Inoviridae prophage dynamics during diversification, succession and Atlantic invasion of Pacific-native *Vibrio parahaemolyticus*.

Jillian Means[1,2*], Ashley L. Marcinkiewicz[1,2,3*], Randi L. Foxall[1,4*], Christopher Schillaci[1,5,6], Kristin DeRosia-Banick[1,7], Feng Xu[1,8], Jeffrey A. Hall[1,9], Stephen H. Jones[1,5], Vaughn S. Cooper[1,10], Cheryl A. Whistler[1,4,§]

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

[2] Graduate program in Microbiology, University of New Hampshire, Durham, New Hampshire, USA

[3] Current address: Division of Infectious Disease, Wadsworth Center, New York State, Department of Health, Albany, New York, USA

[4] Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, USA

[5]Department of Natural Resources and the Environment, University of New Hampshire, Durham, New Hampshire, USA

[6] Current address: NOAA Fisheries Service, Greater Atlantic Regional Fisheries Office, Gloucester Massachusetts

[7]Milken School of Public Health, George Washington University, Washington D.C., USA

[8]Current address: Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

[9] Hubbard Center for Genome Studies, University of New Hampshire, Durham, NH, USA

[10] Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

* Contributed equally to this work

[§]Corresponding Author: Cheryl.Whistler@unh.edu

Running Title: Inovirus-mediated evolution of ST36 *Vibrio parahaemolyticus*

Abstract: The epidemiology of *Vibrio parahaemolyticus,* the leading cause of seafood-borne gastroenteritis world-wide, dramatically changed in the United States following the translocation and establishment of a Pacific lineage of *Vibrio parahaemolyticus*, sequence type (ST) 36, into the Atlantic. In this study we used phylogeography based on traceback locations and comparative genomics to identify features that promoted evolution, dispersal, and competitive dominance of ST36. This revealed a striking shuffling of filamentous prophage in the family *Inoviridae* (inoviruses), including loss of a prophage that had been maintained for decades, accompanied modern diversification and competitive replacement of the endemic north Pacific population. Subsequently, at least five distinct progenitors arising from this diversification translocated from the Pacific into the Atlantic and established four geographically defined clonal subpopulations with remarkably low migration or mixing. Founders of two prevailing Atlantic subpopulations each acquired new diagnostic inoviruses while other subpopulations that apparently declined did not. Broader surveys indicate inoviruses are common and active among the global population of *V. parahaemolyticus* and though parallel inovirus replacements, such as in ST36, appear infrequent, they are notable in pathogenic lineages that dispersed.

Importance: An understanding of the processes that contribute to emergence of pathogens from environmental reservoirs is critical as changing climate precipitates pathogen evolution and population expansion. Phylogeographic analysis of *Vibrio parahaemolyticus* hosts combined with analysis of their *Inoviridae* phage resolved ambiguities of diversification dynamics which preceded successful Atlantic invasion by the epidemiologically predominant ST36 lineage. It has been established experimentally that filamentous phage can limit host recombination, but here we show that phage loss is linked to rapid bacterial host diversification during epidemic spread in natural ecosystems alluding to a potential role for ubiquitous inoviruses in the adaptability of

pathogens. This work paves the way for functional analyses to define the contribution of inoviruses in the evolutionary dynamics of environmentally transmitted pathogens.

Introduction

Cases of *V. parahaemolyticus*, the leading cause of bacterial seafood-borne gastroenteritis world-wide, recently increased along the United States (US) Atlantic coast (1-3). This rise in illnesses was tied to warming ocean temperatures, a growing aquaculture industry, and the emergence of endemic pathogens (3-6). But incursion of a Pacific endemic lineage called sequence type (ST) 36 into the Atlantic is arguably the most important driver of this epidemiological shift (3, 5, 7). ST36 began causing infections in 1979 (8) and was originally limited to the Pacific Northwest (PNW) of North America (9). Although sporadic infections were occasionally reported outside the PNW, including New York in 2006, local sources were rarely implicated (7, 9, 10). This changed when cases by ST36 were traced to the Atlantic including Long Island Sound (LIS) in 2012 (5, 7). Subsequently, ST36 caused a 13-state outbreak traced to multiple Northeast US harvest areas in 2013 (5, 7). Unlike previous non-endemic strains causing outbreaks from Atlantic sources, including the pandemic ST3 from Long Island Sound (LIS) (11), and ST8 from Chesapeake Bay (12), ST36 continues to cause sporadic disease from Northeastern US locations. A better understanding of how ST36 established new populations in the northwest Atlantic could aid in ongoing management to limit illness. Furthermore, the expansion of ST36 provides a unique opportunity to broaden our understanding of the population and environmental context by which pandemic strains arise and spread.

Spatiotemporal analyses of the epidemiology of ST36 identified that a new population arose through recombination and replaced the original PNW population by 2000 (9). The epidemiological patterns implied potential reciprocal transfer and mixing of populations between both US coasts, though that inference was largely informed by a few individual cases where

infection occurred on the opposite coast from where illnesses were reported and further complicated by modeling that presumed pathogen diffusion through human populations over land (9), though *V. parahaemolyticus* is not generally contagious but rather is environmentally vectored most often by raw seafood. This highlights a continuing challenge where the lack of environmental traceback in clinical reporting that uncouples strains from their source bodies of water obscures environmental distribution (5, 9). As with other environmentally transmitted human pathogens, this leaves a critical gap in understanding of the ecological context of pathogen evolution and population expansion.

Bacteriophages, the viruses that infect bacteria and archaea, can shape and quell environmentally-transmitted human pathogen populations (13-19). Lytic phage alter the dynamics of bacterial competition though selective predation, sometimes of the most numerous bacteria, thereby maintaining population diversity (20-22). Phages also impact diversity through transfer of novel DNA including toxin-encoding genes and lysogenic conversion (23-28). Chromosomally integrated prophages are double-edged swords in that they can benefit their host by excluding superinfections by related phage and by attacking their hosts' competitors, but they can also harm their hosts by diverting resources for virion production or by host killing upon lytic induction (25-27, 29, 30). Lytic phages are abundant in habitats of *V. parahaemolyticus* (31), and also under consideration for detection and biological control (32-37). Though the role of phage in bacterial competition among *Vibrio cholerae* is well appreciated (14, 38), with the exception of a few descriptive reports (39-44), the contribution of prophage to *V. parahaemolyticus* population structure is undetermined.

Previous epidemiological studies have produced valuable insight into the growing problem of the ST36 lineage's increasing clinical prevalence and potential to spread globally (9, 45), but the adaptations that prompted lineage replacement and dispersal are not evident nor are the ecological factors that spurred diversification. In this study, we examined the

phylogenomics of the Atlantic ST36 incursion--mediated by interactions with filamentous phage--using a collection of genomes capturing a broad geographic and temporal distribution of the population. By curating strains with environmental traceback, we described the phylogeography of ST36 to include multiple strains that first diversified in the Pacific, subsequently translocated to the Atlantic, and established spatially distinct, recurring and non-mixing subpopulations therein. This study also revealed that changes in *Inoviridae* accompanied diversification and lineage replacement in the PNW and preceded clonal expansion by distinct founders of the two northernmost and persisting Atlantic ST36 populations that continue to cause sporadic disease.

Results

Multiple distinct ST36 lineages clonally expanded to form geographically stable subpopulations in the Atlantic

To document the evolution of ST36 as it translocated from the PNW to multiple new locations in the Atlantic, we curated a large collection of genomes and mapped the environmental origin of isolates as determined by trace-back onto their constructed whole genome phylogenies. Geographic locations include two in the Pacific (PNW and California [CA]), and five in the Atlantic (Galicia Spain, the Gulf of Maine [GOM], South of Cape Cod, MA [SCC], LIS and the Mid-Atlantic coast [MAC] [Table S1]). The whole genome phylogenies exemplify how multiple clades emerged from the old-PNW population (Fig. 1) representing diversification and replacement leading to a new-PNW population (9). A single relic of the presumed extinct old-PNW lineage (2013 isolate CTVP44C from LIS) and isolates from seven lineages arising from the modern ST36 diversification trace to environmental locations outside the PNW (Fig. 1, translocation clades I-VII). Notably, most of these translocation clades either contain a PNW-traced isolate or are most closely related to clades dominated by PNW isolates

suggesting these different lineages likely arose in the PNW prior to translocation rather than after arrival in their current locations in the Atlantic.

Assignment of isolates to environmental sources revealed striking geographic structure of the persisting Atlantic subpopulations in GOM, SCC, and LIS which are relatively close to one another indicating that each of these populations was founded by different individuals (Fig. 1). All but one isolate traced to the GOM are clonal (clade II) and share recent ancestry with isolates from British Columbia Canada and WA (9, 46-48), and an early 2006 New York (LIS) clinical isolate VpG-1 (9)(Fig. 1, Table S1). Most isolates traced to areas in the Atlantic south of Cape Cod, MA (all locations south of the GOM) and those from CA share recent common ancestry with PNW isolates first reported from British Columbia Canada in 2005, and later traced to WA in 2011-2012 (Fig. 1 and Table S1). All but one SCC-traced isolate (clade III) are clonal and only two members of this clade were traced to other locations in the Atlantic. All but six LIS-traced isolates are clonal (76%) (clade VII), though two distinct LIS isolates, both from 2018, group with a mixed-location clade (clade V) whose earliest isolates were traced to the PNW in 2011, and from oysters consumed in CA in 2015 (MAVP-86). Phylogenies built with core non-recombining variation mostly agreed with whole genome phylogenies, though a change in clonal assignment of two strains (CTVP25C and MAVP-48) alluded to the possibility that horizontally acquired variation shared in common with the GOM and SCC clades respectively may have obscured these two strain's distinct heritage (Fig. S1 and S2).

Given the strong linkage of clonal subpopulations to locations (Fig. 1), we examined the population genetic structure of isolates with known traceback to identify patterns of coancestry and admixture among ST36 strains using a technique that predicts and allows visualization of the ancestry and mixed ancestry of individuals (49). The conservative use of core genome variation and eight populations (which best explains the genomic data, see Fig. S3A & B) identified two old-PNW populations that were replaced by a single modern day PNW population,

where based on whole genome content, most isolates form two distinct lineages: one related to GOM isolates and the other to SCC/LIS/MAC isolates (Fig. 1 & Fig. 2). Six PNW isolates were exceptions, in that they belong to the same population and form clades with isolates that translocated to Spain, GOM, and LIS. Although no PNW isolates belong to the population that translocated to CA, two PNW isolates exhibit substantial coancestry and a shared history of admixture with the CA isolates providing clues to the origin of this lineage. The clonal clades that are in close geographic proximity to one another from SCC and LIS also belong to the same population, whereas the isolates from MAC form their own population. Atlantic-derived isolates from locations south of the GOM exhibit little coancestry with other populations, whereas GOM clonal clade isolates display mixed coancestry with the prevalent modern PNW population and the remaining northwest Atlantic populations (Fig. 2). In addition to the two PNW isolates that group with the GOM population, three additional isolates that are members new-PNW population and trace to the PNW also exhibited the same pattern of mixed heritage though with proportions that are dominated by the PNW population.

Analysis that incorporated horizontally acquired novel content in the displayed admixture patterns separated the two Atlantic populations on opposite sides of Cape Cod, MA (GOM and SCC) from the LIS population to the south and identified that the CA, MAC and LIS clades are one population. This also relocated a single GOM clade isolate (MAVP-20) into the CA/MAC/LIS population (Fig. S4). This suggests that even though all translocated lineages are related, distinct horizontally-acquired variation in the SCC and GOM lineages (but absent in MAVP-20) may define their uniqueness. In contrast to the GOM and SCC populations that exhibit relatively little admixture, the modern PNW ST36 isolates exhibit more complex admixture as do many of the LIS population (Fig. S4). However, a few individuals within the GOM and SCC population display admixture with each other. Thus, population structures inferred from horizontal variation better aligned many strains to their modern geographic locations even as horizontal content

overshadowed some interesting ancestral patterns, specifically, the shared mixed heritage of the GOM clade with modern PNW strains (Fig. 2).

