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

Randi L. Foxall[a,b*], Jillian Means[a,c,d*], Ashley L. Marcinkiewicz[a,c,e*], Christopher Schillaci[a,f,g], Kristin DeRosia-Banick[a,h], Feng Xu[a,b,i], Jeffrey A. Hall[a,j], Stephen H. Jones[a,f], Vaughn S. Cooper[a,b,k], Cheryl A. Whistler[a,b,#]

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

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

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

[d] Current address: Synlogic Inc., Cambridge, MA, USA

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

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

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

[h] Current address: Milken School of Public Health, George Washington University, Washington D.C., USA

[i] Current address: Children's Hospital of Philadelphia, Philadelphia, PA, USA

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

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

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

* Contributed equally to this work. Ms. Marcinkiewicz initiated the study and documented the phage distribution in clinical and environmental samples, associated phage with aquaculture areas, and conducted initial phylogeographic analyses. Ms. Means extended the research and assembled inovirus genomes, conducted phylogenetic and bioinformatics analyses of phage, produced visuals and tables for the manuscript, and completed the first draft of the manuscript. Ms. Foxall conducted the final genome analyses of both host and inoviruses, completed population structure analyses, identified inoviruses from public databases, conducted comprehensive analyses of global patterns of phage association and relatedness, conducted all analyses requested by reviewers, and produced both visuals and methods for final publication.

Running Title: Inovirus mediated evolution of *Vibrio parahaemolyticus*

Abstract word count: 190

Importance word count: 120

Manuscript word count: 4987

Abstract: The epidemiology of *Vibrio parahaemolyticus,* the leading cause of seafood-borne bacterial gastroenteritis of humans world-wide, dramatically changed in the United States following the establishment of a Pacific native lineage called sequence type (ST) 36 in the Atlantic. In this study we used phylogeography based on traceback to environmental source locations and comparative genomics to identify features that promoted evolution, dispersal, and competitive dominance of ST36. The major genomic differentiation and competitive success of ST36 was associated with a striking succession of filamentous prophage in the family *Inoviridae* (inoviruses), including loss of an inovirus prophage that had been maintained for decades in 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 stable and 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 inovirus replacements, such as in ST36, appear to be 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 of humans world-wide recently increased along the United States (US) Atlantic coast (1-3). Rising seasonal illnesses were tied to warming ocean temperatures, a growing aquaculture industry, and the emergence of endemic pathogens (3-6). But incursion of a Pacific Northwest (PNW) lineage called sequence type (ST) 36 into the Atlantic following an anomalously mild winter is the most important driver of this shift (3, 5, 7). ST36 began causing infections in 1979 (8) and was originally limited to the Pacific Northwest (PNW) (9). Although sporadic infections were occasionally reported outside the PNW, local sources were rarely implicated (7, 9, 10). This changed in 2012 when ST36 began causing illnesses traced to Atlantic sources and in 2013 it caused a 13-state outbreak traced to multiple northeast US locations (5, 7). Unlike previous non-endemic strains causing outbreaks from Atlantic sources, including the pandemic ST3 (11), and ST8 (12), ST36 continues to cause sporadic disease from a few northeast US locations. A better understanding of how and where ST36 established populations is needed to aid in management to limit illness. Furthermore, this 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). Disease patterns implied potential reciprocal transfer between both US coasts, though inference was informed by a few cases where infection occurred from shellfish consumed on the opposite coast from where illnesses were reported and further complicated by modeling of a mode of

spread through human populations (9), whereas seasonal illnesses by *V. parahaemolyticus* are typically vectored by environmentally contaminated raw seafood. This highlights a continuing challenge where a lack of environmental traceback in clinical reporting that uncouples strains from their source bodies of water obscures environmental distribution (5, 9). As with other human pathogens originating from the natural environment, this leaves a critical gap of knowledge of the ecological context of pathogen evolution and expansion.

Bacteriophages, the viruses that infect bacteria, can shape, and quell environmentally vectored human pathogen populations (13-19). Lytic phages 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 death 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 roles of phage in mediating 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.

Prior 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. Here, we examined the phylogenomics of the Atlantic ST36 incursion—accompanied by interactions with filamentous *Inoviridae* phage—by capturing a broad geographic and temporal distribution of the population. By curating strains

with environmental traceback, we determined the phylogeography of ST36 wherein multiple strains that first diversified in the Pacific, subsequently translocated to the Atlantic, and established spatially distinct and non-mixing subpopulations therein. This also revealed that changes in inovirus content accompanied diversification and lineage replacement in the PNW and preceded clonal expansion by distinct founders of two persisting Atlantic ST36 populations that continue to cause sporadic disease.

Results

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

To investigate the evolution of ST36 as it translocated from the PNW to the Atlantic, we curated a collection of genomes and mapped their environmental sources onto their constructed whole genome phylogenies. Locations include two Pacific (PNW and California [CA]), and five Atlantic (Galicia Spain, the Gulf of Maine [GOM], an Island south of Cape Cod, [SCC], Long Island Sound [LIS] and the mid-Atlantic coast [MAC] [Table S1]). The phylogenies exemplify how multiple clades emerged from the old-PNW population (Fig. 1) and gave rise to the new-PNW population (9). A single relic of the old-PNW population and 103 isolates from seven lineages arising from the modern ST36 diversification were traced to environmental locations outside the PNW (Fig. 1, clades I-VII). Notably, most translocated clades either contain a PNW-traced isolate or are most closely related to clades dominated by PNW isolates suggesting these differentiated in the PNW prior to translocation rather than after arrival in the Atlantic.

Assignment of isolates to environmental sources revealed striking geographic structure of the Atlantic subpopulations in GOM, SCC, and LIS, indicating that each was founded by a genetically unique individual (Fig. 1). All but one isolate traced to the GOM are clonal (clade II) and share ancestry with PNW isolates (9, 46-48), and the 2006 New York isolate VpG-1 (9)(Fig.

1, Table S1). Most isolates traced to other Atlantic locations and those from CA share common ancestry with isolates reported from British Columbia Canada in 2005, and later in WA in 2011-2012 (Fig. 1 and Table S1). All but one SCC-traced isolate (clade III) are clonal and only two clade members trace to other nearby locations. LIS-traced isolates are also mostly clonal (76%) (clade VII), though two distinct 2018 isolates group within a mixed-location clade (clade V) with isolates trace to the PNW in 2011, and from oysters consumed in CA in 2015. 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 these shared in common with the GOM and SCC clades respectively may have obscured their distinct heritage (Fig. 1, S1 & S2; Tables S2- S4).

Given the geographic linkage of clonal subpopulations (Fig. 1), we examined the population genetic structure of isolates from known locations to identify patterns of coancestry and admixture (49). The conservative use of core variation and eight ancestral populations (which best explains the data, see Fig. S3A & B) identified two old-PNW populations that were replaced by a single modern-day PNW population, comprised of two geographically distinct lineages: one related to GOM isolates and the other to most 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 other translocated isolates, including those traced to Spain. Although no PNW isolates belong to the population that translocated to CA, two PNW isolates exhibit a shared history of admixture with the CA isolates providing clues to this lineage's origin. The clonal clades from SCC and LIS, which are in very close proximity to each other and without geographic barriers, also belong to the same population, whereas two MAC isolates form a distinct population. Atlantic-derived isolates from locations south of the GOM exhibit little coancestry with other populations, whereas the GOM clade displays near equal mixed

171 coancestry between the prevalent modern PNW population and the other Atlantic populations
172 (Fig. 2). Five PNW isolates also exhibited this same pattern of mixed heritage.
173 Analysis of admixture that incorporated horizontally acquired content separated the two
174 Atlantic populations on opposite sides of Cape Cod, MA (GOM and SCC) from the rest of those
175 in the Atlantic which grouped with strains from CA (Fig. S4). This also relocated a single GOM
176 clade isolate (MAVP-20) into the CA/MAC/LIS population. This suggested that horizontally
177 acquired variation in the SCC and GOM lineages (that MAVP-20 lacks) may define their
178 uniqueness. In contrast to the GOM and SCC populations that exhibit little admixture, the
179 modern PNW ST36 isolates exhibit complex admixture as do members of the LIS population
180 (Fig. S4). Relatively few individuals within the GOM and SCC populations display some
181 admixture with each other. Thus, population structures inferred from horizontal variation better
182 aligned most strains to their geographic locations even as horizontal content overshadowed
183 some interesting ancestral patterns, specifically, the shared mixed heritage of the GOM clade
184 with modern PNW strains (Fig. 2).

