Fall 2020

IDENTIFICATION OF TYPE THREE SECRETION SYSTEM EFFECTORS LINKED TO CLINICAL AND ENVIRONMENTAL PREVALENCE OF VIBRIO PARAHAEMOLYTICUS

Sarah Eggert
University of New Hampshire, Durham

IDENTIFICATION OF TYPE THREE SECRETION SYSTEM EFFECTORS LINKED TO CLINICAL AND ENVIRONMENTAL PREVALENCE OF *VIBRIO PARAHAEMOLYTICUS*

BY

SARAH E. EGGERT

B.S. IN BIOLOGICAL SCIENCES, CELLULAR AND MOLECULAR BIOLOGY,

CALIFORNIA STATE UNIVERSITY, CHICO, 2015

THESIS

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirements for the Degree of

Master of Science

in

Microbiology

September 2020

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

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

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

Cheryl Andam, Assistant Professor, Molecular, Cellular, and Biomedical Sciences

On October 1st 2019

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

DEDICATION

To my love. None of this would have been possible without your continuous adventure, inspiration, and light.

ACKNOWLEDGEMENTS

To my PI, Cheryl Whistler, thank you for all the beyond-textbook knowledge throughout the past few years and always continuing to push me as a scientist.

To my committee members, thank you for your expertise and enthusiasm, especially in the time crunch at the very end.

A special thank you my fellow UNH graduate students, especially those I had the pleasure of teaching with. You always kept a smile on my face despite our stressful times.

To the MCBS faculty and staff, especially Randi, Amy, Tim, Jobriah, Paul and Pam, thank you for all your wisdom, encouragement, laughs, and unending help.

To Carter, my best friend and other half, I cannot thank you enough for this adventure. Thank you for the continuous energy you pour into our never-ending journey, and for always keeping me anchored.

Thank you to my parents, sister and her family, brother, and family for always supporting me through the years, even when my dreams led me across the country.

Lastly, I would like to thank my funding sources, SeaGrant New Hampshire and UNH Agricultural Experiment Station for making this research possible.

TABLE OF CONTENTS

ABSTRACT

IDENTIFICATION OF TYPE THREE SECRETION SYSTEM EFFECTORS LINKED TO CLINICAL AND ENVIRONMENTAL PREVALENCE OF *VIBRIO PARAHAEMOLYTICUS*

By

Sarah Eggert

University of New Hampshire

Vibrio parahaemolyticus is an emergent human pathogen that is the leading cause of seafood-borne bacterial infections in the United States and worldwide. Recently, clinical prevalence of *V. parahaemolyticus* has increased in the United States, especially from the North Atlantic Ocean. The majority of clinical isolates of *V. parahaemolyticus* harbor a hemolysin gene (*tdh* and/or *trh*) and a Type III Secretion System (T3SS)-containing mobile Vibrio Pathogenicity Island (VPaI). The VPaIs are evolutionary related yet distinct. These VPaIs include VPaIα (*tdh*$^+$), VPaIβ (*trh*$^+$), and a mosaic VPaIγ (*tdh*$^+$/*trh*$^+$). Strains harboring VPaIα cause the most infections globally, whose major effectors are known. In the United States, in particular from the North Atlantic, human infections with *V. parahaemolyticus* are predominantly caused by strains harboring the relatively uncharacterized VPaIγ. There are four major VPaIγ lineages in *V. parahaemolyticus* strains of North Atlantic, identified by the *tdh* allele they harbor. Strains harboring *tdh*3-VPaIγ are isolated most clinically, though strains harboring *tdh*5-VPaIγ are environmentally predominant. Using a dual approach of bioinformatics and bioassays, in this study we identify five novel T3SS2 effectors on VPaIγ. We provide evidence that suggests a *V. parahaemolyticus* population-based contribution towards its pathogenicity and environmental prevalence. Patterns of recombination in toxin orthologs from *tdh*3-VPaIγ and *tdh*5-VPaIγ indicate

that the surrounding population is a major contributor to the content of these VPaIs. The divergence seen in these orthologs were assessed for contribution towards their effectiveness in pathogenesis or environmental fitness. Our predatory grazing assay suggests that one effector may have evolved for survival against eukaryotic predators in the environment. Utilizing bioassays to assess toxicity and virulence, we found evidence that four of the VPaIγ orthologs likely contribute to the *tdh*3-VPaIγ clinical prevalence. This work lays the foundation for understanding VPaIγ effector fitness benefits in relation to their environment and population.