Inovirus loss and reacquisition accompanied ST36 population replacement and expansion

One explanation for the differences in population structure when comparing core to whole genomes is that differential and potentially adaptive horizontal gains or losses preceded clonal expansion. To characterize such features of the Atlantic lineages, we compared the genomes of a member of each clade, (MAVP-1 from MAC, MAVP-23 from LIS, MAVP-36 from SCC, and MAVP-26 from the GOM), to a PNW strain of the new lineage (12310) and two reference genomes from the old-PNW clade (10290 and 10329). This revealed MAVP-26 and MAVP-36 each contain a unique, but related, prophage integrated into the *dif* site of chromosome I. These phages are similar to the filamentous phage f237 present in pandemic strains (39, 41) and a phage from strains 10290 and 10329 that is absent in strain 12310 and MAVP-1 (Fig. 3). The ST36 phages are classified in the family *Inoviridae* (hereafter inoviruses) and we subsequently assigned unique names, including “vB” for virus infecting Bacteria, and “Vipa” in reference to the host *Vibrio parahaemolyticus* and isolate name hereafter Vipa26 in MAVP-26, Vipa36 in MAVP-36, and Vipa10290 in strain 10290 (50).

The Inoviruses have a conserved central core (ORF 1-7) and the conserved ORF9 and ORF10 (Fig. 3, gray), flanking two variable regions. High divergence in core nucleotides represented up to 46% of the variant sites in the bacterial host’s whole genome which likely explains observed differences between whole and core genome phylogenies (Fig. 1 and Fig. S2) and population structures (Fig. 2 and Fig. S4). As divergent as these phages are, this variation may not impair phage function: the 142 variant nucleotides in Vipa26 generated only four non-synonymous substitutions, whereas the 201 variant nucleotides in Vipa36 produced only eight non-synonymous mutations when compared with Vipa10290. ORF1-7 are under purifying selection (codon-based Z test =9.425, $p < 0.001$) implying these are essential genes.

This further suggests these are productive infections. Potential functions were assigned to the core phage genes (Table 1), whereas functions of the variable regions were not discernable.

With only one exception (2005 strain A5Z905), all old-PNW clade members, including the last isolate CTVP44C which was traced to LIS in 2013 contained Vipa10290 or closely related phage, (Fig. 1, square). ST39, who is a member of the same clonal complex as ST36 (CC36) also contains Vipa10290 whereas other CC36 members (ST59 and ST21), share a distinct inovirus (Fig. S5, Tables 2, S2, & S3). (Table 2). Despite this long history of phage carriage among the endemic population, no modern-day and definitively traced PNW ST36 isolates harbor inoviruses (Fig. 1 & Table S1) and the timing of phage loss coincides with remarkable population diversification (Fig. 1). However, like the GOM and SCC lineages, a total of nine other modern ST36 isolates collected between 2007-2016 some that dispersed to new locations and others without definitive trace, acquired eight unique inoviruses, for a total of eleven distinct inoviruses among ST36 isolates (Fig. 1, Table 2).

The identities and distribution of community-acquired phage offer clues of residency and dispersal

For the Atlantic populations of ST36, inovirus content was remarkably diagnostic of ST36 lineage and geography. Vipa26 was absent from only one member (MAVP-20) of the ST36 GOM clonal clade and present in only one distinctive isolate from a different location CTVP25C representing a either basal or independent acquisition (Fig. 1 circles). Vipa36 (Fig. 1 triangles) is in all members of the SCC clonal clade and only two isolates from this clonal clade were traced to nearby locations. Because these specific inoviruses, like f237, are not widespread among global *V. parahaemolyticus* populations and have been maintained by these lineages, they are diagnostic of these ST36 subpopulations. In contrast, and like the modern PNW traced-ST36 isolates, no isolates of the LIS or MAC clonal clades harbor inoviruses.

The association of distinguishing inoviruses with two localized clades of the Atlantic, and their absence in others, suggests that the founding subpopulation progenitors acquired phage upon Atlantic invasion but did not do so in all locations. Environmental surveys revealed that although 22% of all Northeast environmental isolates of *V. parahaemolyticus* harbor inoviruses, this proportion varied by location (50). Proportions were low in LIS (10%), higher in GOM locations in MA (30%), and peaked in SCC (42%). Though no environmental isolates other than ST36 harbored Vipa36, four non-ST36 isolates from the GOM, including two collected in 2008 from NH contained Vipa26, and two others contained remnants of Vipa26 (Table S2, S3, Fig. S6). These findings suggest Vipa26 is a resident phage that was acquired locally, a premise that is supported by its absence from VpG-1 (Fig. 1, Table S1,S2 & S3).

Because comparative phage phylogeny could elucidate the history of the ST36 members as they moved through locations and interacted with other populations, we expanded our search for related inoviruses among the publicly available genomes of *V. parahaemolyticus.* This survey added to the remarkable diversity among *Inoviridae* based on whole (Fig. 4) and central core (ORF1-7, Fig. S6) phylogenies. Vipa26, which has only been found in strains from the GOM, is most closely related to phage from environmental isolates from SCC, whereas Vipa36 is most closely related to phages from isolates originating in Asia (Fig. 4). Two ST36 isolates harbor two different inoviruses maintained by members of the PNW co-resident and clinically prevalent ST43 lineage that is not related to ST36 (Table 2). Though most members of the two ST36 lineages reported in Peru also lack inoviruses, including the single 2011 environmental isolate that shares ancestry with the Spain lineage (45), the second clinical member of this lineage harbors an inovirus that is most closely related to the ancestral PNW Vipa10290, whereas one other clinical isolate (45) harbors the pandemic clonal complex inovirus f237.

Given the diversity of inoviruses (eleven distinct, Table 2) and evidence of their repeated horizontal acquisition and replacement by ST36, we assessed to what extent inoviruses in other

lineages of *V. parahaemolyticus* reflects vertical inheritance or horizontal transmission. This survey, which is predominated by clinical isolates, revealed inoviruses are common (Fig. 5, Fig. S5, Fig. S6, and Table S3) with 46% of genomes containing one or more inovirus, although the distribution is uneven. For instance, many lineages generally correlate with an endemic and persistent inovirus (Fig. 5 and Fig. S6). A few lineages harbored two evolutionary distinct inoviruses concurrently suggesting protection from superinfection, a common attribute of both temperate phage and inoviruses (51, 52), is not absolute (Fig. S5). One regional phage lineage other than Vipa26 is highly successful and present in multiple, unrelated STs, some from LIS (Vipa4291; check mark, Fig. S6). Even though most STs generally maintain and propagate their prophage vertically over multiple years, some isolates harbored diverse inoviruses, and apparently underwent prophage replacement, like ST36 (Table 3, Fig. S6). Notable among these are emergent pathogenic lineages (e.g. ST43 the prevalent Atlantic endemic lineage that has caused illnesses along the entire North American Atlantic coast (ST631)(4) the pandemic complex (ST3) and a diverse and far-spreading Asian lineage that is now an Atlantic resident (ST8).

Discussion

Here we demonstrate that shuffling of inovirus prophage accompanied diversification, competitive replacement, dispersal, and Atlantic invasion of ST36. This suggests that ubiquitous persistently-infecting inoviruses could be key players in structuring populations that lead to emergence of new lineages with the potential to spread more broadly. The misleading nature of disease reporting that does not align illnesses with source location of exposure and gaps in population coverage due to the general decrease in available ST36 genomes from the US PNW alluding to a decrease in clinical prevalence have been obstacles to deciphering the ecological context of ST36 evolution and spread. We addressed these limitations by including publicly

available genomes from isolates reported in Canada and other US states and by augmenting publicly available genomes with a large collection of ST36 clinical isolates from four Northeast US states. We carefully curated trace back for geographic assignment to enhance the resolution ST36 population distribution and advance our understanding of ST36 ecological expansion into the Atlantic Ocean.

Our analysis indicates at least five unique ST36 progenitors diversified in and then translocated from the Pacific into the Atlantic to found resident subpopulations causing infections in two or more years (Fig. 1, S2, 2, and S4). The earliest sequenced ST36 isolate traced to the Atlantic (VpG-1) from New York (LIS) and several PNW strains group with the northernmost GOM clade and share a similar pattern of mixed ancestry suggesting that the GOM progenitor may have been introduced prior to 2006 with little effect on epidemiology until after the anomalously mild winter of 2011-2012 (1, 2). Isolates traced to SCC, LIS and MAC, are closely related to each other and both their phylogeography and the presence of multiple Pacific traced isolates (both PNW and CA) within or adjacent to their clades is most simply explained by introduction of three related but already divergent ancestors (as well as a relic of the old population) from the Pacific: one as the progenitor of the persisting SCC population, a second as the common progenitor of both the predominant LIS and MAC populations and a third progenitor of two LIS isolates (clade V) from 2018 (Fig. 1, 2 & S2). That this newer lineage (clade V) that shares ancestry with 2012 isolates from CA, and which also caused infections traced to the PNW further implicates the PNW as the source population, rather than multiple re-introductions of Atlantic lineages into more than one location of the Pacific. And though imperfect trace-back or mislabeled genomes could explain some of the anomalous phylogeographic patterns, nearly every Pacific-traced close relative of Atlantic residential populations were isolated, reported, and sequenced independently and prior to Atlantic isolation (Table S1). Though our interpretations of population evolution and dispersal contrast with

inferences of previous epidemiological analyses (9), this is mostly explained by differences in geographic assignment, in our case based on environmental location of traceback with the goal of understanding relationships between environment populations that are potential reservoirs of emergent pathogenic lineages (See details of Table S1). Even with this different emphasis, traceback confirmation from CA (Table S1) disease reporting in Pacific coastal Mexico (53) and environmental isolation from Peru (9, 45) corroborate that pre-diversified ST36 lineages spread to both locations from the PNW. And though missing intermediate populations could suggest unknown reservoirs exist, it is even more likely that sampling bias (i.e. reported disease), a decrease in availability of genomes from the PNW, decreased clinical prevalence of ST36 in the PNW region, as well as a lack of environmental surveillance have compounded challenges with deciphering the evolution and spread of this pathogen. Based on mounting evidence presented herein that most if not all lineages arose in the PNW, this warrants further examination of the ecological context in the PNW and emphasizes the importance for resolving population dynamics in environmental reservoirs to help manage ongoing spread.

The stable and non-mixing spatial distribution of Atlantic ST36 lineages indicates that environmental fitness that promotes persistence in these local ecosystems, rather than mobility in of itself, may be the key to the success of ST36 in the Atlantic. Members of the GOM, SCC, and LIS geographic and clonal lineages were repeatedly traced to their same location, but infrequently traced to each other's locations nor have any of these lineages established in new aquaculture areas. Population perturbation by an exceptionally mild winter in 2011-2012 likely enabled establishment of some of these non-natives (2). Illnesses from Pacific sources and long-distance movement of *V. parahaemolyticus* have been tied to climate, such as El Niño events (9, 54-57), but in BC Canada where many of the diversified ST36 isolates were first reported and a decline in clinical prevalence of ST36 has been suggested, rising infection rates were not fully explained by sea surface temperature models and the spike in infections in 2015

was attributed to new pathogenic strains (58, 59). Furthermore, the MAC experienced a warming trend in 2012 that was similar to the Northeast US (9), and yet this location has not experienced continuing problems with ST36, suggesting those populations have not persisted or their levels remain below a threshold necessary for disease (1, 10). Combined, we believe underexplored population interactions and dynamics are at play. Notably, even in locations in the Atlantic with recurrent illness, ST36 is extremely rare, with only one environmental isolate to date (MAVP-123E from SCC, Table S1) further illustrating how low-level populations that persist can change epidemiology. Identification of unique ecological and population contexts of the MAC compared to others in the north Atlantic could provide insight into the unique ecological context that influences disease risk.