185 *Inovirus loss and reacquisition accompanied ST36 population replacement and expansion*

186 One explanation for the differences in core versus whole genome population structure is
187 that differential genetic gains or losses preceded clonal expansion. To identify such features, we
188 compared high-quality genomes of four Atlantic clade members, (MAVP-1, MAVP-23, MAVP-
189 36, and MAVP-26), to PNW strains of the new-PNW (12310) and old-PNW (10290) clades. This
190 revealed MAVP-26, MAVP-36, and 10290 each contain a unique prophage integrated into the
191 *dif* site at the replication terminus of chromosome I, similar to prophage f237 found in pandemic
192 O3:K6 strains (39, 41) (Fig.3). In contrast, the other two Atlantic clade members and the new-
193 PNW isolate lack a phage in this location (Fig. 3). These are classified in the family *Inoviridae*
194 (hereafter inoviruses) and we 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).

These inoviruses have a conserved central core (ORF 1-7) and ORF9 and ORF10 (Fig. 3A), and two variable regions. Whereas core phage gene functions were identifiable (Table 1), the functions encoded in variable regions were not discernable. Nucleotide variation in the inoviruses represented up to 46% of their host’s genome variation helping to explain differences between whole and core genome phylogenies (Fig. 1, S1 & S2) and population structures (Fig. 2 and Fig. S4). Though this content would be credited to recombination, surprisingly, genomes that lack inoviruses had a higher proportion of gene content assigned to blocks of recombination: 2.7% in 12310 and MAVP-1, and 2.0% in MAVP-23, compared to 1.3% in 10290, 1.4% in MAVP-26, and 1.8% in MAVP-36 (Fig. S1 and Table S3). Though divergent from one another, inovirus variation likely did not alter function: the 74 variant sites in the core of Vipa26 generated only three non-synonymous mutations, whereas the 78 variant sites in Vipa36 produced 13 non-synonymous mutations (Fig. 3B, Table S6). ORF1-7 are under purifying selection (codon-based Z test =9.425, $p < 0.001$) implying these are essential genes, and that their function is preserved. This also suggests the infections not cryptic, and the phage are likely still functional. Comparisons of core inovirus content from additional ST36 lysogens revealed 100% identity within the Vipa26 and Vipa36 lineages, whereas Vipa10290 had 18 non-conserved variable sites, perhaps reflecting its long history with a sizeable ST36 population (Fig. 3B).

With only one exception (A5Z905), all old-PNW clade members, including CTVP44C, contained Vipa10290 or related phage (Fig. 1, square). ST39, a member of the same clonal complex as ST36, also contains Vipa10290 whereas other clonal complex members (ST59 and ST21), share a distinct inovirus (Fig. S5, Tables 2, S9, & S10). Despite this history of lysogeny among the endemic population, no modern-day and definitively traced PNW ST36 isolates harbor inoviruses (Fig. 1 & Table S1 & S7) and this loss coincides with population diversification

(Fig. 1). However, like GOM and SCC clade members, a total of nine other modern ST36 isolates collected between 2007-2016, some that dispersed to new locations and all others untraced, acquired eight other inoviruses (Fig. 1, Table 2).

The distribution of community-acquired phage offers clues of residency and dispersal.

For the Atlantic ST36 populations, inovirus content was diagnostic of 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 either basal or independent acquisition (Fig. 1 circles). Vipa36 (Fig. 1 triangles) is in all members of the SCC clade and only two isolates from this clade were traced to nearby locations. Because these inoviruses, like f237, are not widespread and have been maintained by these lineages tied to two locations, 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 inoviruses with two Atlantic clades, and their absence in others, suggests that the founding subpopulation progenitors acquired phage upon Atlantic invasion only in some locations perhaps reflecting phage predation pressure. Environmental surveys revealed that although 22% of Northeast US *V. parahaemolyticus* isolates harbor inoviruses, this proportion varied by location (50). Proportions were lower in LIS (10%), than in GOM and SCC (30% and 42% respectively). Whereas no environmental isolates other than ST36 harbor Vipa36, multiple non-ST36 isolates collected in NH as early as 2008 contain Vipa26 (Table S1, Fig. 4, & S6). This suggests Vipa26 was acquired locally as supported by its absence from VpG-1 (Fig. 1, Table S1 & S7).

Because comparative phage phylogeny could elucidate the history of ST36 as members moved through other populations, we expanded our search for inoviruses among available genomes of *V. parahaemolyticus.* These prophage genomes added to the *Inoviridae* diversity

(Fig. & S5) but display little discernable phylogeographic population structure in that related phage are found in both the Atlantic and Pacific *V. parahaemolyticus* populations, and phage in Asian isolates group among most branches, with the exception of Vipa10290. Two ST36 isolates harbor two different inoviruses maintained by members of the PNW co-resident but unrelated ST43 (Table 2, Fig. S5). Though most members of the two ST36 lineages reported in Peru also lack inoviruses, including a 2011 environmental isolate related to the Spain lineage (45), one clinical isolate of this lineage harbors a Vipa10290-related inovirus, whereas a second clinical isolate (1.146-15), harbors f237 (Fig. 4).

The phylogenomic distribution of inoviruses in all publicly available *V. parahaemolyticus* genomes indicates they are common, with 46% of genomes containing one or more inovirus (Fig. 5, Fig. S6, and Table S10) yet distributed unevenly. Many lineages harbor a persisting inovirus (Fig. 5 and Fig. S6). A few lineages harbor two evolutionary distinct inoviruses concurrently suggesting that protection from superinfection, a common attribute of both temperate phage and inoviruses (51, 52), is not absolute (Fig. S5). In addition to Vipa26, a second inovirus is present in multiple, unrelated STs (Fig. S6). Even though most STs have an inovirus prophage that they propagate vertically, the inoviruses of some isolates were replaced by another, like ST36 (Table 3, Fig. S6). Notable among these are emergent pathogenic lineages including ST43 and the prevalent Atlantic endemic ST631 lineage that has caused illnesses along the North American Atlantic coast (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 changes in inovirus prophage accompanied diversification, competitive replacement, dispersal, and Atlantic invasion of *V. parahaemolyticus* ST36. This

suggests that inoviruses could countribute to the emergence of new lineages with the potential to spread more broadly. The misleading nature of disease reporting that does not align illnesses with environmental reservoirs and declining availability of recent ST36 genomes from the US PNW have been obstacles to deciphering the ecological context of ST36 evolution and spread. We addressed these limitations using genomes from isolates reported in Canada and other US states that were publicly available and by augmenting available genomes with a collection of ST36 clinical isolates from the Northeast US. We carefully curated environmental traceback to advance our understanding of ST36 ecological expansion into the Atlantic Ocean.

Our analysis indicates five or more unique ST36 progenitors diversified in the Pacific and then translocated into the Atlantic to found subpopulations (Fig. 1, S2, 2, and S4). The earliest sequenced ST36 isolate traced to the Atlantic in 2006 (VpG-1) and several PNW strains share a similar pattern of mixed ancestry with this isolate and group with the GOM clade suggesting its progenitor may have arrived 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 more related to each other, and multiple Pacific-traced isolates group within or adjacent to their clades (Fig. 1, 2, & S2). This pattern is most simply explained by divergence of these populations' ancestors in the Pacific prior to introduction into the Atlantic (Fig. 1, 2 & S2). Clade V's shared ancestry with CA and PNW isolates further implicates the PNW as the source population. And although imperfect trace back or mislabeled genomes could explain some unexpected patterns, nearly every Pacific-traced relative of Atlantic residential populations were isolated, reported, and sequenced independently and prior to Atlantic isolation (Table S1). Although our interpretations of population evolution and dispersal contrast with previous reports (9), this is mostly explained by differences in geographic assignment of isolates to environmental sources instead of the readily available location of disease reporting as our goal was to understand dynamics affecting reservoirs of emergent pathogenic lineages (See details of Table S1).

Traceback confirmation from CA (Table S1), disease reporting in Mexico (53) and environmental isolation in Peru (9, 45) further corroborate that these lineages diversified in the Pacific. And though missing intermediate populations could suggest unknown reservoirs exist, it is more likely that sampling bias, a decrease in availability of PNW genomes, decreased clinical prevalence of ST36 in the PNW, and lack of surveillance there have compounded challenges with deciphering this pathogen's evolution and spread. Based on mounting evidence that most if not all lineages arose in the PNW, this warrants further examination of the PNW environmental reservoir to manage further spread.

Even as increasing ocean temperatures and weather events have been tied to both seasonal illnesses by *V. parahaemolyticus* and its spread to new locations (9, 54-57), the stable and non-mixing distribution of Atlantic ST36 lineages suggests that environmental fitness that promotes persistence may be their key to success. It is notable that in BC Canada where many diversified ST36 isolates were first reported, rising infection rates were not fully explained by temperature models but rather attributed to new pathogenic strains (58, 59). Like the Northeast US, the MAC experienced a warming trend in 2012 (9), nonetheless this location has not experienced persistent problems with ST36 (1, 10). Combined, we believe warming ocean temperatures are only part of the picture and that underexplored ecological interactions are at play in driving *V. parahaemolyticus* invasion and persistence (60). Even in locations in the Atlantic with recurrent illness, environmental prevalence of ST36 is extremely rare (Table S1) illustrating how low-level persisting populations can change epidemiology. Identification of unique ecological and population contexts of the MAC compared to more northern locations could provide useful insight.

