyticus as a possible clue to genetic transmission. J Bacteriol.

178. Shapiro JW, Turner PE. 2018. Evolution of mutualism from parasitism in experimental virus populations. Evolution (N Y).

179. Chibani-Chennoufi S, Bruttin A, Dillmann ML, Brüssow H. 2004. Phage-host interaction: An ecological perspective. J Bacteriol.

180. Obeng N, Pratama AA, Elsas JD van. 2016. The Significance of Mutualistic Phages for Bacterial Ecology and Evolution. Trends Microbiol.

181. Rakonjac J. 2012. Filamentous Bacteriophages: Biology and ApplicationseLS.

182. Horabin JI, Webster RE. 1986. Morphogenesis of f1 filamentous bacteriophage. Increased expression of gene I inhibits bacterial growth. J Mol Biol.

183. Kuo T teh, Tan M shin, Su M tsan, Yang M kwei. 1991. Complete nucleotide sequence of filamentous phage Cf1C from Xanthomonas campestris pv. citri. Nucleic Acids Res.

184. Kimsey HH, Waldor MK. 2009. Vibrio cholerae LexA coordinates CTX prophage gene expression. J Bacteriol.

185. McLeod SM, Kimsey HH, Davis BM, Waldor MK. 2005. CTXϕ and Vibrio cholerae: Exploring a newly recognized type of phage-host cell relationship. Mol Microbiol.

186. Laskowski-Arce MA, Orth K. 2008. Acanthamoeba castellanii promotes the survival of Vibrio parahaemolyticus. Appl Environ Microbiol.

187. Nanda AM, Thormann K, Frunzke J. 2015. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. J Bacteriol.

188. García K, Bastías R, Higuera G, Torres R, Mellado A, Uribe P, Espejo RT. 2013. Rise and fall of pandemic Vibrio parahaemolyticus serotype O3: K6 in southern Chile. Environ Microbiol.

189. Acanthamoeba castellanii (Douglas) Page (ATCC® 30234™).

190. Cafeteria roenbergensis Fenchel and Patterson (ATCC® 50301™).

191. Kolter R, Helinski DR. 1978. Construction of plasmid R6K derivatives in vitro: Characterization of the R6K replication region. Plasmid.

192. Stabb E V., Ruby EG. 2002. RP4-based plasmids for conjugation between Escherichia coli and members of the vibrionaceae. Methods Enzymol.

193. B F. 2000. The identification and characterization of a Vibrio fischeri hemagglutination factor. University of Hawaii.

194. Dunn AK, Millikan DS, Adin DM, Bose JL, Stabb E V. 2006. New rfp- and pES213-derived tools for analyzing symbiotic Vibrio fischeri reveal patterns of infection and lux expression in situ. Appl Environ Microbiol.

195. Sabrina Pankey M, Foxall RL, Ster IM, Perry LA, Schuster BM, Donner RA, Coyle M, Cooper VS, Whistler CA. 2017. Host-selected mutations converging on a global regulator drive an adaptive leap towards symbiosis in bacteria. Elife.

196. New England Biolabs. M13 Amplification.

197. 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 Environ Microbiol.

198. O'Toole GA, Kolter R. 1998. Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. Mol Microbiol.

199. Albers U, Reus K, Shuman HA, Hilbi H. 2005. The amoebae plate test implicates a paralogue of lpxB in the interaction of Legionella pneumophila with Acanthamoeba castellanii. Microbiology.

200. Bondy-Denomy J, Davidson AR. 2014. When a virus is not a parasite: The beneficial effects of prophages on bacterial fitness. J Microbiol.

201. Solovyev V, Salamov A. 2011. Automatic annotation of microbial genomes and metagenomic sequencesMetagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies.

202. McCarter LL. 1998. OpaR, a homolog of Vibrio harveyi LuxR, controls opacity of Vibrio parahaemolyticus. J Bacteriol.

203. Gotoh K, Kodama T, Hiyoshi H, Izutsu K, Park KS, Dryselius R, Akeda Y, Honda T, Iida T. 2010. Bile acid-induced virulence gene expression of Vibrio parahaemolyticus reveals a novel therapeutic potential for bile acid sequestrants. PLoS One.

204. Horabin JI, Webster RE. 1988. An amino acid sequence which directs membrane insertion causes loss of membrane potential. J Biol Chem.

205. Mitchell AM, Silhavy TJ. 2019. Envelope stress responses: balancing damage repair and toxicity. Nat Rev Microbiol.

206. Rakonjac J, Bennett NJ, Spagnuolo J, Gagic D, Russel M. 2011. Filamentous bacteriophage: biology, phage display and nanotechnology applications. Curr Issues Mol Biol.

207. Yamada T. 2013. Filamentous phages of Ralstonia solanacearum: Double-edged swords for pathogenic bacteria. Front Microbiol.

208. Martínez E, Campos-Gómez J. 2016. Pf Filamentous Phage Requires UvrD for Replication in *Pseudomonas aeruginosa*. mSphere.

209. Martínez E, Paly E, Barre FX. 2015. CTXϕ Replication Depends on the Histone-Like HU Protein and the UvrD Helicase. PLoS Genet.

210. Quinones M, Kimsey HH, Waldor MK. 2005. LexA cleavage is required for CTX prophage induction. Mol Cell.

211. Genthner FJ, Volety AK, Oliver LM, Fisher WS. 1999. Factors influencing in vitro killing of bacteria by hemocytes of the eastern oyster (Crassostrea virginica). Appl Environ Microbiol.

212. Williams HN, Lymperopoulou DS, Athar R, Chauhan A, Dickerson TL, Chen H, Laws E, Berhane TK, Flowers AR, Bradley N, Young S, Blackwood D, Murray J, Mustapha O, Blackwell C, Tung Y, Noble RT. 2016. Halobacteriovorax, an underestimated predator on bacteria: Potential impact relative to viruses on bacterial mortality. ISME J.

213. Sunnotel O., Snelling WJ., McDonough N., Browne L., Moore JE., Dooley JSG., Lowery CJ. d. 2007. Effectiveness of standard UV depuration at inactivating Cryptosporidium parvum recovered from spiked Pacific oysters (Crassostrea gigas). Appl Environ Microbiol.