Sequential phage replacement is one of the most striking and common features of ST36 lineage succession in the PNW and population expansion into the Atlantic. All but one member of the old PNW lineage harbor Vipa10290 (Fig. 1 and Table S1), whereas strains of the new ST36 lineages isolated in the PNW after 2002 lack inoviruses. The presence of Vipa10290 in other CC36 members could indicate a basally acquired phage was lost during lineage diversification, though inovirus excision is infrequently reported (60, 61). The lack of phage-harboring PNW ST36 population members indicates they ultimately were at a disadvantage in the context of a changing population even though other residents concurrently maintained inoviruses (e.g. ST43, Table 2) and potentially transmitted these to a few ST36 members. The subsequent acquisition of eleven different inoviruses by diverging ST36 lineages (Table 2) indicates ST36 remained susceptible to at least some inoviruses, though the persisting invirus-free state of modern ST36 lineages traced to the PNW population suggests a potential that resistance may have evolved during their rapid recombination mediated diversification, as has been described in other sympatric Vibrio populations (62). In contrast, though the progenitors of two clinically prevailing Atlantic ST36 lineages acquired inoviruses prior to clonal expansion,

while ST36 populations in the LIS where phage may be less environmentally abundant did not, indicates phage acquisition was not requisite for success in the Atlantic, perhaps reflecting differences in phage predation pressure and other population dynamics at these locations. Though resident ST36 from SCC and GOM have continued to cause sporadic disease, infections from LIS have precipitously declined in recent years, in part due to management.

Even though inovirus prophage confer obvious risks and fitness costs most often due to the persistent production of virions (60), their maintenance for years (Fig. 5), and their reacquisition and replacement in several successful lineages (Table 2 & 4, Fig. 5) emphasizes that inoviruses may confer key advantages in some ecological contexts. Shed inovirus particles contribute to biofilm matrix (63) and prophage can regulate biofilm production which not only confers enhanced virulence through immune evasion and antimicrobial resistance, but also could contribute to survival during long distance movement through attachment to marine particles and plankton (52, 54, 64-66). Variable content in inoviruses is also linked to anti-predation (17, 67) and virulence (22, 27, 39). Shed virions can also mislead human immune responses thereby decreasing the ability of hosts to clear bacterial infections, which could further enhance virulence (68). It is intriguing to consider that the succession in the PNW and loss of inoviruses from the ST36 population may have accompanied a decline in clinical prevalence of ST36.

Prophage also can also modify their hosts in ways that alter their capacity for horizontal gene transfer (HGT). Prophage protect against superinfection by competing phages, including unrelated lytic phages (27, 60, 69) thereby blocking transduction and also modulate, and concurrently block conjugation through the same modifications of Type IV secretion, which has been shown both in phage M13 (70) and for Pf4 of *Pseudomonas aeruginosa* (51, 52). The concept that prophage conferred immunity contributed to ST36 population succession is attractive considering DNA release by predation of non-lysogen susceptible hosts can promote

transformation through natural competence (71-73), and phage transduction and conjugation mediated by Type IV secretion also facilitate horizontal gene transfer (HGT) all of which could drive rapid diversification such as observed in the native ST36 population (Fig. 1 and (9)). In keeping with this premise, the newly diversified ST36 lineages would be vulnerable to lytic infections, phage transduction, and conjugal transfer alike and would likely not have evolved mechanisms to resist local phage populations as they invaded the Atlantic (62).. Aquaculture areas of the MAC are teeming with lytic phage that attack pathogenic *V. parahaemolyticus* (31), suggesting these could have impeded the long-term establishment of ST36. Some highly divergent ST36 individuals lacking phage evolved by multiple recombination events including an isolate of the MAC clade lacking inoviruses (VP30). Accumulation of non-inovirus prophage was also a notable feature of the early diverging Spain lineage (clade I) indicating its susceptibility to infection (9). In further support of this premise, the inovirus-harboring SCC and the GOM clades remain stable and exhibit strikingly little diversification or local population admixture whereas the single GOM clade member lacking Vipa26 (MAVP-20) and many members of the LIS population exhibit more complex admixture (Fig. S4). The old-PNW population relic strain CTVP44C harboring Vipa10290 and isolated in LIS in 2013 is virtually unchanged from the last known members of this lineage from the PNW that still harbored this inovirus, isolated more than a decade earlier in 2002 (Fig. 1, Table S1). In contrast a 2005 isolate of this same lineage that lacked Vipa10290, A5Z905, is among the most highly divergent ST36 strains (Fig. 1, 2, S2). Though these data only establish correlation of phage absence and diversification, it alludes to the possibility that carriage of inovirus prophage that provide some level of resistance to superinfection and block conjugation could thwart HGT, and by contrast, loss of inoviruses could remove barriers to recombination and promote rapid diversification. The selective conditions that favor inovirus acquisition and maintenance or loss would thereby influence some

mechanism that could promote rapid evolution, potentially of lineages with more or even less pathogenic potential.

These analyses reveal an intriguing connection of inovirus prophage to pathogen evolution dynamics, alluding to the possibility that inoviruses could be gatekeepers of HGT in evolving populations in natural ecosystems driving dramatic changes in human disease epidemiology. With broader population analysis, this study could lay the foundation for a new evolutionary paradigm where inovirus prophage-mediated immunity is a major governing force beyond the appreciated roles of phage in population diversity and pathogen evolution through selective predation (20-22) and phage conversion (23-27). The apparent ability of ST36 and some other lineages (Table 2 & 3) to transition between phage infection states or replace their phage could generate flexibility in the balancing of fitness tradeoffs under different selective regimes that are not yet understood (29, 62). The high prevalence of inoviruses not only among the global population of *V. parahaemolyticus* (Fig. 5) but universally in other bacteria (27) with little clarity of how inoviruses impact the ecology of their hosts calls for more mechanistic analyses which could reveal varied roles for inoviruses in promoting evolution and competitive dominance of emerging pathogens.

Materials and Methods

Vibrio parahaemolyticus *isolates and sequencing*

ST36 genomes from our own collection and acquired from NCBI (https://www.ncbi.nlm.nih.gov/genome/681; August 2019) are in Table 1, whereas additional genomes included in whole genome phylogenies are in Table S1 and those identified as harboring inoviruses are in Table S2. Clinical strains were acquired from Massachusetts, Maine, New Hampshire, and Connecticut State public health laboratories and spanned infections occurring between 2010-2018 (2). Trace-back data was provided by public health

entities. Only isolates with a single geographic source trace were assigned to a location and all others were considered unknown. Environmental strains were isolated from New Hampshire, Connecticut, Maine and Massachusetts oysters, water, sediment and plankton between 2007-2015 and grown routinely in LBS medium (1, 2). Clinical and environmental strains were identified as *V. parahaemolyticus* by detection of the thermolabile hemolysin gene (*tlh*) (74, 75) and identity confirmed by genome sequencing.

Genomic DNA was extracted from overnight LBS broth cultures using the Wizard Genomic DNA purification Kit (Promega, Madison WI USA) or by organic extraction (2). Libraries for DNA sequencing were prepared using a high-throughput Nextera DNA preparation protocol (76). Genomic DNA was sequenced using an Illumina – HiSeq2500 device at the Hubbard Center for Genome Studies at the University of New Hampshire, using a 150bp paired-end library and reads de-multiplexing and quality filtered prior to analysis. The *de novo* genome assembly were performed using the A5 pipeline (77), and annotations assigned with Prokka1.9 using "*Vibrio"* for the reference database (78). The sequence types were determined using the SRST2 pipeline (79) or using assemblies (80) referencing https://pubmlst.org/vparahaemolyticus/(81).

Relationships of V. parahaemolyticus *isolates*.

The relationships of strains was determined using reference-free, whole genome alignments using kSNP 3.1 (82) and maximum likelihood phylogenetic trees built by RaxML (83) and visualized with iTOL(84). 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 (85, 86).

To reduce the confounding bias of variation generated by frequent recombination events, homoplasic regions were identified using Gubbins 2.3.4 (87) and removed, and the

relationships of strains subsequently determined with alignments generated with snippy 4.3.6 (88). Phylogenies were inferred using RaxML with 1000 bootstraps, and a random seed set to 12345, and visualized using iTOL (84).

Estimation of the population structure of ST36 was performed using both whole and core genome genetic matrixes in the R statistical package LEA (49). To determine the number of genetic clusters best explained by the distribution of genomic variation, we explored ancestral population (*K*) numbers of between 1-20 and used the entropy criterion to evaluate the quality of fit of the statistical model to the data. The probable number of populations was selected based on minimal entropy, and where entropy was comparably low for two different ancestral population numbers, each and all flanking population structures were visualized. Once strains were assigned to populations, the least-squares estimates of ancestry proportions were plotted and organized in sequential order by population, within population by geographic location, and within location by year of isolation using ggplot2 (89).

Identification and classification of inoviruses

Strains MAVP-26, MAVP-36, 12310 (a modern PNW isolate) were compared to 10290 (a pre-diversified ancestral PNW isolate) (5), as a reference using Breseq (90). The unique prophage in MAVP-26, MAVP-36 and 10290 were subsequently classified belonging to the family *Inoviridae* proposed subfamily Protoinoviridae (27) and named as *Vibrio* phage according to current recommendations, following confirmation that the chosen names were unique (91, 92). 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 designation representing the strain identity (Table 2). All sequenced *V. parahaemolyticus* strains were evaluated for f237-like phage content (ORF1 – ORF7) and phage type using BLASTn(93) (https://www.ncbi.nlm.nih.gov/genome/genomes/691; accessed Aug. 2019). To identify whether phage were unique, maximum parsimony phylogenies were constructed on SNPs generated by

kSNP with 1000 bootstraps and a random seed 12345. For all unique *Inoviridae* in ST36 genomes, the complete prophage assemblies were constructed and named from the highest quality assembly (Table 2) and other prevalent inoviruses were also identified and named (Table S4).

To determine if phage evolved under selection, a Nei-Gojobori codon-based Z test (94) was performed in MEGA 6 (95) on the seven central core genes. Protein sequences for the phages Vf33 (NC_005948.1) (96, 97), CTXΦ (MF155889.1)(27), and the type strain for *Inoviridae* M13, which infects *Escherichia coli* (GCF_000845205.1) were compared using BLASTp (98) or by reference to its relative location in the phage genome regardless of sequence identity.

Phage relatedness with abbreviated sets of whole phage genomes was examined using the Genome-BLAST Distance Phylogeny (GBDP) method (99) to conduct pairwise comparisons of the nucleotide sequences under settings recommended for prokaryotic viruses (100). Branch support was inferred from 100 pseudo-bootstrap replicates. Trees were rooted at the midpoint (28) and visualized using iTOL(84).

Phage typing by PCR

ST36 phage and f237 alignments produced in MEGA 6 (95) allowed identification of sequences shared by f237-like phages for assay development. Oligonucleotide primers were designed for multiplex amplification with the species-specific *tlh* primers (74, 75) as an amplification control. Inovirus presence was identified with primers ST36Phage F2 (5'-AGCAACGAAAACGCCTGT-3') annealing within ORF3 and ST36Phage R2 (5'-ACCGTATCACCAATGGACTGT-3') annealing within ORF5, to produce a ~1000bp amplicon. Subsequent identification of Vipa26 and Vipa36 was conducted using PhHypDF3 (5'-AAGTGCTACATGAATGAAAGTGCT-3') with PhORF9R1 (5'-

TCAATGAAGTATCACGAAATGACTA -3') which produce different sized amplicons (1440bp or 854bp respectively).

PCR amplification was performed with 30 cycles of 10µL reactions with 1x AccuStart II Supermix (Quantabio, Beverly MA USA), 1µL lysate (101), and 0.2µM of each primer annealed at 55°C with a 1 minute amplification at 72°C for f237-like amplicons and 1.5 minutes for discrimination between Vipa26 and Vipa36. The primers were validated on a variety of strains with known phage content and prophage identity only assigned following whole genome sequencing and analysis.