Phage loss and later 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. Nearly all old PNW lineage members harbor Vipa10290 (Fig. 1 and Table S1), whereas

strains of the new ST36 lineages isolated in the PNW after 2002 lack inoviruses, though other phage are not uncommon (Table S3). The presence of Vipa10290 in other clonal complex members could indicate a basally acquired phage was lost during lineage diversification. Though inovirus excision is infrequently reported (61, 62), laboratory passage of ST36 can lead to both phage curing and loss of protection from superinfection (63). The lack of inovirus-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 inoviruses by diverging ST36 lineages (Table 2) indicates ST36 remained susceptible to infection, though the persisting inovirus-free state of modern-PNW ST36 lineages suggests that resistance may have evolved during their diversification, as has been described in other sympatric *Vibrio* populations (64). In contrast, that the progenitors of two clinically prevailing Atlantic ST36 lineages acquired inoviruses prior to clonal expansion, while ST36 populations in the LIS where inoviruses are less abundant did not, indicates phage acquisition was not requisite for success, perhaps reflecting differences in phage predation and other population dynamics. Even so, whereas resident ST36 from SCC and GOM have continued to cause sporadic disease, infections from LIS have precipitously declined, in part reflective of successful management.

Inovirus prophage confer obvious costs through persistent non-lethal virion production (61), suggesting their loss would be favored; and yet inoviruses have been maintained by prevalent lineages for decades (Table 2 & 4, Fig. 5). This could reflect their fastidious nature, but the success of the two northernmost Atlantic populations harboring inoviruses contrasting with the unsuccessful lineages that lacked them signals alternatives are possible: that inoviruses may confer advantages in some as yet undefined context. For example, shed inovirus virions contribute to biofilm matrix (65) thereby promoting virulence and antimicrobial

resistance, but also potentially enhancing survival during long-distance movement on particles (52, 54, 66-68). Shed inoviruses can mislead human immune responses and decrease host ability to clear bacterial infections, thereby enhancing virulence (69). Furthermore, variable content in some inoviruses is linked to anti-predation (17, 70) and virulence (22, 27, 39). We find it intriguing that succession in the PNW and loss of inoviruses by ST36 accompanied its clinical decline there. Future functional studies with these inoviruses are essential to discern whether any of these possibilities are at play.

Prophage can also modify their hosts in ways that alter their capacity for horizontal gene transfer (HGT). Prophage can protect against superinfection, including by unrelated lytic phages (27, 61, 71) thereby blocking transduction. They also modulate, and concurrently block conjugative pili (72). The concept that loss of prophage-conferred immunity contributed to ST36 evolution is attractive considering DNA release by predation of non-lysogen susceptible hosts could promote transformation through natural competence, though the conditions of natural competence by *V. parahaemolyticus* are unknown (73-75). Type IV pili which are the receptor for some inoviruses also facilitate transformation mediated HGT (51, 52). If the newly inovirus-less ST36 lineages were more vulnerable to phage infections, phage transduction, and conjugal transfer alike this could have contributed to its rapid diversification (Fig. 1) whereas translocating lineages would likely not have evolved mechanisms to resist resident Atlantic phage facilitating re-acquisition (31, 52). In keeping with this premise, many of the most highly divergent ST36 individuals lacking inoviruses evolved by multiple recombination events including an isolate of the MAC clade (VP30). Accumulation of non-inovirus prophage was notable in the early diverging Spain lineage (clade I) (9). And whereas the inovirus-harboring SCC and the GOM clades remain remarkably stable, as do their phage, Atlantic isolates lacking inoviruses, including the single GOM clade member (MAVP-20) and many members of the LIS population exhibit complex admixture (Fig. S4). In contrast, the inovirus-harboring old-PNW relic

CTVP44C isolated in LIS in 2013 is virtually unchanged from the last known members of this lineage from the PNW that were isolated in 2002 (Fig. 1, Table S1). Furthermore, a 2005 isolate of this same lineage lacking Vipa10290, A5Z905, is among the most highly divergent ST36 strains (Fig. 1, 2, S2). Similarly, pandemic O3:K6 strains, which stably maintain f237, reported in the Americas between 1996-2012 display remarkably little genetic variation from the type strain for this lineage RIMD 2210366 (76). Though these data only establish correlation of phage absence and diversification, it alludes to the possibility that 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 mechanisms that could promote rapid evolution, potentially of more or even less pathogenic potential.

These analyses suggest an intriguing connection of inovirus prophage to pathogen evolutionary dynamics, alluding to the possibility that inoviruses could be gatekeepers of HGT in natural ecosystems driving changes in human disease epidemiology. With broader population analysis and functional studies, this 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 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, 64). 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 genomes

Regional clinical strains spanning 2010-2018, and trace back were acquired from public health laboratories (2). Only isolates with single source traces were assigned a location. Environmental isolates were collected between 2007-2015 (1, 2). Isolates were identified as *V. parahaemolyticus* by *tlh* gene amplification (77, 78) and identity confirmed by genome sequencing. Publicly available genomes were acquired from NCBI (https://www.ncbi.nlm.nih.gov/genome/681; August 2019).

DNA was extracted using the Wizard Genomic DNA purification Kit (Promega, Madison WI USA) or by organic extraction (2). Sequencing libraries were prepared as described (79). 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 reads with de-multiplexing and quality filtering prior to analysis. The *de novo* genome assembly were performed using the A5 pipeline (80), and annotations assigned with Prokka1.9 using "*Vibrio"* for the reference database (81). The sequence types were determined using the SRST2 pipeline (82) or using assemblies (83) referencing https://pubmlst.org/vparahaemolyticus/(84). Reference inoviruses were extracted from genomes sequenced using the Pacific Biosciences RSII technology and with Illumina short read error correction as described (4).

V. parahaemolyticus *Phylogenetic Relationships*.

Because all ST36 strains are closely related (clonal) and recombination contributed to the recent lineage divergence, ST36 strain relationships (Fig. 1) and all *V. parahaemolyticus* relationships (Fig. 5) were determined using reference-free, whole genome alignments using kSNP 3.1 (85) and maximum likelihood phylogenetic trees built by RAxML (86) and visualized with iTOL(87). The optional Kchooser script provided in kSNP3.1 determined the optional kmer

size of 19, for each alignment dataset. Whole genome trees were rooted with *Vibrio alginolyticus* ARGOS_108 (hidden after rooting in iTOL). Maximum likelihood trees were inferred using the GTR-GAMMA model of nucleotide substitution, and the convergence criteria autoMRE option for automatically determining a sufficient number of rapid bootstrap replicates. RAxML completed a thorough ML search after optimization on every 5th bootstrapped tree, which were mapped on the ML tree with the best likelihood ("raxmlHPC-PTHREADS-AVX -m GTRGAMMA -f a -N autoMRE -x 12345 -p 12345" options). Gene/features were visualized with EasyFig 2.2.0 (88, 89).

In parallel, homoplastic regions resulting from recombination were identified in genomes by alignment to reference strain 10296 (90) (summary table S2) using Gubbins 2.3.4 (91) and removed, and the relationships of strains determined based on nucleotide SNP alignments generated with Gubbins 2.3.4 (91) (Fig. 2). Phylogenies were inferred as above, except using the GTR-CAT nucleotide substitution model (because the alpha parameter was >10). ML search and optimization, rapid bootstrap criterion, and mapping of bootstraps were done as described above. Unmapped reads were assembled using SPAdes 3.13.1 (92) and annotated using Prokka 1.14(81) (Table S4).

Population structure of ST36 was performed using whole and core SNPs genetic matrixes generated from SNP alignments in the R statistical package LEA (49, 93, 94) (Fig. 2). To determine the number of genetic clusters best explained by the distribution of genomic variation, we explored ancestral populations (*K*) of between 1-20 and used the entropy criterion to evaluate the quality of fit of the statistical model to the data and selected the probable number of populations based on minimal entropy. Once strains were assigned to populations, the least-squares estimates of ancestry proportions were plotted using ggplot2 (95).