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Vibrio parahaemolyticus

Vibrio parahaemolyticus is a gram-negative, halophilic bacterium found in brackish water throughout the world. *V. parahaemolyticus* usually exists in a free-swimming state throughout the water column in marine, estuarine, and coastal environments, but it also attaches to animate and inert surfaces underwater such as suspended solids, sediments, shellfish, phytoplankton and zooplankton, fish, and crustaceans [1–5]. *V. parahaemolyticus* is an emergent human pathogen and is the leading seafood-borne bacterial pathogen in the United States and worldwide. Food poisoning caused by *V. parahaemolyticus* is generally caused by consuming raw, mishandled, or undercooked seafood, including crab, shrimp, shellfish, lobster, fish, and oysters [6,7]. Due to their filter-feeding mechanisms, shellfish, especially oysters, can concentrate bacteria, including *V. parahaemolyticus*, up to 100-fold higher than the surrounding water [6,8–11].

Pathogenic *Vibrio* spp. and their epidemiology

The genus *Vibrio*, collectively called Vibrios, belong to the *Gammaproteobacteria* class of bacteria and are members of the *Vibrionaceae* family [4]. They are ubiquitous in aquatic environments, and though mostly harmless, some species, or more accurately, some strains of these species, cause disease in humans and other animals [4]. Vibriosis is a human illness caused by Vibrios that do not harbor the cholera toxin. Many bacteria of the *Vibrio* genus can cause vibriosis including most notably *Vibrio vulnificus*, *Vibrio alginolyticus* and *V. parahaemolyticus* [4,8,12–14], with the latter three species contributing to the highest amount of human illnesses in the United

States [15]. Historically, *Vibrio cholerae* is the most well-known pathogenic *Vibrio* species since it is the causal agent of the disease cholera [16,17]. Cholera is associated with contaminated water and poor sanitization, causes severe diarrhea and dehydration, and has caused seven pandemics since 1817 [4]. The Centers for Disease Control and Prevention (CDC) monitors *V. cholerae* infections separately from other vibriosis cases due to it high infection rate and death toll [16,17].

V. parahaemolyticus is estimated to cause 45,000 of the 80,000 vibriosis illnesses each year in the United States,[15] and is the leading cause of seafood-borne bacterial infections in the United States and worldwide [15,18,19]. Infection with *V. parahaemolyticus* can take on three forms: gastroenteritis, wound infections, and, rarely, lethal septicemia [18]. Wound infections occur through the exposure of open wounds to seawater containing *V. parahaemolyticus* and are most common among fishermen followed by those partaking in recreational aquatic activities [6,8,20–22]. These wound infections can lead to life-threatening septicemia to immunocompromised individuals [6,23] and require hospitalization and in rare cases can lead to death [2,8,18,24] or necrotizing fasciitis resulting in widespread tissue damage [20,21,25].

The most common form of *V. parahaemolyticus* infection is gastroenteritis, caused by consumption of raw, undercooked, or mishandled seafood that is contaminated with pathogenic strains [1,6,7,9,10,20,22]. *V. parahaemolyticus*-induced gastroenteritis is typically mild and self-limiting, lasting for 2–10 days [6,7]. Symptoms include acute dysentery, abdominal pain, watery diarrhea, nausea, vomiting, fever, and chills [6,7,26–28]. Pathological changes can vary from mild destruction of the jejunum and ileum, and gastric inflammation, to severe cases including internal organ damage (liver, spleen, lung congestion, etc.) [6]. In rare cases, septicemia can result if *V. parahaemolyticus* crosses the intestinal epithelium [25]. Due to the self-limiting nature of gastroenteritis, infection with *V. parahaemolyticus* is often under-reported and most commonly treated at home with oral

rehydration [10,18]. If reported, antibiotics may be prescribed [6]. There have been very few studies on non-human *V. parahaemolyticus* infections. Recent studies reported the damage to the shrimp aquaculture industry from *V. parahaemolyticus* causing Acute Hepatopancreatic Necrosis Disease (AHPND), also known as Early Mortality Syndrome (EMS), in penaeid shrimp [29–31].