Data Availability

GenBank accession numbers for genomes are listed in Table S1 & S2 and for Inoviridae are listed in Table 2.

Acknowledgments: Partial funding was provided by the National Sea Grant College Program and projects R/SFA-4, R-HCE-3, and R/SFA-13, and by the New Hampshire Agricultural Experimentation Station. This work (#####) was supported by the USDA National Institute of Food and Agriculture Hatch projects NH00574, NH00609 [accession number 233555], and NH00625 [accession number 1004199]) and NH00658 [accession number 1013479]. We thank Narjol Gonzalez-Escalona for helpful insight and feedback; Kanwit Kohl, Stephen Combes and Heather Grieser from the Maine Center for Disease Control and Prevention for strains and information; Tracy Stiles, Eric Hickey, Coleen Murphy, Michael Moore, and Jana Ferguson from the MA Department of Public Health for helpful insight, strains and epidemiological information; Jeff Kennedy and Christian Pepitas from the MA division of Marine Fisheries for insight and guidance; Christopher Nash and Robert Atwood from the New Hampshire Department of Environmental Services, Colleen Smith, New Hampshire Department of Health and Human

Services for information and discussion. We also thank Zoe Pavlik for curation of genome accessions.

Role of authors: JM, AM, RF, FX, CS, KD, JH, and ZP collected samples, performed and interpreted analyses, curated databases, and constructed visuals. CW, JM and AM drafted the manuscript. SJ, VC, and CW designed and directed the studies. All coauthors reviewed and edited the manuscript for accuracy and completeness.

Figure 1. Maximum-likelihood phylogeny of ST36 *Vibrio parahaemolyticus*. ML-phylogenies were built with whole genome variation of select quality ST36 strains. Isolates are colored to correspond to geographic origin trace-back source, where available (Table S1) and where no color indicates unknown or ambiguous origin. Strains acquired from four Northeast US States acquired as part of this study are named uniquely by a combination of state (Maine [ME], New Hampshire [NH], Massachusetts [MA], and Connecticut [CT]) of reporting and sequential numbers. Symbols next to strains indicate unique inovirus content. The ancestral PNW ST36 population is identified by a black bar, whereas seven different lineages associated with translocation events (I – VII) are identified by greyscale bars.

Figure 2. Population structure of ST36 strains. Coancestry estimates were inferred from core non-recombining genomes of ST36 strains by LEA (49). Colored bars represent proportion of genetic variation derived from 8 ancestral populations (see Fig S3A). Geographic locations are marked by colors below strain name, followed by phage content (symbol) and year of isolation.

Figure 3. Structure of *V. parahaemolyticus* Inoviruses. ORFs with the same color are homologous between the different phages with the exception of the blue hypothetical proteins. The core genome (highlighted in gray) is highly conserved, whereas accessory regions vary, with some ORFs conserved in two or more of the phages.

Table 1. Assignment of potential functions of core inovirus genes [1].

ORF	Vf33 (E-value/%Identity)	CTXΦ (E-value/%Identity)	M13 (E-value/%Identity)	Predicted Functions[a]
ORF1	Vpf402 (0.0/99%)	RstA ($7x10^{-66}$/36%)	gII (size and location)	Replication initiation protein
ORF2 Integrase	Vpf117 ($1x10^{-85}$/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
ORF6	Vpf104 (0.2/19%)	Ace ($5x10^{-10}$/18.3%)	gVI (4.1/8.7%)	Minor coat protein, termination of assembly
ORF7	Vpf380 ($1x10^{-4}$/13.9%)	Zot ($2x10^{-27}$/17.8%)	gI ($5x10^{-4}$/8.5%)	Assembly protein, maturation
ORF9	Vpf122 ($4x10^{-66}$/68%)	RstR (0.17/5.7%)	-	Transcriptional repressor, regulator
ORF10	Vpf152 (1^{-44}/88-6%)	-	-	Function unknown

[1]ORF number corresponds to numbers designated for f237, and comparative analyses between homologous ORFs completed with Vipa26 as a reference.

[a] Functions summarized from Mai-Prochnow et al., 2015 (36) and Chang et al., 1998 (37)

Table 2. Summary of *Inoviridae* prophage in ST36 isolates

Strain	Phage name	NCBI Accession (Genome coordinates)	Additional ST36 strains with phage	Other STs with related Phage
10290	vB_Vipa10290	AVOH01000008.1 (1370845..1378511)	Many	38, 39

MAVP-26	vB_Vipa26	MT188662	Many	1574, NF
MAVP-36	vB_Vipa36	MT188663	Many	None
CDC_K5323G	vB_Vipa5323	MIUF01000019.1 (188973..190181) MIUF01000072.1 MIUF01000065.1 MIUF01000002.1 (4141..6171)	None	752, 1132, NF
CDC_K5308	vB_Vipa5308	MIUE01000114.1 (7376..8729) MIUE01000031.1 (5201..8319)	None	43
10-4255	vB_Vipa71[1]	MT193890 NIXZ01000007.1 (181573..191867)	None	43
CDC_A8962	vB_Vipa8962	LHRO01000017.1 (70737..72090)	None	327
CDC_K5512	vB_Vipa5512	MIUZ01000070.1 (628..7658)	CDC_K5345G, CDC_K5280	1713
MEVP-10	vB_Vipa10	MT188666	None	None
1.146-15	f237	WSRX01000192.1, WSRX01000239.1, WSRX01000194.1, WSRX01000150.1	None	3, 152, 809, 1473, NF
1.220-16	other	AAXNNA010000036.1	None	None

NF: Sequence type that is not yet assigned; None: no strains harboring phage.

[1]Named for ST43 strain MAVP-71

Figure 4. Phylogenies of representative inoviruses from 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 1. Shape next to isolate name corresponds to phage type (Fig. 1); circle is Vipa26, square is Vipa10290, triangle is Vipa36 and grey square is other (Vipa8296). Isolates in bold are ST36.

Figure 5. Maximum-likelihood tree of *V. parahaemolyticus* isolates. A maximum-likelihood (ML) phylogeny was built on core genome SNPs where 100 bootstraps were mapped onto the best scoring ML tree. Isolates are colored to correspond to geographic region as in Fig. 1, where no color indicates unknown or ambiguous origin. Symbols next to strains indicate unique inovirus content.

Table 3. Strain lineages other than ST36 with diverse phage genotypes.

Sequence type	Number of unique Inovirus prophage in lineage	Sequenced strains without prophage
631	4	Yes
674	4	No
43	5	Yes
12	4	No
3	2	Yes
8	2	Yes
308	2	Yes
120	2	Yes

Literature Cited:

1. Urquhart EA, Jones SH, Jong WY, Schuster BM, Marcinkiewicz AL, Whistler CA, Cooper VS. Environmental Conditions Associated with Elevated *Vibrio parahaemolyticus* Concentrations in Great Bay Estuary, New Hampshire. PloS ONE. 2016;11(5):e0155018.
2. Xu F, Ilyas S, Hall JA, Jones SH, Cooper VS, Whistler CA. Genetic characterization of clinical and environmental *Vibrio parahaemolyticus* from the Northeast USA reveals emerging resident and non-indigenous pathogen lineages. Name: Frontiers in Microbiology. 2015;6:272.
3. Baker-Austin C, Trinanes JA, Taylor NG, Hartnell R, Siitonen A, Martinez-Urtaza J. Emerging *Vibrio* risk at high latitudes in response to ocean warming. Nature Climate Change. 2013;3(1):73-7.
4. Xu F, Gonzalez-Escalona N, Sebra RP, Dress KP, Cooper VS, Jones SH, Whistler CA. Parallel evolution of two clades of a major Atlantic endemic *Vibrio parahaemolyticus* pathogen lineage by independent acquisition of related pathogenicity islands. Applied and Environmental Microbiology. 2017. doi: https://doi.org/10.1101/155234.
5. Martinez-Urtaza J, Baker-Austin C, Jones JL, Newton AE, Gonzalez-Aviles GD, DePaola A. Spread of Pacific Northwest *Vibrio parahaemolyticus* strain. N Engl J Med. 2013;369(16):1573-4.

6. Vezzulli L, Grande C, Reid PC, Hélaouët P, Edwards M, Höfle MG, Brettar I, Colwell RR, Pruzzo C. Climate influence on *Vibrio* and associated human diseases during the past half-century in the coastal North Atlantic. Proceedings of the National Academy of Sciences. 2016:201609157.

7. Newton AE, Garrett N, Stroika SG, Halpin JL, Turnsek M, Mody RK, Division of Foodborne W, Environmental D. Notes from the field: Increase in *Vibrio parahaemolyticus* infections associated with consumption of Atlantic coast shellfish—2013. MMWR Morbidity Mortality Weekly Report. 2014;63(15):335-6.

8. Abbott SL, Powers C, Kaysner CA, Takeda Y, Ishibashi M, Joseph SW, Janda JM. 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. 1989;27(12):2891-3. Epub 1989/12/01. PubMed PMID: 2592553; PMCID: PMC267157.

9. Martinez-Urtaza J, van Aerle R, Abanto M, Haendiges J, Myers RA, Trinanes J, Baker-Austin C, Gonzalez-Escalona N. Genomic Variation and Evolution of *Vibrio parahaemolyticus* ST36 over the Course of a Transcontinental Epidemic Expansion. MBio. 2017;8(6). Epub 2017/11/16. doi: 10.1128/mBio.01425-17. PubMed PMID: 29138301; PMCID: PMC5686534.

10. Martinez-Urtaza J, Powell A, Jansa J, Rey JL, Montero OP, Campello MG, Lopez MJ, Pousa A, Valles MJ, Trinanes J, Hervio-Heath D, Keay W, Bayley A, Hartnell R, Baker-Austin C. Epidemiological investigation of a foodborne outbreak in Spain associated with U.S. West Coast genotypes of *Vibrio parahaemolyticus*. Springerplus. 2016;5:87. Epub 2016/02/06. doi: 10.1186/s40064-016-1728-1. PubMed PMID: 26848427; PMCID: PMC4729754.

11. CDC. Outbreak of *Vibrio parahaemolyticus* infection associated with eating raw oysters and clams harvested from Long Island Sound--Connecticut, New Jersey, and New York, 1998. MMWR Morb Mortal Wkly Rep. 1999;48(3):48-51.

12. Haendiges J, Jones J, Myers RA, Mitchell CS, Butler E, Toro M, Gonzalez-Escalona N. A Nonautochthonous U.S. Strain of *Vibrio parahaemolyticus* Isolated from Chesapeake Bay Oysters Caused the Outbreak in Maryland in 2010. Appl Environ Microbiol. 2016;82(11):3208-16. Epub 2016/03/20. doi: 10.1128/AEM.00096-16. PubMed PMID: 26994080; PMCID: PMC4959243.

13. Zahid MS, Udden SM, Faruque AS, Calderwood SB, Mekalanos JJ, Faruque SM. Effect of phage on the infectivity of *Vibrio cholerae* and emergence of genetic variants. Infect Immun. 2008;76(11):5266-73. Epub 2008/09/17. doi: 10.1128/IAI.00578-08. PubMed PMID: 18794293; PMCID: PMC2573317.

14. Jensen MA, Faruque SM, Mekalanos JJ, Levin BR. Modeling the role of bacteriophage in the control of cholera outbreaks. Proc Natl Acad Sci U S A. 2006;103(12):4652-7. Epub 2006/03/16. doi: 10.1073/pnas.0600166103. PubMed PMID: 16537404; PMCID: PMC1450226.

15. Faruque SM, Bin Naser I, Fujihara K, Diraphat P, Chowdhury N, Kamruzzaman M, Qadri F, Yamasaki S, Ghosh AN, Mekalanos JJ. Genomic sequence and receptor for the *Vibrio cholerae* phage KSF-1phi: evolutionary divergence among filamentous vibriophages mediating lateral gene transfer. J Bacteriol. 2005;187(12):4095-103. Epub 2005/06/07. doi: 10.1128/JB.187.12.4095-4103.2005. PubMed PMID: 15937172; PMCID: PMC1151723.