Inovirus Identification

ST36 strains were compared to reference 10290 (5) using Breseq (96). Unique prophage were classified as belonging to the family *Inoviridae* proposed subfamily *Protoinoviridae* (27) and assigned unique names (97, 98) (Table 2). All *V. parahaemolyticus* genomes were independently searched against a BLASTn (99) database constructed using the central core (ORF1 – ORF7) of vB Vipa-26. Nucleotide sequences of 745 phage were extracted and to determine if they were unique, maximum likelihood phylogenies using the GTR-GAMMA model for substitution and rapid bootstraps were constructed via RAxML on nucleotide SNPs generated by kSNP (Fig. S6). Rapid bootstraps reached criterion after 350 replicates and were mapped on the tree as described above. The complete prophage assemblies of unique ST36 inoviruses were named from the highest quality assembly as were other prevalent inoviruses (Table S9). Relatedness of select whole phage genomes was visualized using the Genome-BLAST Distance Phylogeny (GBDP) method (100) to conduct pairwise comparisons of the nucleotide sequences under settings recommended for prokaryotic viruses (101). Trees were rooted at the midpoint (28) and visualized using iTOL (87) (Fig. S5).

To determine if phage evolved under selection, a Nei-Gojobori codon-based Z test (102) was performed in MEGA 6 (103). Protein sequences for the phages Vf33 (NC_005948.1) (104, 105), CTXΦ (MF155889.1)(27), and the type strain for *Inoviridae* M13, which infects *Escherichia coli* (GCF_000845205.1) were compared using BLASTp (99).

Oligonucleotide primers were designed for multiplex amplification with the species-specific *tlh* primers (77, 78) where primers annealing within ORF3 within ORF5 identified inovirus presence, and primers annealing within HypD and ORF9 produced different size amplicons diagnostic of Vipa26 and Vipa36 (Table 4). Isolates were screened with 0.2µM primer annealed at 55°C with 1.5 min extension and phage identity confirmed by genome sequencing.

Data Availability

Genome accessions are listed in Tables S1, S7, S8 and 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 (scientific contribution number 2995) was supported by the USDA National Institute of Food and Agriculture Hatch projects NH00609 [accession number 233555], NH00625 [accession number 1004199], NH00658 [accession number 1013479], and NH00698 [accession number 1023165]. 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 and the Fish and Game Department, and Colleen Smith from the 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, VC, SJ, CW collected samples, performed and interpreted analyses, curated databases, and constructed visuals. JM and AM drafted the manuscript and CW and RF completed writing 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 1,025,281 aligned variable nucleotide sites identified in quality ST36 isolates. Isolates are colored to correspond to geographic environmental traceback, where available (Table S1) and where no color indicates unknown or ambiguous origin. Isolates

497 acquired from four Northeast US States as part of this study are named uniquely by a
498 combination of state (Maine [ME], New Hampshire [NH], Massachusetts [MA], and Connecticut
499 [CT]) of reporting and sequential numbers. Symbols next to isolates indicate identity of inovirus
500 content. The ancestral PNW ST36 population is identified by a black bar, whereas seven
501 different lineages associated with translocation events (I – VII) are identified by greyscale bars.
502 Bootstraps are from 250 replicates (criterion reached).

503 **Figure 2. Population structure of ST36 strains.** Coancestry estimates were inferred from
504 SNP matrices using 741 SNPs identified from core non-recombining genomes of ST36 isolates
505 with known environmental source and at least one close relative (see Table S5 for excluded
506 genomes) by LEA (49). Colored bars represent proportion of genetic variation derived from 8
507 ancestral populations (see Fig S3A). Geographic locations are marked by colors below strain
508 name, followed by phage content (symbol) and year of isolation.

509 **Figure 3. Comparisons of inoviruses in *V. parahaemolyticus* ST36.** (A) Alignments of
510 reference inovirus genomes depicting unique and shared content, and % nucleotide identity
511 between Vipa36 (9,721 bp), Vipa26 (10,893 bp) and Vipa10290 (9,414 bp) as compared to f237
512 (8,784 bp). Orthologous ORFs are depicted by the same color except for non-homologous light
513 blue hypothetical proteins. (B) Core inovirus genome (ORFs 1-7, and ORF9-10) single
514 nucleotide polymorphisms (SNPs) were identified by comparison of quality ST36 genomes to
515 the alignment consensus of reference inovirus genomes for Vipa36 (21 genomes), Vipa26 (24
516 genomes), and Vipa10290 (23 genomes) and used to assess spatial distribution of variation
517 between the inoviruses and within each inovirus lineage. SNPs are mapped by location where
518 ORF identities are labeled and colored to match those in (A), and SNP type designated by
519 colored blocks (see Table S6). This indicates few non-synonymous mutations (see text for
520 details) and uneven distribution of variation with ORF3, an intergenic region between ORF4 and

521 ORF5, and the 5' end of ORF5 being identical in all three phage. Only Vipa10290 displayed
522 non-conserved core genome variation among ST36 prophage.

523 **Table 1. Inferred functions of 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}/88-86%)	-	-	Function unknown

524 [1]ORF number corresponds to numbers designated for f237, and comparative analyses between
525 homologous ORFs completed with Vipa26 as a reference.

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

527

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

Strain	Inovirus name	NCBI Accession (Genome coordinates)	Additional ST36 genomes with inovirus	Other STs with related inoviruses
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

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

530 [1]Named for ST43 strain MAVP-71

531

532 **Figure 4. Whole genome phylogenies of representative inoviruses from diverse locations**

533 **and sequence types.** Whole genome phylogenies of select inoviruses from environmental (E)

534 and clinical (C) isolates of *V. parahaemolyticus* labeled by name and year of isolation.

535 Bootstraps criterion met after 350 replicates. Environmental traceback color coding and

536 symbols, representing named phage corresponds to those in Figure 1 where circle is Vipa26,

537 square is Vipa10290, triangle is Vipa36 and grey square is other (Vipa8296). ST36 isolates are
538 in bold.

539 **Figure 5. Maximum-likelihood tree of diverse *V. parahaemolyticus* isolates with their**
540 **inovirus content.** A maximum-likelihood (ML) phylogeny was built on 1,025,281 genome SNPs
541 where 250 bootstraps (criterion met) were mapped onto the best scoring ML tree. Isolates are
542 colored to correspond to geographic region as in Fig. 1, where no color indicates unknown or
543 ambiguous origin. Symbols next to strains indicate unique inovirus content. Outermost label
544 indicates the sequence type (ST) where NF or unlabled = sequence type not found or not
545 known, followed by year of isolation, if known.

546

547 **Table 3. 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

548

549 **Table 4. Multiplex PCR for detection of inoviruses**

Primer	Sequence (5'-3')	Amplicon Size (bp)	
		Host (*V.p.*)	Inovirus
TLH-F2 TLH-R	AGAACTTCATCTTGATGACACTGC GCTACTTTCTAGCATTTTCTCTGC	401	N/A
ST36Phage F2 ST36Phage R2	AGCAACGAAAACGCCTGT ACCGTATCACCAATGGACTGT	N/A	~1000
PhHypDF3 PhORF9R1	AAGTGCTACATGAATGAAAGTGCT TCAATGAAGTATCACGAAATGACTA	N/A	Vipa26: 1440 Vipa36: 854

550 N/A: not an amplification template

551

552 **Literature Cited:**

1. Urquhart EA, Jones SH, Jong WY, 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 11:e0155018.
2. Xu F, Ilyas S, Hall JA, Jones SH, Cooper VS, Whistler CA. 2015. Genetic characterization of clinical and environmental *Vibrio parahaemolyticus* from the Northeast USA reveals emerging resident and non-indigenous pathogen lineages. Front Microbiol 6:272.
3. Baker-Austin C, Trinanes JA, Taylor NG, Hartnell R, Siitonen A, Martinez-Urtaza J. 2013. Emerging *Vibrio* risk at high latitudes in response to ocean warming. Nature Climate Change 3:73-77.
4. Xu F, Gonzalez-Escalona N, Sebra RP, Dress KP, Cooper VS, Jones SH, Whistler CA. 2017. Parallel evolution of two clades of a major Atlantic endemic *Vibrio parahaemolyticus* pathogen lineage by independent acquisition of related pathogenicity islands. Appl Environ Microbiol doi:https://doi.org/10.1101/155234.
5. Martinez-Urtaza J, Baker-Austin C, Jones JL, Newton AE, Gonzalez-Aviles GD, DePaola A. 2013. Spread of Pacific Northwest *Vibrio parahaemolyticus* strain. N Engl J Med 369:1573-1574.
6. Vezzulli L, Grande C, Reid PC, Hélaouët P, Edwards M, Höfle MG, Brettar I, Colwell RR, Pruzzo C. 2016. 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:201609157.
7. Newton AE, Garrett N, Stroika SG, Halpin JL, Turnsek M, Mody RK, Division of Foodborne W, Environmental D. 2014. Notes from the field: Increase in *Vibrio parahaemolyticus* infections associated with consumption of Atlantic coast shellfish—2013. MMWR Morbidity Mortality Weekly Report 63:335-336.
8. 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 27:2891-3.
9. 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 8.
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. 2016. Epidemiological investigation of a foodborne outbreak in Spain associated with U.S. West Coast genotypes of *Vibrio parahaemolyticus*. Springerplus 5:87.
11. CDC. 1999. Outbreak of *Vibrio parahaemolyticus* infection associated with eating raw oysters and clams harvested from Long Island Sound--Connecticut, New Jersey, and New York, 1998. MMWR Morb Mortal Wkly Rep 48:48-51.
12. 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 82:3208-3216.
13. Zahid MS, Udden SM, Faruque AS, Calderwood SB, Mekalanos JJ, Faruque SM. 2008. Effect of phage on the infectivity of *Vibrio cholerae* and emergence of genetic variants. Infect Immun 76:5266-73.
14. 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 103:4652-7.

15. Faruque SM, Bin Naser I, Fujihara K, Diraphat P, Chowdhury N, Kamruzzaman M, Qadri F, Yamasaki S, Ghosh AN, Mekalanos JJ. 2005. Genomic sequence and receptor for the *Vibrio cholerae* phage KSF-1phi: evolutionary divergence among filamentous vibriophages mediating lateral gene transfer. J Bacteriol 187:4095-103.
16. Nanda AM, Thormann K, Frunzke J. 2015. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. J Bacteriol 197:410-9.
17. Brussow H. 2007. Bacteria between protists and phages: from antipredation strategies to the evolution of pathogenicity. Mol Microbiol 65:583-9.
18. Molina-Quiroz RC, Camilli A, Silva-Valenzuela CA. 2023. Role of bacteriophages in the evolution of pathogenic *Vibrios* and lessons for phage therapy. Adv Exp Med Biol 1404:149-173.
19. Hampton HG, Watson BNJ, Fineran PC. 2020. The arms race between bacteria and their phage foes. Nature 577:327-336.
20. Bohannan BJM, Lenski RE. 2000. The relative importance of competition and predation varies with productivity in a model community. Am Nat 156:329-340.
21. Joo J, Gunny M, Cases M, Hudson P, Albert R, Harvill E. 2006. Bacteriophage-mediated competition in *Bordetella* bacteria. Proc Biol Sci 273:1843-8.
22. Ritchie JM, Wagner PL, Acheson DW, 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 69:1059-66.
23. Taylor MJ, Thorne CB. 1963. Transduction of *Bacillus licheniformis* and *Bacillus subtilis* by each of two phages. J Bacteriol 86:452-61.
24. Romig WR, Brodetsky AM. 1961. Isolation and preliminary characterization of bacteriophages for *Bacillus subtilis*. J Bacteriol 82:135-41.
25. Buckling A, Rainey PB. 2002. Antagonistic coevolution between a bacterium and a bacteriophage. Proc Biol Sci 269:931-6.
26. Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F, Mira A. 2009. Explaining microbial population genomics through phage predation. Nat Rev Microbiol 7:828-36.
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. 2019. Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. Nat Microbiol doi:10.1038/s41564-019-0510-x.
28. Breitbart M, Rohwer F. 2005. Here a virus, there a virus, everywhere the same virus? Trends Microbiol 13:278-84.
29. Askora A, Yamada T. 2015. Two different evolutionary lines of filamentous phages in *Ralstonia solanacearum*: their effects on bacterial virulence. Front Genet 6:217.
30. Yen M, Camilli A. 2017. Mechanisms of the evolutionary arms race between *Vibrio cholerae* and Vibriophage clinical isolates. Int Microbiol 20:116-120.
31. Richards GP, Chintapenta LK, Watson MA, Abbott AG, Ozbay G, Uknalis J, Oyelade AA, Parveen S. 2019. Bacteriophages against pathogenic *Vibrios* in Delaware Bay oysters (*Crassostrea virginica*) during a period of high levels of pathogenic *Vibrio parahaemolyticus*. Food Environ Virol 11:101-112.
32. Onarinde BA, Dixon RA. 2018. Prospects for biocontrol of *Vibrio parahaemolyticus* contamination in blue mussels (*Mytilus edulus*)-a year-long study. Front Microbiol 9:1043.
33. Zhang H, Yang Z, Zhou Y, Bao H, Wang R, Li T, Pang M, Sun L, Zhou X. 2018. Application of a phage in decontaminating *Vibrio parahaemolyticus* in oysters. Int J Food Microbiol 275:24-31.

34. Wang W, Li M, Li Y. 2015. Intervention strategies for reducing *Vibrio parahaemolyticus* in seafood: a review. J Food Sci 80:R10-9.
35. Cao Y, Zhang Y, Lan W, Sun X. 2021. Characterization of vB_VpaP_MGD2, a newly isolated bacteriophage with biocontrol potential against multidrug-resistant *Vibrio parahaemolyticus*. Arch Virol 166:413-426.
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. 2022. Coastal water bacteriophages infect various sets of *Vibrio parahaemolyticus* sequence types. Front Microbiol 13:1041942.
37. Ding T, Sun H, Pan Q, Zhao F, Zhang Z, Ren H. 2020. Isolation and characterization of *Vibrio parahaemolyticus* bacteriophage vB_VpaS_PG07. Virus Res 286:198080.
38. Faruque SM, Naser IB, Islam MJ, Faruque AS, 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 102:1702-7.
39. Nasu H, Iida T, Sugahara T, Yamaichi Y, Park KS, Yokoyama K, Makino K, Shinagawa H, Honda T. 2000. A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:K6 strains. J Clin Microbiol 38:2156-61.
40. 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 7:477-8.
41. Iida T, Makino K, Nasu H, Yokoyama K, Tagomori K, Hattori A, Okuno T, Shinagawa H, Honda T. 2002. Filamentous bacteriophages of vibrios are integrated into the dif-like site of the host chromosome. J Bacteriol 184:4933-5.
42. Chan B, Miyamoto H, Taniguchi H, Yoshida S. 2002. 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 46:565-9.
43. Garin-Fernandez A, Wichels A. 2020. Looking for the hidden: characterization of lysogenic phages in potential pathogenic *Vibrio* species from the North Sea. Mar Genomics 51:100725.
44. Garin-Fernandez A, Glockner FO, Wichels A. 2020. Genomic characterization of filamentous phage vB_VpaI_VP-3218, an inducible prophage of *Vibrio parahaemolyticus*. Mar Genomics 53:100767.
45. Abanto M, Gavilan RG, Baker-Austin C, Gonzalez-Escalona N, Martinez-Urtaza J. 2020. Global expansion of Pacific Northwest *Vibrio parahaemolyticus* sequence type 36. Emerg Infect Dis 26:323-326.
46. Banerjee SK, Kearney AK, Nadon CA, Peterson C-L, Tyler K, Bakouche L, Clark CG, Hoang L, Gilmour MW, Farber JM. 2014. Phenotypic and genotypic characterization of Canadian clinical isolates of *Vibrio parahaemolyticus* collected from 2000 to 2009. Journal of Clinical Microbiology 52:1081-1088.
47. Ronholm J, Petronella N, Kenwell R, Banerjee S. 2015. Draft whole-genome sequences of 14 *Vibrio parahaemolyticus* clinical isolates with an ambiguous K serogroup. Genome Announc 3.
48. Ronholm J, Petronella N, Chew Leung C, Pightling AW, Banerjee SK. 2016. Genomic features of environmental and clinical *Vibrio parahaemolyticus* isolates lacking recognized virulence factor are dissimilar. Appl Environ Microbiol 82:1102-1113.
49. Frichot E, François O. 2015. LEA: an R package for landscape and ecological association studies. Methods in Ecology and Evolution 6:925-929.
50. Marcinkiewicz AL. 2016. Bacterial and phage interactions influencing *Vibrio parahaemolyticus* ecology. Master of Science in Microbiology. University of New Hampshire.