16. Nanda AM, Thormann K, Frunzke J. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. J Bacteriol. 2015;197(3):410-9. Epub 2014/11/19. doi: 10.1128/JB.02230-14. PubMed PMID: 25404701; PMCID: PMC4285972.

17. Brussow H. Bacteria between protists and phages: from antipredation strategies to the evolution of pathogenicity. Mol Microbiol. 2007;65(3):583-9. Epub 2007/07/05. doi: 10.1111/j.1365-2958.2007.05826.x. PubMed PMID: 17608793.

18. Molina-Quiroz RC, Camilli A, Silva-Valenzuela CA. Role of Bacteriophages in the Evolution of Pathogenic Vibrios and Lessons for Phage Therapy. Adv Exp Med Biol. 2023;1404:149-73. doi: 10.1007/978-3-031-22997-8_8. PubMed PMID: 36792875.

19. Hampton HG, Watson BNJ, Fineran PC. The arms race between bacteria and their phage foes. Nature. 2020;577(7790):327-36. Epub 20200115. doi: 10.1038/s41586-019-1894-8. PubMed PMID: 31942051.

20. Bohannan BJM, Lenski RE. The Relative Importance of Competition and Predation Varies with Productivity in a Model Community. Am Nat. 2000;156(4):329-40. Epub 2000/10/01. doi: 10.1086/303393. PubMed PMID: 29592139.

21. Joo J, Gunny M, Cases M, Hudson P, Albert R, Harvill E. Bacteriophage-mediated competition in Bordetella bacteria. Proc Biol Sci. 2006;273(1595):1843-8. Epub 2006/06/23. doi: 10.1098/rspb.2006.3512. PubMed PMID: 16790419; PMCID: PMC1634791.

22. Ritchie JM, Wagner PL, Acheson DW, Waldor MK. Comparison of Shiga toxin production by hemolytic-uremic syndrome-associated and bovine-associated Shiga toxin-producing *Escherichia* coli isolates. Appl Environ Microbiol. 2003;69(2):1059-66. Epub 2003/02/07. doi: 10.1128/aem.69.2.1059-1066.2003. PubMed PMID: 12571029; PMCID: PMC143677.

23. Taylor MJ, Thorne CB. Transduction of *Bacillus licheniformis* and *Bacillus subtilis* by each of two phages. J Bacteriol. 1963;86:452-61. Epub 1963/09/01. PubMed PMID: 14066421; PMCID: PMC278456.

24. Romig WR, Brodetsky AM. Isolation and preliminary characterization of bacteriophages for *Bacillus subtilis*. J Bacteriol. 1961;82:135-41. Epub 1961/07/01. PubMed PMID: 13743075; PMCID: PMC279126.

25. Buckling A, Rainey PB. Antagonistic coevolution between a bacterium and a bacteriophage. Proc Biol Sci. 2002;269(1494):931-6. Epub 2002/05/25. doi: 10.1098/rspb.2001.1945. PubMed PMID: 12028776; PMCID: PMC1690980.

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

Microbiol. 2009;7(11):828-36. Epub 2009/10/17. doi: 10.1038/nrmicro2235. PubMed PMID: 19834481.

27. 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. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. Nat Microbiol. 2019. Epub 2019/07/25. doi: 10.1038/s41564-019-0510-x. PubMed PMID: 31332386.

28. Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? Trends Microbiol. 2005;13(6):278-84. doi: 10.1016/j.tim.2005.04.003. PubMed PMID: 15936660.

29. Askora A, Yamada T. Two different evolutionary lines of filamentous phages in *Ralstonia solanacearum*: their effects on bacterial virulence. Front Genet. 2015;6:217. Epub 2015/07/08. doi: 10.3389/fgene.2015.00217. PubMed PMID: 26150828; PMCID: PMC4471427.

30. Yen M, Camilli A. Mechanisms of the evolutionary arms race between *Vibrio cholerae* and Vibriophage clinical isolates. Int Microbiol. 2017;20(3):116-20. doi: 10.2436/20.1501.01.292. PubMed PMID: 29446802; PMCID: PMC7114818.

31. Richards GP, Chintapenta LK, Watson MA, Abbott AG, Ozbay G, Uknalis J, Oyelade AA, Parveen S. Bacteriophages Against Pathogenic Vibrios in Delaware Bay Oysters (*Crassostrea virginica*) During a Period of High Levels of Pathogenic *Vibrio parahaemolyticus*. Food Environ Virol. 2019;11(2):101-12. Epub 2019/02/02. doi: 10.1007/s12560-019-09365-5. PubMed PMID: 30706411.

32. Onarinde BA, Dixon RA. Prospects for Biocontrol of *Vibrio parahaemolyticus* Contamination in Blue Mussels (*Mytilus edulus*)-A Year-Long Study. Front Microbiol. 2018;9:1043. Epub 2018/06/21. doi: 10.3389/fmicb.2018.01043. PubMed PMID: 29922246; PMCID: PMC5996151.

33. Zhang H, Yang Z, Zhou Y, Bao H, Wang R, Li T, Pang M, Sun L, Zhou X. Application of a phage in decontaminating *Vibrio parahaemolyticus* in oysters. Int J Food Microbiol.

2018;275:24-31. Epub 2018/04/06. doi: 10.1016/j.ijfoodmicro.2018.03.027. PubMed PMID: 29621738.

34. Wang W, Li M, Li Y. Intervention strategies for reducing *Vibrio parahaemolyticus* in seafood: a review. J Food Sci. 2015;80(1):R10-9. Epub 2014/12/05. doi: 10.1111/1750-3841.12727. PubMed PMID: 25472618.

35. Cao Y, Zhang Y, Lan W, Sun X. Characterization of vB_VpaP_MGD2, a newly isolated bacteriophage with biocontrol potential against multidrug-resistant *Vibrio parahaemolyticus*. Arch Virol. 2021;166(2):413-26. Epub 20210103. doi: 10.1007/s00705-020-04887-x. PubMed PMID: 33389104.

36. Brossard Stoos KA, Ren J, Shields-Cutler RR, Sams KL, Caldwell S, Ho MB, Rivara G, Whistler CA, Jones SH, Wiedmann M, DeMent J, Getchell RG, Marquis H. Coastal water bacteriophages infect various sets of *Vibrio parahaemolyticus* sequence types. Front Microbiol. 2022;13:1041942. Epub 20221219. doi: 10.3389/fmicb.2022.1041942. PubMed PMID: 36601403; PMCID: PMC9807174.

37. Ding T, Sun H, Pan Q, Zhao F, Zhang Z, Ren H. Isolation and characterization of *Vibrio parahaemolyticus* bacteriophage vB_VpaS_PG07. Virus Res. 2020;286:198080. Epub 20200629. doi: 10.1016/j.virusres.2020.198080. PubMed PMID: 32615132.

38. Faruque SM, Naser IB, Islam MJ, Faruque AS, Ghosh AN, Nair GB, Sack DA, Mekalanos JJ. Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. Proc Natl Acad Sci U S A. 2005;102(5):1702-7. Epub 2005/01/18. doi: 10.1073/pnas.0408992102. PubMed PMID: 15653771; PMCID: PMC547864.

39. Nasu H, Iida T, Sugahara T, Yamaichi Y, Park KS, Yokoyama K, Makino K, Shinagawa H, Honda T. A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:K6 strains. J Clin Microbiol. 2000;38(6):2156-61. Epub 2000/06/02. PubMed PMID: 10834969; PMCID: PMC86752.

40. Iida T, Hattori A, Tagomori K, Nasu H, Naim R, Honda T. Filamentous phage associated with recent pandemic strains of *Vibrio parahaemolyticus*. Emerg Infect Dis. 2001;7(3):477-8. Epub 2001/06/01. doi: 10.3201/eid0703.010325. PubMed PMID: 11384535; PMCID: PMC2631806.

41. Iida T, Makino K, Nasu H, Yokoyama K, Tagomori K, Hattori A, Okuno T, Shinagawa H, Honda T. Filamentous bacteriophages of vibrios are integrated into the dif-like site of the host chromosome. J Bacteriol. 2002;184(17):4933-5. Epub 2002/08/10. doi: 10.1128/jb.184.17.4933-4935.2002. PubMed PMID: 12169621; PMCID: PMC135294.

42. Chan B, Miyamoto H, Taniguchi H, Yoshida S. Isolation and genetic characterization of a novel filamentous bacteriophage, a deleted form of phage f237, from a pandemic *Vibrio parahaemolyticus* O4:K68 strain. Microbiol Immunol. 2002;46(8):565-9. Epub 2002/10/05. doi: 10.1111/j.1348-0421.2002.tb02734.x. PubMed PMID: 12363020.

43. Garin-Fernandez A, Wichels A. Looking for the hidden: Characterization of lysogenic phages in potential pathogenic *Vibrio* species from the North Sea. Mar Genomics. 2020;51:100725. Epub 20191119. doi: 10.1016/j.margen.2019.100725. PubMed PMID: 31757758.

44. Garin-Fernandez A, Glockner FO, Wichels A. Genomic characterization of filamentous phage vB_VpaI_VP-3218, an inducible prophage of *Vibrio parahaemolyticus*. Mar Genomics. 2020;53:100767. Epub 20200312. doi: 10.1016/j.margen.2020.100767. PubMed PMID: 32171709.

45. Abanto M, Gavilan RG, Baker-Austin C, Gonzalez-Escalona N, Martinez-Urtaza J. Global Expansion of Pacific Northwest *Vibrio parahaemolyticus* Sequence Type 36. Emerg Infect Dis. 2020;26(2):323-6. doi: 10.3201/eid2602.190362. PubMed PMID: 31961301; PMCID: PMC6986845.

46. Banerjee SK, Kearney AK, Nadon CA, Peterson C-L, Tyler K, Bakouche L, Clark CG, Hoang L, Gilmour MW, Farber JM. Phenotypic and genotypic characterization of Canadian

clinical isolates of *Vibrio parahaemolyticus* collected from 2000 to 2009. Journal of Clinical Microbiology. 2014;52(4):1081-8.

47. Ronholm J, Petronella N, Kenwell R, Banerjee S. Draft Whole-Genome Sequences of 14 Vibrio parahaemolyticus Clinical Isolates with an Ambiguous K Serogroup. Genome Announc. 2015;3(2). Epub 2015/04/04. doi: 10.1128/genomeA.00217-15. PubMed PMID: 25838480; PMCID: PMC4384484.

48. Ronholm J, Petronella N, Chew Leung C, Pightling AW, Banerjee SK. Genomic Features of Environmental and Clinical *Vibrio parahaemolyticus* Isolates Lacking Recognized Virulence Factors Are Dissimilar. Appl Environ Microbiol. 2016;82(4):1102-13. Epub 2015/12/08. doi: 10.1128/AEM.03465-15. PubMed PMID: 26637607; PMCID: PMC4751838.

49. Frichot E, François O. LEA: an R package for landscape and ecological association studies. Methods in Ecology and Evolution. 2015;6(8):925-9.

50. Marcinkiewicz AL. Bacterial and phage interactions influencing *Vibrio parahaemolyticus* ecology.: University of New Hampshire; 2016.

51. Schmidt AK, Fitzpatrick AD, Schwartzkopf CM, Faith DR, Jennings LK, Coluccio A, Hunt DJ, Michaels LA, Hargil A, Chen Q, Bollyky PL, Dorward DW, Wachter J, Rosa PA, Maxwell KL, Secor PR. A Filamentous Bacteriophage Protein Inhibits Type IV Pili To Prevent Superinfection of *Pseudomonas aeruginosa*. mBio. 2022;13(1):e0244121. Epub 20220118. doi: 10.1128/mbio.02441-21. PubMed PMID: 35038902; PMCID: PMC8764522.

52. Ismail MH, Michie KA, Goh YF, Noorian P, Kjelleberg S, Duggin IG, McDougald D, Rice SA. The Repressor C Protein, Pf4r, Controls Superinfection of *Pseudomonas aeruginosa* PAO1 by the Pf4 Filamentous Phage and Regulates Host Gene Expression. Viruses. 2021;13(8). Epub 20210815. doi: 10.3390/v13081614. PubMed PMID: 34452479; PMCID: PMC8402870.

53. Leon-Sicairos N, Zatarain-Lopez R, Angulo-Zamudio UA, Velazquez-Roman J, Flores-Villasenor H, Martinez-Garcia JJ, Moreno-Perez MA, Buelna-Romero A, Hernandez-Monroy I, Lopez-Martinez I, Cuen-Diaz HM, Diaz-Quinonez JA, Canizalez-Roman A. *Vibrio*

parahaemolyticus Is Associated with Diarrhea Cases in Mexico, with a Dominance of Pandemic O3:K6 Clones. Int J Environ Res Public Health. 2022;19(16). Epub 20220819. doi: 10.3390/ijerph191610318. PubMed PMID: 36011953; PMCID: PMC9408606.

54. Rosales D, Ellett A, Jacobs J, Ozbay G, Parveen S, Pitula J. Investigating the Relationship between Nitrate, Total Dissolved Nitrogen, and Phosphate with Abundance of Pathogenic Vibrios and Harmful Algal Blooms in Rehoboth Bay, Delaware. Appl Environ Microbiol. 2022;88(14):e0035622. Epub 20220706. doi: 10.1128/aem.00356-22. PubMed PMID: 35862751; PMCID: PMC9317868.

55. Martinez-Urtaza J, Huapaya B, Gavilan RG, Blanco-Abad V, Ansede-Bermejo J, Cadarso-Suarez C, Figueiras A, Trinanes J. Emergence of Asiatic Vibrio diseases in South America in phase with El Nino. Epidemiology. 2008;19(6):829-37. doi: 10.1097/EDE.0b013e3181883d43. PubMed PMID: 18854707.

56. McLaughlin JB, DePaola A, Bopp CA, Martinek KA, Napolilli NP, Allison CG, Murray SL, Thompson EC, Bird MM, Middaugh JP. Outbreak of *Vibrio parahaemolyticus* gastroenteritis associated with Alaskan oysters. N Engl J Med. 2005;353(14):1463-70. doi: 10.1056/NEJMoa051594. PubMed PMID: 16207848.

57. Konrad S, Paduraru P, Romero-Barrios P, Henderson SB, Galanis E. Remote sensing measurements of sea surface temperature as an indicator of *Vibrio parahaemolyticus* in oyster meat and human illnesses. Environ Health. 2017;16(1):92. Epub 20170831. doi: 10.1186/s12940-017-0301-x. PubMed PMID: 28859689; PMCID: PMC5580290.

58. Galanis E, Otterstatter M, Taylor M. Measuring the impact of sea surface temperature on the human incidence of *Vibrio* sp. infection in British Columbia, Canada, 1992-2017. Environ Health. 2020;19(1):58. Epub 20200527. doi: 10.1186/s12940-020-00605-x. PubMed PMID: 32460848; PMCID: PMC7251872.

59. Taylor M, Cheng J, Sharma D, Bitzikos O, Gustafson R, Fyfe M, Greve R, Murti M, Stone J, Honish L, Mah V, Punja N, Hexemer A, McIntyre L, Henry B, Kendall P, Atkinson R,

Buenaventura E, Martinez-Perez A, Galanis E, Team TOI. Outbreak of *Vibrio parahaemolyticus* Associated with Consumption of Raw Oysters in Canada, 2015. Foodborne Pathog Dis. 2018;15(9):554-9. Epub 20180629. doi: 10.1089/fpd.2017.2415. PubMed PMID: 29958009.

60. Mai-Prochnow A, Hui JG, Kjelleberg S, Rakonjac J, McDougald D, Rice SA. 'Big things in small packages: the genetics of filamentous phage and effects on fitness of their host'. FEMS Microbiol Rev. 2015;39(4):465-87. Epub 2015/02/12. doi: 10.1093/femsre/fuu007. PubMed PMID: 25670735.

61. Das B, Bischerour J, Barre FX. VGJphi integration and excision mechanisms contribute to the genetic diversity of *Vibrio cholerae* epidemic strains. Proc Natl Acad Sci U S A. 2011;108(6):2516-21. Epub 2011/01/26. doi: 10.1073/pnas.1017061108. PubMed PMID: 21262799; PMCID: PMC3038760.

62. Hussain FA, Dubert J, Elsherbini J, Murphy M, VanInsberghe D, Arevalo P, Kauffman K, Rodino-Janeiro BK, Gavin H, Gomez A, Lopatina A, Le Roux F, Polz MF. Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. Science. 2021;374(6566):488-92. Epub 20211021. doi: 10.1126/science.abb1083. PubMed PMID: 34672730.

63. Secor PR, Sweere JM, Michaels LA, Malkovskiy AV, Lazzareschi D, Katznelson E, Rajadas J, Birnbaum ME, Arrigoni A, Braun KR, Evanko SP, Stevens DA, Kaminsky W, Singh PK, Parks WC, Bollyky PL. Filamentous Bacteriophage Promote Biofilm Assembly and Function. Cell Host Microbe. 2015;18(5):549-59. doi: 10.1016/j.chom.2015.10.013. PubMed PMID: 26567508; PMCID: PMC4653043.

64. Martinez-Urtaza J, Blanco-Abad V, Rodriguez-Castro A, Ansede-Bermejo J, Miranda A, Rodriguez-Alvarez MX. Ecological determinants of the occurrence and dynamics of *Vibrio parahaemolyticus* in offshore areas. ISME J. 2012;6(5):994-1006. Epub 20111117. doi: 10.1038/ismej.2011.156. PubMed PMID: 22094349; PMCID: PMC3329108.

65. Doni L, Oliveri C, Lasa A, Di Cesare A, Petrin S, Martinez-Urtaza J, Coman F, Richardson A, Vezzulli L. Large-scale impact of the 2016 Marine Heatwave on the plankton-associated microbial communities of the Great Barrier Reef (Australia). Mar Pollut Bull. 2023;188:114685. Epub 20230203. doi: 10.1016/j.marpolbul.2023.114685. PubMed PMID: 36739716.

66. Xu Y, Yang L, Wang Y, Zhu Z, Yan J, Qin S, Chen L. Prophage-encoded gene VpaChn25_0734 amplifies ecological persistence of *Vibrio parahaemolyticus* CHN25. Curr Genet. 2022;68(2):267-87. Epub 20220122. doi: 10.1007/s00294-022-01229-z. PubMed PMID: 35064802; PMCID: PMC8783578.

67. Erken M, Weitere M, Kjelleberg S, McDougald D. In situ grazing resistance of *Vibrio cholerae* in the marine environment. FEMS Microbiol Ecol. 2011;76(3):504-12. Epub 2011/02/15. doi: 10.1111/j.1574-6941.2011.01067.x. PubMed PMID: 21314704.

68. Sweere JM, Van Belleghem JD, Ishak H, Bach MS, Popescu M, Sunkari V, Kaber G, Manasherob R, Suh GA, Cao X, de Vries CR, Lam DN, Marshall PL, Birukova M, Katznelson E, Lazzareschi DV, Balaji S, Keswani SG, Hawn TR, Secor PR, Bollyky PL. Bacteriophage trigger antiviral immunity and prevent clearance of bacterial infection. Science. 2019;363(6434). doi: 10.1126/science.aat9691. PubMed PMID: 30923196; PMCID: PMC6656896.

69. Paff ML, Nuismer SL, Ellington A, Molineux IJ, Bull JJ. Virus wars: using one virus to block the spread of another. PeerJ. 2016;4:e2166. Epub 2016/07/15. doi: 10.7717/peerj.2166. PubMed PMID: 27413636; PMCID: PMC4933091.

70. Lin A, Jimenez J, Derr J, Vera P, Manapat ML, Esvelt KM, Villanueva L, Liu DR, Chen IA. Inhibition of bacterial conjugation by phage M13 and its protein g3p: quantitative analysis and model. PLoS One. 2011;6(5):e19991. Epub 2011/06/04. doi: 10.1371/journal.pone.0019991. PubMed PMID: 21637841; PMCID: PMC3102678.

71. Molina-Quiroz RC, Dalia TN, Camilli A, Dalia AB, Silva-Valenzuela CA. Prophage-Dependent Neighbor Predation Fosters Horizontal Gene Transfer by Natural Transformation.

mSphere. 2020;5(6). Epub 20201111. doi: 10.1128/mSphere.00975-20. PubMed PMID: 33177216; PMCID: PMC7657591.

72. Matthey N, Stutzmann S, Stoudmann C, Guex N, Iseli C, Blokesch M. Neighbor predation linked to natural competence fosters the transfer of large genomic regions in *Vibrio cholerae*. Elife. 2019;8. Epub 20190903. doi: 10.7554/eLife.48212. PubMed PMID: 31478834; PMCID: PMC6783263.

73. Dalia AB, McDonough E, Camilli A. Multiplex genome editing by natural transformation. Proc Natl Acad Sci U S A. 2014;111(24):8937-42. Epub 20140602. doi: 10.1073/pnas.1406478111. PubMed PMID: 24889608; PMCID: PMC4066482.

74. Panicker G, Call DR, Krug MJ, Bej AK. Detection of pathogenic *Vibrio spp.* in shellfish by using multiplex PCR and DNA microarrays. Appl Environ Microbiol. 2004;70(12):7436-44.

75. Whistler CA, Hall JA, Xu F, Ilyas S, Siwakoti P, Cooper VS, Jones SH. Use of Whole-Genome Phylogeny and Comparisons for Development of a Multiplex PCR Assay To Identify Sequence Type 36 *Vibrio parahaemolyticus*. Journal of clinical microbiology. 2015;53(6):1864-72.

76. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive multiplexed library preparation for megabase-sized genomes. PloS one. 2015;10(5):e0128036.

77. Tritt A, Eisen JA, Facciotti MT, Darling AE. A5. An integrated pipeline for *de novo* assembly of microbial genomes. PLoS ONE. 2012;7(9):e42304 DOI: 10.1371/journal.pone.0042304.

78. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014:DOI:10.1093/bioinformatics/btu153.

79. Inouye M, Conway TC, Zobel J, Holt KE. Short read sequence typing (SRST): multi-locus sequence types from short reads. BMC Genomics. 2012;13(1):338.

80. Jolley KA, Maiden MC. Using MLST to study bacterial variation: prospects in the genomic era. Future microbiology. 2014;9(5):623-30.

81. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 2014;6(11):90. Epub 2014/11/26. doi: 10.1186/s13073-014-0090-6. PubMed PMID: 25422674; PMCID: PMC4237778.

82. Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics. 2015;31(17):2877-8. Epub 2015/04/29. doi: 10.1093/bioinformatics/btv271. PubMed PMID: 25913206.

83. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688-90.

84. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256-W9. Epub 2019/04/02. doi: 10.1093/nar/gkz239. PubMed PMID: 30931475; PMCID: PMC6602468.

85. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010;5(6):e11147. Epub 2010/07/02. doi: 10.1371/journal.pone.0011147. PubMed PMID: 20593022; PMCID: PMC2892488.

86. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27(7):1009-10. Epub 2011/02/01. doi: 10.1093/bioinformatics/btr039. PubMed PMID: 21278367; PMCID: PMC3065679.

87. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic acids research. 2014:gku1196.

88. T. S. Snippy: fast bacterial variant calling from NGS reads [Internet]. Available from https://github.com/tseemann/snippy 2015.

89. Wickham H. ggplot2: elegant graphics for data analysis: Springer; 2016.

90. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol. 2014;1151:165-88.

Epub 2014/05/20. doi: 10.1007/978-1-4939-0554-6_12. PubMed PMID: 24838886; PMCID: PMC4239701.

91. 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 MR, Barylski J, Ackermann HW, Kuhn JH. Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. Arch Virol. 2016;161(4):1095-9. Epub 2016/01/07. doi: 10.1007/s00705-015-2728-0. PubMed PMID: 26733293.