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. 2022. A filamentous bacteriophage protein inhibits Type IV pili to prevent superinfection of *Pseudomonas aeruginosa*. mBio 13:e0244121.
52. Ismail MH, Michie KA, Goh YF, Noorian P, Kjelleberg S, Duggin IG, McDougald D, Rice SA. 2021. The repressor C protein, Pf4r, controls superinfection of *Pseudomonas aeruginosa* PAO1 by the Pf4 filamentous phage and regulates host gene expression. Viruses 13.
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. 2022. *Vibrio parahaemolyticus* is associated with diarrhea cases in Mexico, with a dominance of pandemic O3:K6 clones. Int J Environ Res Public Health 19.
54. Rosales D, Ellett A, Jacobs J, Ozbay G, Parveen S, Pitula J. 2022. 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 88:e0035622.
55. 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 Nino. Epidemiology 19:829-37.
56. McLaughlin JB, DePaola A, Bopp CA, Martinek KA, Napolilli NP, Allison CG, Murray SL, Thompson EC, Bird MM, Middaugh JP. 2005. Outbreak of Vibrio parahaemolyticus gastroenteritis associated with Alaskan oysters. N Engl J Med 353:1463-70.
57. Konrad S, Paduraru P, Romero-Barrios P, Henderson SB, Galanis E. 2017. Remote sensing measurements of sea surface temperature as an indicator of *Vibrio parahaemolyticus* in oyster meat and human illnesses. Environ Health 16:92.
58. Galanis E, Otterstatter M, Taylor M. 2020. Measuring the impact of sea surface temperature on the human incidence of *Vibrio* sp. infection in British Columbia, Canada, 1992-2017. Environ Health 19:58.
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. 2018. Outbreak of *Vibrio parahaemolyticus* associated with consumption of raw oysters in Canada, 2015. Foodborne Pathog Dis 15:554-559.
60. Hartwick MA, Urquhart EA, Whistler CA, Cooper VS, Naumova EN, Jones SH. 2019. Forecasting seasonal *Vibrio parahaemolyticus* concentrations in New England shellfish. Int J Environ Res Public Health 16.
61. Mai-Prochnow A, Hui JG, 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 39:465-87.
62. Das B, Bischerour J, Barre FX. 2011. VGJphi integration and excision mechanisms contribute to the genetic diversity of *Vibrio cholerae* epidemic strains. Proc Natl Acad Sci U S A 108:2516-21.
63. Means J. 2019. Filamentous bacteriophage associated with shaping community struture and fitness of invasive *Vibrio parahaemolyticus* ST36 Master's Theses and Capstones. University of New Hampshire.
64. 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. 2021. Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. Science 374:488-492.

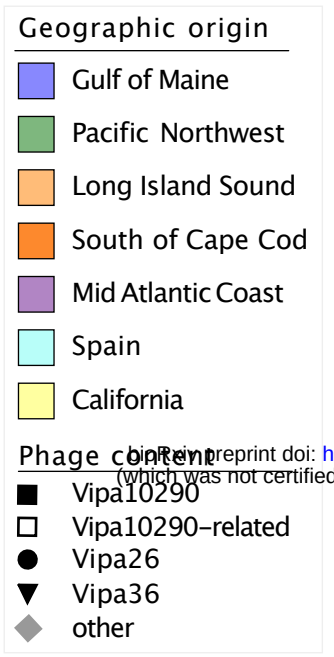
65. Secor PR, Sweere JM, Michaels LA, Malkovskiy AV, Lazzareschi D, Katznelson E, Rajadas J, Birnbaum ME, Arrigoni A, Braun KR, Evanko SP, Stevens DA, Kaminsky W, Singh PK, Parks WC, Bollyky PL. 2015. Filamentous bacteriophage promote biofilm assembly and function. Cell Host Microbe 18:549-59.
66. Martinez-Urtaza 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 6:994-1006.
67. Doni L, Oliveri C, Lasa A, Di Cesare A, Petrin S, Martinez-Urtaza J, Coman F, Richardson A, Vezzulli L. 2023. Large-scale impact of the 2016 marine heatwave on the plankton-associated microbial communities of the Great Barrier Reef (Australia). Mar Pollut Bull 188:114685.
68. Xu Y, Yang L, Wang Y, Zhu Z, Yan J, Qin S, Chen L. 2022. Prophage-encoded gene VpaChn25_0734 amplifies ecological persistence of *Vibrio parahaemolyticus* CHN25. Curr Genet 68:267-287.
69. 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. 2019. Bacteriophage trigger antiviral immunity and prevent clearance of bacterial infection. Science 363.
70. Erken M, Weitere M, Kjelleberg S, McDougald D. 2011. In situ grazing resistance of *Vibrio cholerae* in the marine environment. FEMS Microbiol Ecol 76:504-12.
71. Paff ML, Nuismer SL, Ellington A, Molineux IJ, Bull JJ. 2016. Virus wars: using one virus to block the spread of another. PeerJ 4:e2166.
72. Lin A, Jimenez J, Derr J, Vera P, Manapat ML, Esvelt KM, Villanueva L, Liu DR, Chen IA. 2011. Inhibition of bacterial conjugation by phage M13 and its protein g3p: quantitative analysis and model. PLoS One 6:e19991.
73. Molina-Quiroz RC, Dalia TN, Camilli A, Dalia AB, Silva-Valenzuela CA. 2020. Prophage-dependent neighbor predation fosters horizontal gene transfer by natural transformation. mSphere 5.
74. Matthey N, Stutzmann S, Stoudmann C, Guex N, Iseli C, Blokesch M. 2019. Neighbor predation linked to natural competence fosters the transfer of large genomic regions in *Vibrio cholerae*. Elife 8.
75. Dalia AB, McDonough E, Camilli A. 2014. Multiplex genome editing by natural transformation. Proc Natl Acad Sci U S A 111:8937-42.
76. Guerrero A, Gomez-Gil B, Lizarraga-Partida ML. 2021. Genomic stability among O3:K6 *V. parahaemolyticus* pandemic strains isolated between 1996 to 2012 in American countries. BMC Genom Data 22:38.
77. Panicker G, Call DR, Krug MJ, Bej AK. 2004. Detection of pathogenic *Vibrio spp.* in shellfish by using multiplex PCR and DNA microarrays. Appl Environ Microbiol 70:7436-7444.
78. 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*. Journal of clinical microbiology 53:1864-1872.
79. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. 2015. Inexpensive multiplexed library preparation for megabase-sized genomes. PloS one 10:e0128036.
80. Tritt A, Eisen JA, Facciotti MT, Darling AE. 2012. A5. An integrated pipeline for *de novo* assembly of microbial genomes. PLoS ONE 7:e42304 DOI: 10.1371/journal.pone.0042304.

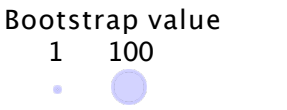
81. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics:DOI:10.1093/bioinformatics/btu153.
82. Inouye M, Conway TC, Zobel J, Holt KE. 2012. Short read sequence typing (SRST): multi-locus sequence types from short reads. BMC Genomics 13:338.
83. Jolley KA, Maiden MC. 2014. Using MLST to study bacterial variation: prospects in the genomic era. Future microbiology 9:623-630.
84. 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 6:90.
85. Gardner SN, Slezak T, Hall BG. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics 31:2877-8.
86. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.
87. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256-W259.
88. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147.
89. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27:1009-10.
90. T. S. 2015. Snippy: fast bacterial variant calling from NGS reads [Internet]. Available from https://github.com/tseemann/snippy Accessed
91. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2014. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic acids research:gku1196.
92. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes De Novo Assembler. Curr Protoc Bioinformatics 70:e102.
93. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. Gigascience 10.
94. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, Harris SR. 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom 2:e000056.
95. Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer.
96. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol 1151:165-88.
97. 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. 2016. Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. Arch Virol 161:1095-9.
98. Adriaenssens E, Brister JR. 2017. How to Name and Classify Your Phage: An Informal Guide. Viruses 9.
99. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403-10.
100. Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60.
101. Meier-Kolthoff JP, Goker M. 2017. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics 33:3396-3404.

855 102. Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous
856 and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418-26.
857 103. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular
858 Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725-9.
859 104. Taniguchi H, Sato K, Ogawa M, Udou T, Mizuguchi Y. 1984. Isolation and
860 characterization of a filamentous phage, Vf33, specific for *Vibrio parahaemolyticus*.
861 Microbiology and immunology 28:327-337.
862 105. Chang B, Yoshida S, Miyamoto H, Ogawa M, Horikawa K, Ogata K, Nishibuchi M,
863 Taniguchi H. 2000. A unique and common restriction fragment pattern of the nucleotide
864 sequences homologous to the genome of vf33, a filamentous bacteriophage, in
865 pandemic strains of *Vibrio parahaemolyticus* O3:K6 O4:K68, and O1:K untypeable.
866 FEMS Microbiol Lett 192:231-6.