92. Adriaenssens E, Brister JR. How to Name and Classify Your Phage: An Informal Guide. Viruses. 2017;9(4). Epub 2017/04/04. doi: 10.3390/v9040070. PubMed PMID: 28368359; PMCID: PMC5408676.

93. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10. Epub 1990/10/05. doi: 10.1016/S0022-2836(05)80360-2. PubMed PMID: 2231712.

94. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 1986;3(5):418-26. Epub 1986/09/01. doi: 10.1093/oxfordjournals.molbev.a040410. PubMed PMID: 3444411.

95. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725-9. Epub 2013/10/18. doi: 10.1093/molbev/mst197. PubMed PMID: 24132122; PMCID: PMC3840312.

96. Taniguchi H, Sato K, Ogawa M, Udou T, Mizuguchi Y. Isolation and characterization of a filamentous phage, Vf33, specific for Vibrio parahaemolyticus. Microbiology and immunology. 1984;28(3):327-37.

97. Chang B, Yoshida S, Miyamoto H, Ogawa M, Horikawa K, Ogata K, Nishibuchi M, Taniguchi H. A unique and common restriction fragment pattern of the nucleotide sequences homologous to the genome of vf33, a filamentous bacteriophage, in pandemic strains of *Vibrio*

parahaemolyticus O3:K6 O4:K68, and O1:K untypeable. FEMS Microbiol Lett. 2000;192(2):231-6. Epub 2000/11/07. doi: 10.1111/j.1574-6968.2000.tb09387.x. PubMed PMID: 11064200.

98. Nair GB, Ramamurthy T, Bhattacharya SK, Dutta B, Takeda Y, Sack DA. Global dissemination of *Vibrio parahaemolyticus* serotype O3: K6 and its serovariants. Clinical Microbiology Reviews. 2007;20(1):39-48.

99. Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013;14:60. Epub 2013/02/26. doi: 10.1186/1471-2105-14-60. PubMed PMID: 23432962; PMCID: PMC3665452.

100. Meier-Kolthoff JP, Goker M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics. 2017;33(21):3396-404. Epub 2017/10/17. doi: 10.1093/bioinformatics/btx440. PubMed PMID: 29036289; PMCID: PMC5860169.

101. Nordstrom JL, Vickery MC, Blackstone GM, Murray SL, DePaola A. 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. 2007;73(18):5840-7. Epub 2007/07/24. doi: 10.1128/AEM.00460-07. PubMed PMID: 17644647; PMCID: PMC2074920.

Figure 1

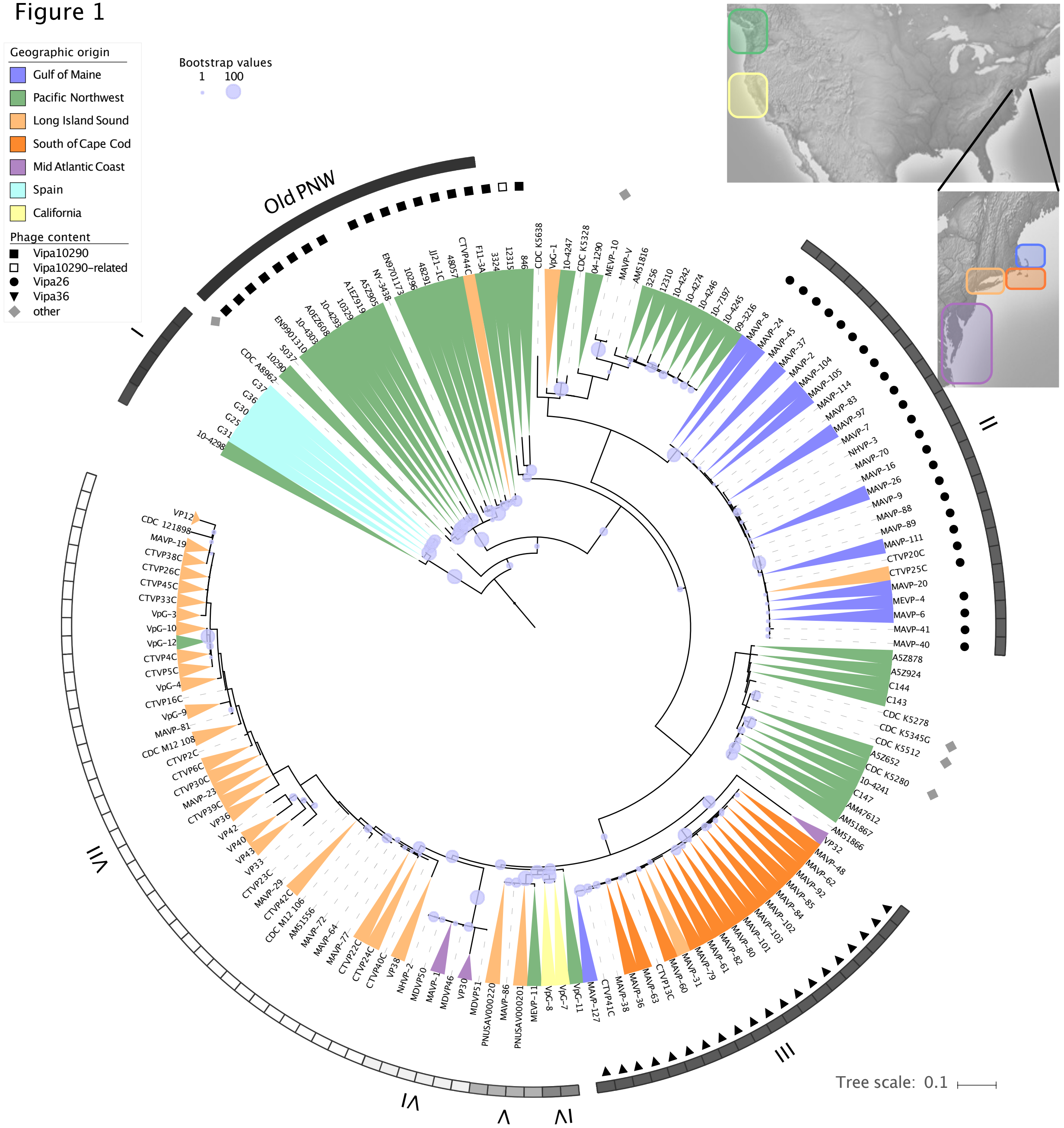

Figure 1. Maximum-likelihood phylogeny of ST36 Vibrio parahaemolyticus. ML-phylogenieswere builtwith whole genome variation of select quality ST36 strains. Isolates are colored to correspond to geographic origin trace-back source, where available (Table S1) with isolates reported in the PNW prior to 2010 are presumed to originate in that region, but thereafter only assigned to region when supported by traceback, and where no color indicates unknown or ambiguous origin. Symbols next to strains indicate unique inovirus content. The ancestral PNW ST36 population is identified by a black bar, whereas seven different lineages associated with translocation events (I - VII) are identified by greyscale bars.

Figure 2

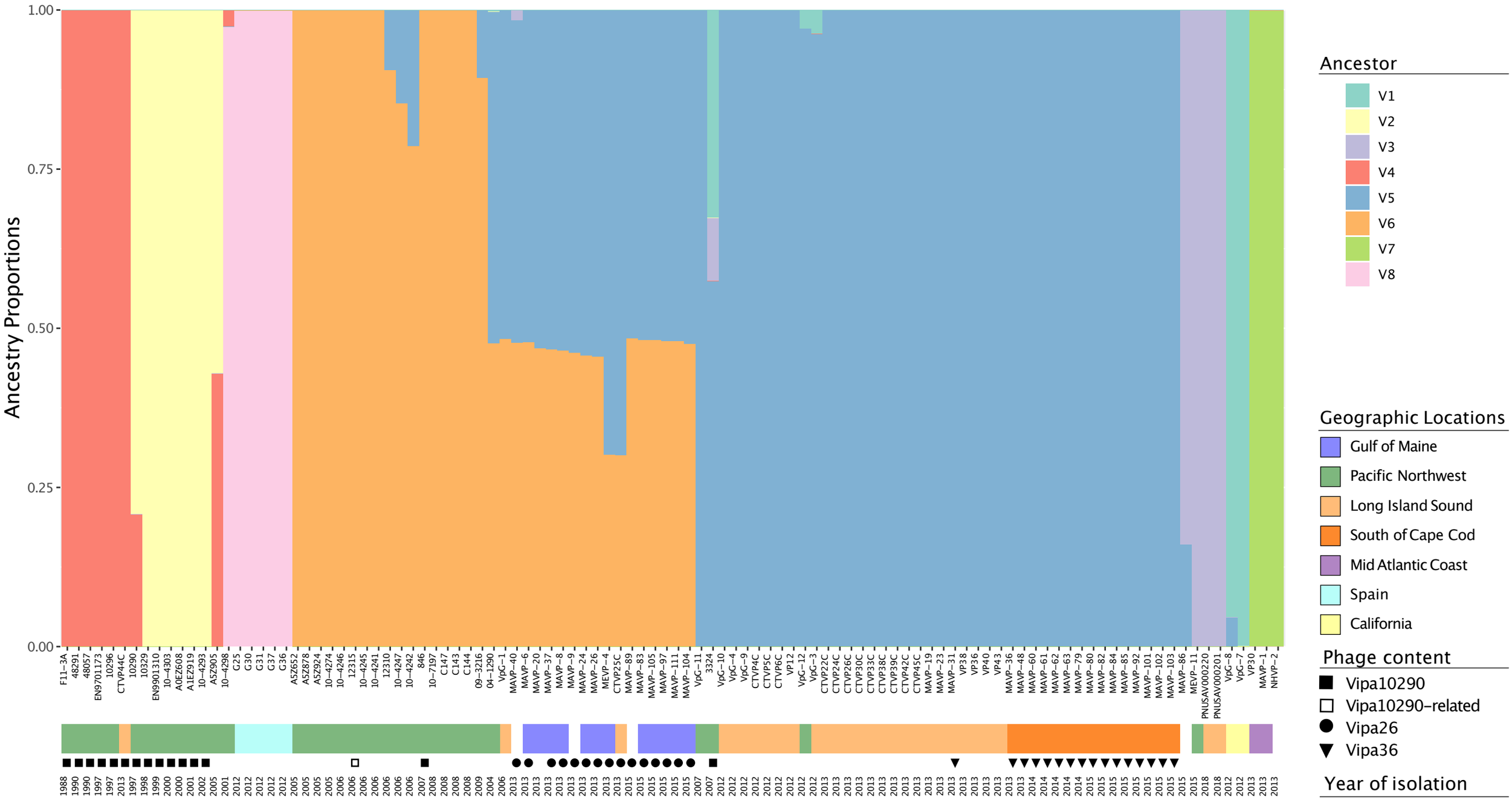

Figure 2. Population structure of ST36 strains. Coancestry estimates were inferred from core genomes of ST36 strains by LEA (Frichot and François, 2015). Colored bars represent proportion of genetic variation derived from 8 ancestral populations (see Fig S3A) with strong temporal/geographic associations. Geographic locations of strains are marked by colors below strain name, followed by phage content (symbol) and year of isolation. ame, followed by phage content (symbol) and year of isolation.

Figure 3

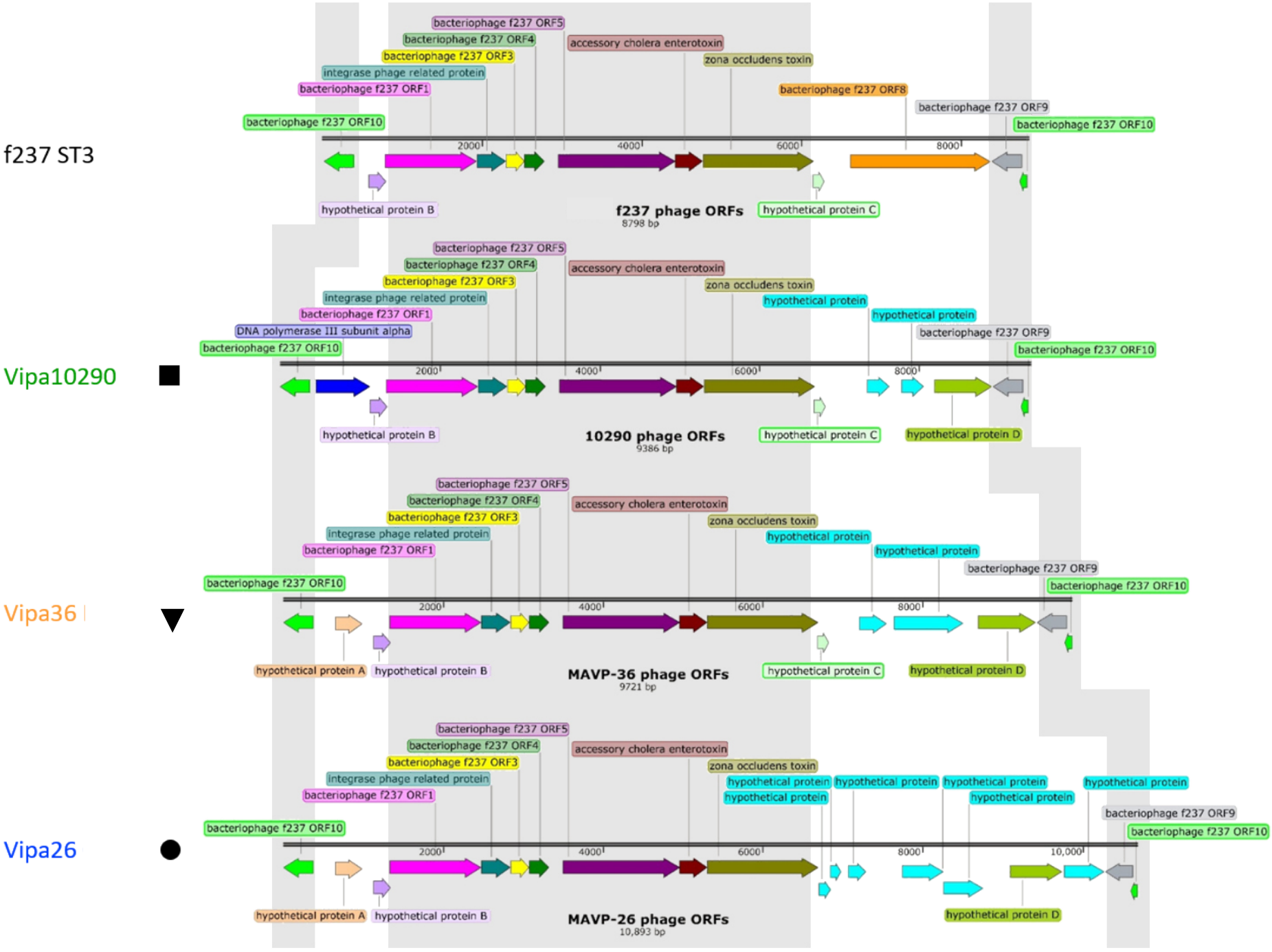

Figure 3. Structure of V. parahaemolyticus Inoviruses. ORFs with the same color are homologous between the different phages with the exception of the blue hypothetical proteins. The core genome (highlighted in gray) is highly conserved, whereas accessory regions vary, with some ORFs conserved in two or more of the phages.

Figure 4

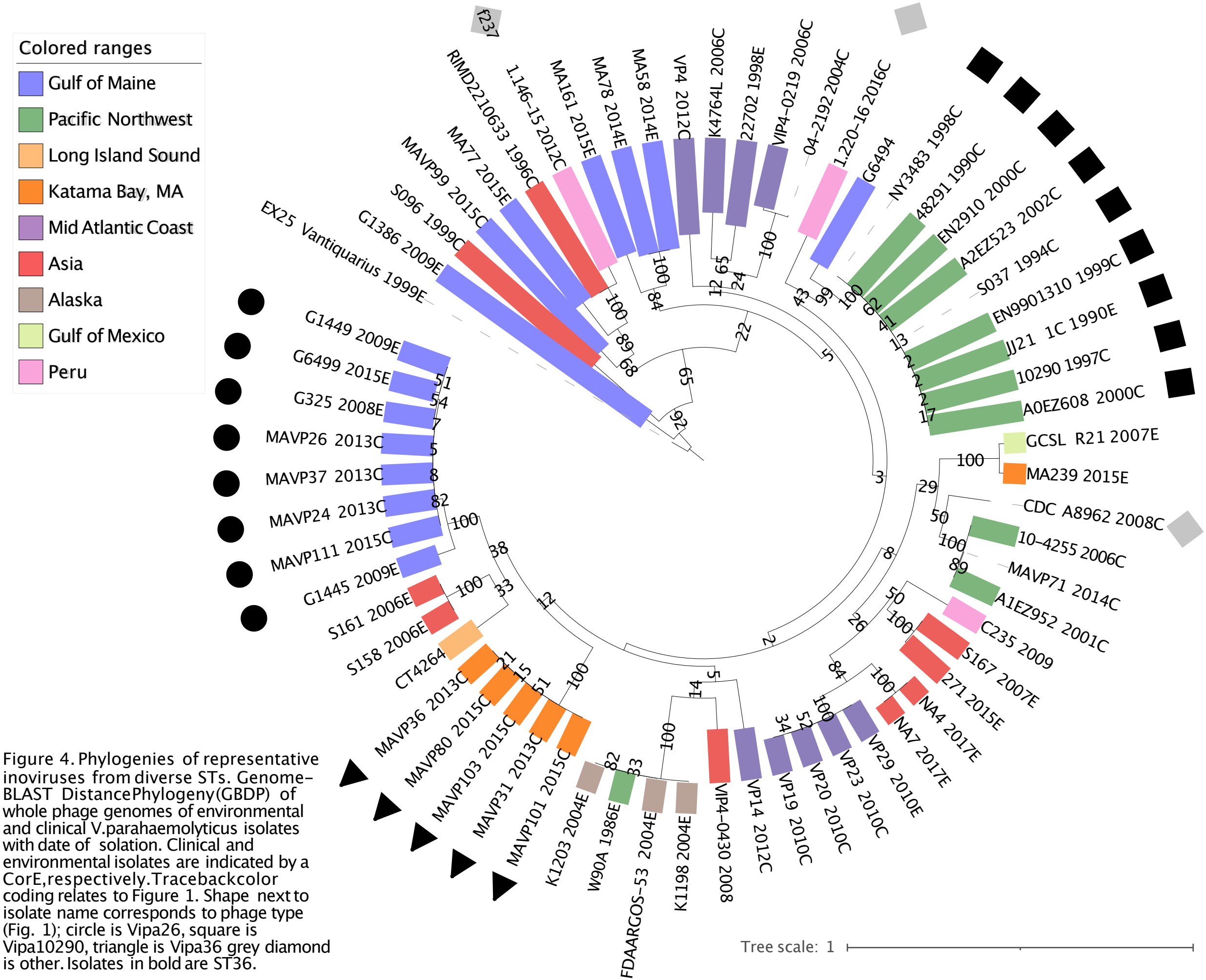

Figure 4. Phylogenies of representative inoviruses from diverse STs. Genome-BLAST DistancePhylogeny(GBDP) of whole phage genomes of environmental and clinical V.parahaemolyticus isolates with date of solation. Clinical and environmental isolates are indicated by a CorE,respectively.Tracebackcolor coding relates to Figure 1. Shape next to isolate name corresponds to phage type (Fig. 1); circle is Vipa26, square is Vipa10290, triangle is Vipa36 grey diamond is other. Isolates in bold are ST36.

Figure 5

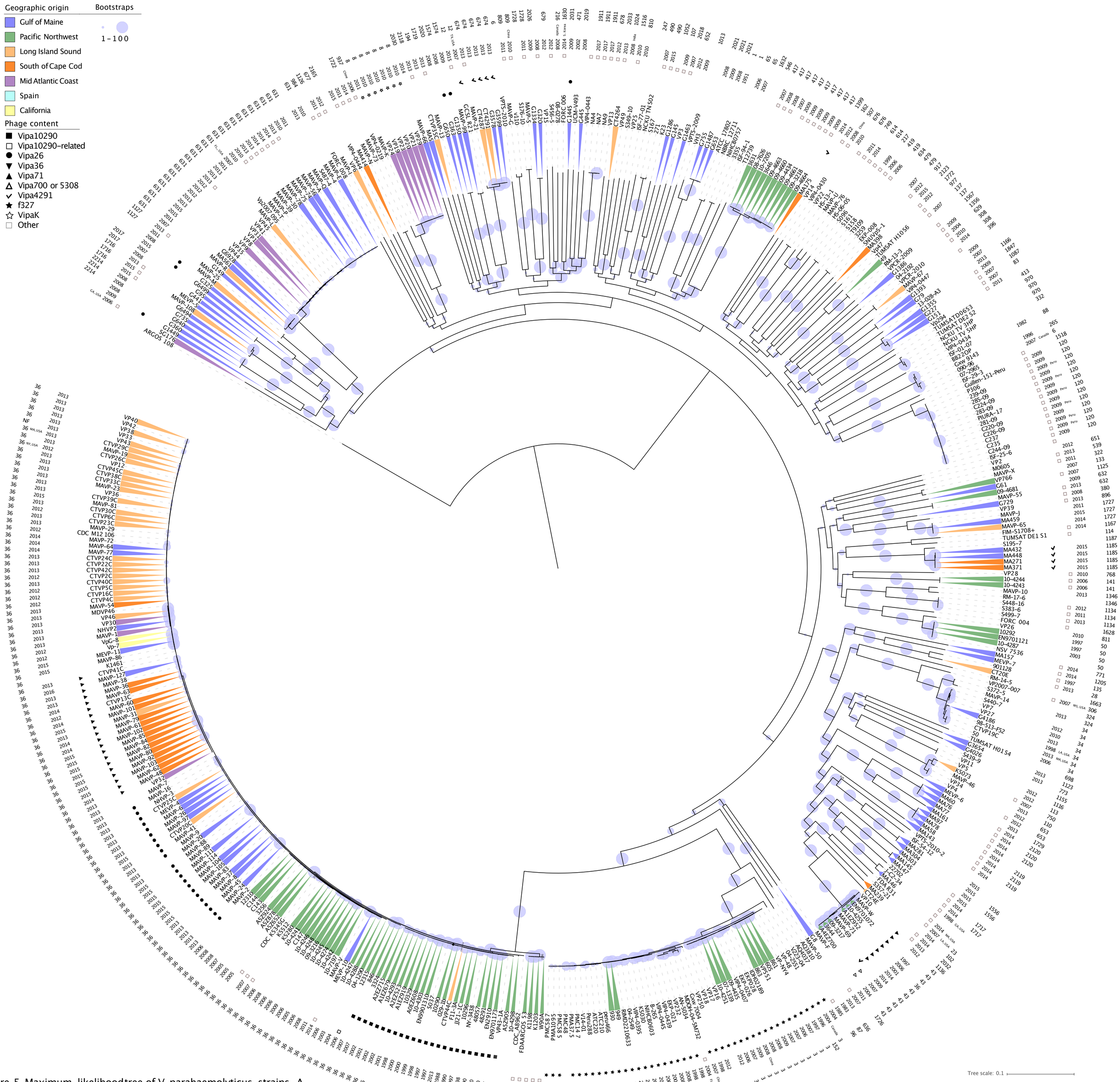

Figure 5. Maximum-likelihood tree of V. parahaemolyticus strains. A maximum-likelihood(ML) phylogeny was built on core genome SNPs where 100 bootstraps were mapped onto the best scoring ML tree. Isolates are colored to correspond to geographic region as in fig. 1, where no color indicates unknown or ambiguous origin. Symbols next to strains indicate unique inovirus content.