867

Figure1





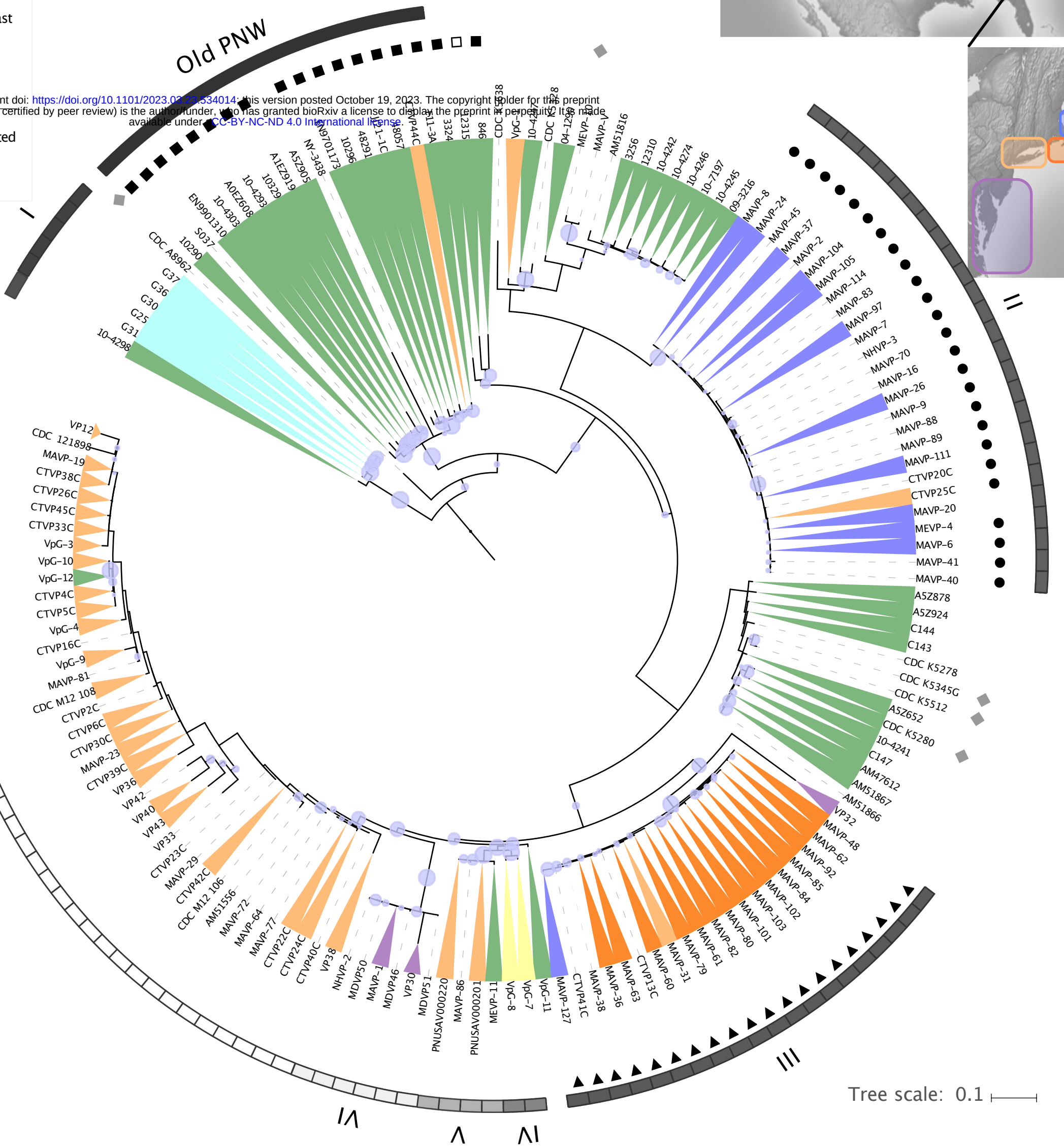


Figure2

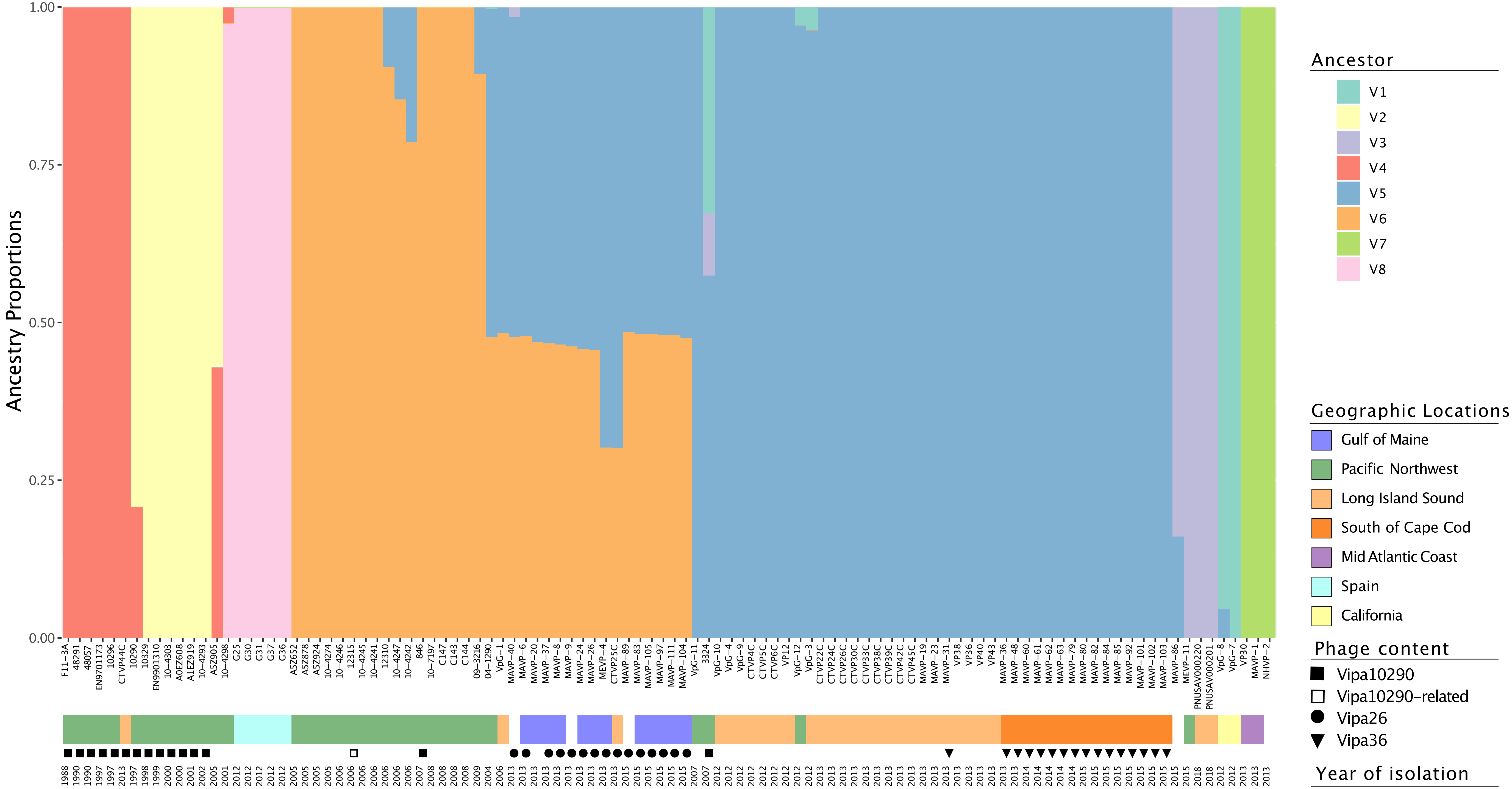

Ancestry Proportions
0.00
0.25
0.50
0.75
1.00
F11−3A
48291
48057
EN9701173
10296
CTVP44C
10290
10329
EN9901310
10−4303
A0EZ608
A1EZ919
10−4293
A5Z905
10−4298
G25
G30
G31
G37
G36
A5Z652
A5Z878
A5Z924
10−4274
10−4246
12315
10−4245
10−4241
12310
10−4247
10−4242
846
10−7197
C147
C143
C144
09−3216
04−1290
VpG−1
MAVP−40
MAVP−6
MAVP−20
MAVP−37
MAVP−8
MAVP−9
MAVP−24
MAVP−26
MEVP−4
CTVP25C
MAVP−89
MAVP−83
MAVP−105
MAVP−97
MAVP−111
MAVP−104
VpG−11
3324
VpG−10
VpG−4
VpG−9
CTVP4C
CTVP5C
CTVP6C
VP12
VpG−12
VpG−3
CTVP22C
CTVP24C
CTVP26C
CTVP30C
CTVP33C
CTVP38C
CTVP39C
CTVP42C
CTVP45C
MAVP−19
MAVP−23
MAVP−31
VP38
VP36
VP40
VP43
MAVP−36
MAVP−48
MAVP−60
MAVP−61
MAVP−62
MAVP−63
MAVP−79
MAVP−80
MAVP−82
MAVP−84
MAVP−85
MAVP−92
MAVP−101
MAVP−102
MAVP−103
MAVP−86
MEVP−11
PNUSAV000220
PNUSAV000201
VpG−8
VpG−7
VP30
MAVP−1
NHVP−2
1988 1990 1990 1997 1997 2013 1997 1998 1999 2000 2000 2001 2002 2005 2001 2012 2012 2012 2012 2012 2005 2005 2005 2005 2006 2006 2006 2006 2006 2006 2006 2007 2008 2008 2008 2008 2009 2004 2006 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 2015 2015 2015 2015 2015 2015 2007 2007 2012 2012 2012 2012 2012 2012 2012 2012 2012 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 2013 2014 2014 2014 2014 2014 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 2018 2018 2012 2012 2013 2013 2013
Ancestor
V1
V2
V3
V4
V5
V6
V7
V8
Geographic Locations
Gulf of Maine
Pacific Northwest
Long Island Sound
South of Cape Cod
Mid Atlantic Coast
Spain
California
Phage content
Vipa10290
Vipa10290−related
Vipa26
Vipa36
Year of isolation

Figure 3

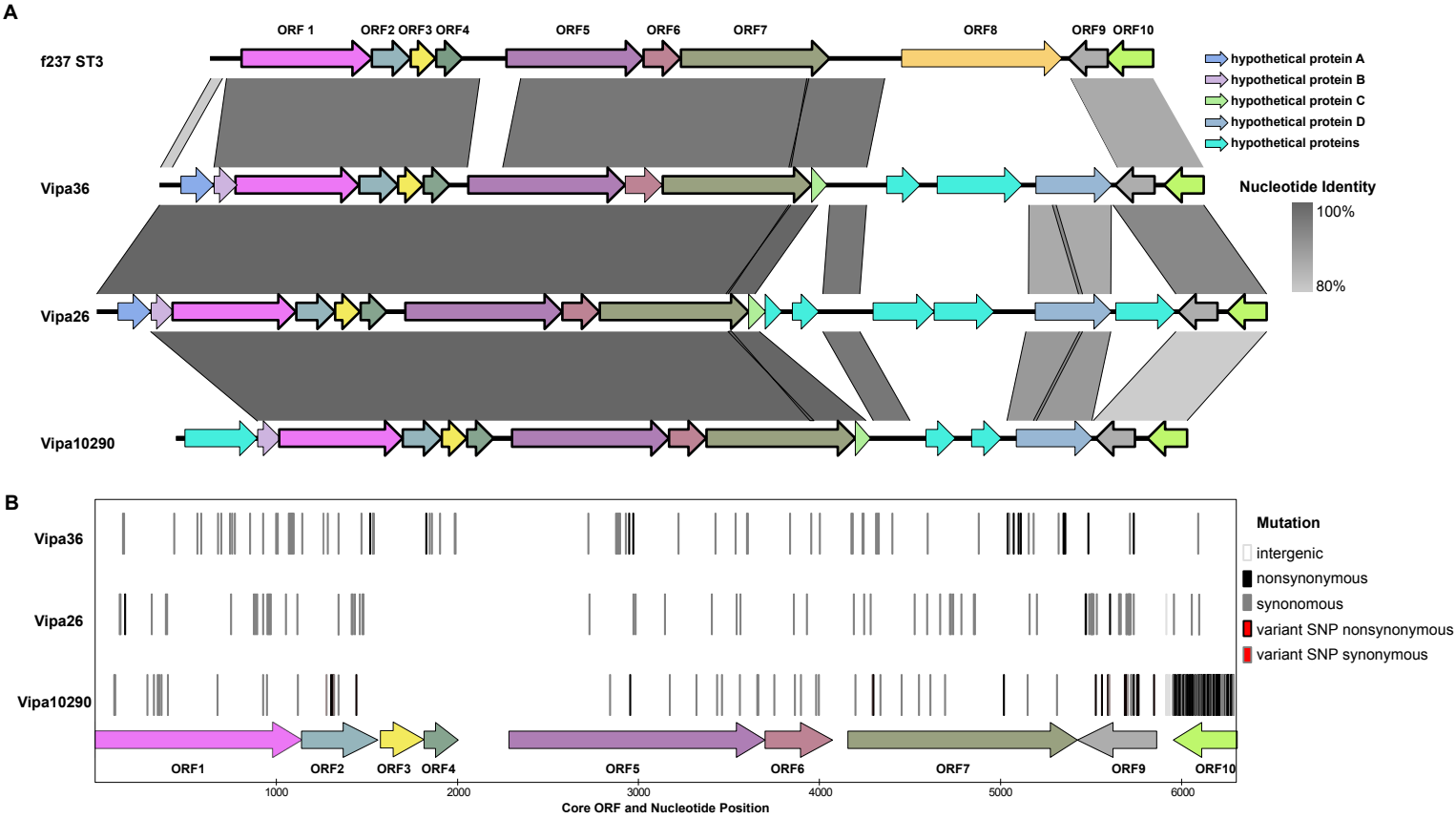

A
ORF 1
ORF2 ORF3 ORF4
ORF5
ORF6
ORF7
ORF8
ORF9 ORF10
f237 ST3
Vipa36
Vipa26
Vipa10290
hypothetical protein A
hypothetical protein B
hypothetical protein C
hypothetical protein D
hypothetical proteins
Nucleotide Identity
100%
80%
B
Vipa36
Vipa26
Vipa10290
ORF1
ORF2
ORF3
ORF4
ORF5
ORF6
ORF7
ORF9
ORF10
1000
2000
3000
4000
5000
6000
Core ORF and Nucleotide Position
Mutation
intergenic
nonsynonymous
synonomous
variant SNP nonsynonymous
variant SNP synonymous

Figure4

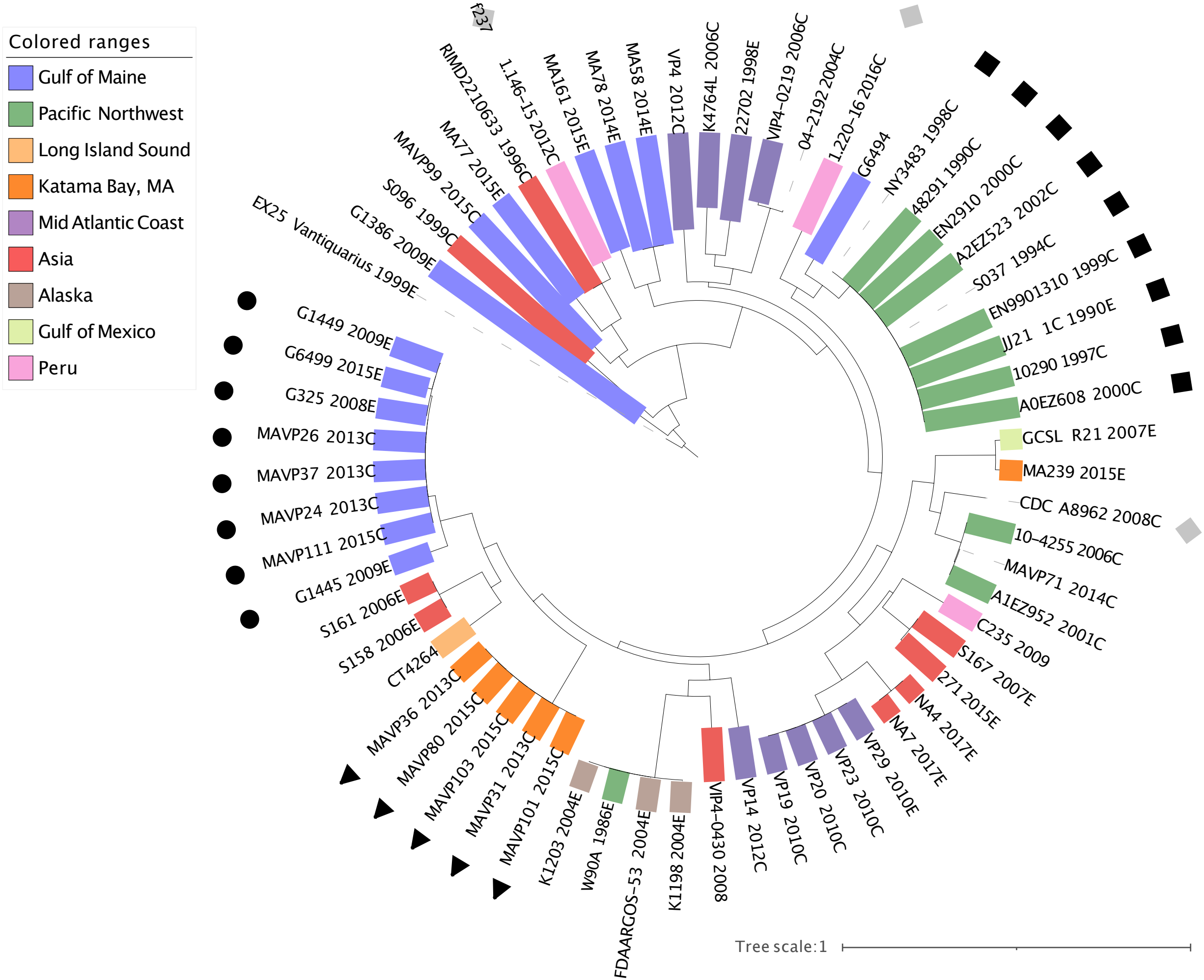


Figure5