Human infection with *V. parahaemolyticus*

V. parahaemolyticus is ubiquitous worldwide in many diverse populations, but only a minority of members within these populations can infect and cause disease [3,19,32–35]. *V. parahaemolyticus* infections typically occur seasonally during summer months (from June to October in the northern hemisphere) [6] when surface water is warmest, facilitating total *V. parahaemolyticus* populations to rise, including pathogen populations [36]. However, *V. parahaemolyticus* infections still occur in cooler water and where there is a low abundance of *V. parahaemolyticus* [19,37,38]. Recurrent infections and outbreaks consistently occur in regions where surface water remains cool throughout the year, in particular, the Pacific Northwest [22,35,39–41]. *V. parahaemolyticus* infections occur worldwide, predominantly in countries with high levels of seafood consumption [1].

Infections of *V. parahaemolyticus* in New England

Vibriosis cases in the United States have increased 120% from 1996 to 2014 (most current data available from CDC as of May 24, 2020 [43]), with *V. parahaemolyticus* responsible for 60-75% of cases [42,43]. This increase in reported cases also coincided with a drastic shift in regional occurrence. Warmer regions, such as the Gulf of Mexico, shifted from 48% of vibriosis infections to 36% [42]. Interestingly, the cooler Northeast Atlantic coast region increased in vibriosis infections

from 3% in 1996 to 22% of vibriosis infections in 2014 [42]. Until recently, there was a long-held belief based on the rare occurrence of illnesses, corroborated by low environmental abundance of hemolysin-harboring members in the environment, that the environmental conditions of the Northeastern US did not sustain pathogenic populations of *V. parahaemolyticus* [33,44]. Except for a large multi-state outbreak in 1998 from a non-indigenous strain of *V. parahaemolyticus* originating from Southeast Asia in oysters harvested from Long Island Sound, *V. parahaemolyticus* has caused few infections associated with shellfish harvested from the cooler waters of the Northeastern US [22,26,45–47]. However, infections associated with Atlantic shellfish suddenly increased. This coincided with warmer than usual ocean temperatures in the region, known to correlate with higher densities of *V. parahaemolyticus* in the water column, and invasion of the Pacific Northwest endemic pathogenic strain Sequence Type (ST) 36 [8,22,48,49]. During this transitional period two outbreaks occurred, one in New York in 2012, and the second in New York, Connecticut and Massachusetts in 2013 [49]. Over the past decade, infection rates have continued to rise in the United States despite the implementation of control measures, with the Northeast included. The cause of this rise is unknown, but one hypothesis attributes this rise to the changing climate on pathogen abundance and distribution [50].

Current molecular detection of pathogenic *V. parahaemolyticus*

Environmental surveillance/enumeration based on the conditionally adopted FDA BAM [51] for *V. parahaemolyticus* includes quantifying the overall abundance of *V. parahaemolyticus* using the thermolabile hemolysin (*tlh*) gene present in all strains, as well as the presence of the two hemolysin genes *tdh* and *trh*, discussed further below, to quantify the number of "pathogenic" *V. parahaemolyticus*. This "pathogenic" *V. parahaemolyticus* detection method does not provide an

accurate measure of the number of pathogenic *V. parahaemolyticus* nor the degree of pathogenicity since 1) not all pathogenic *V. parahaemolyticus* harbor either *tdh* and/or *trh*, 2) not all *V. parahaemolyticus* that harbor these hemolysins are pathogenic, and 3) *V. parahaemolyticus* confers varying degrees of virulence despite the same *tdh*/*trh* profile, as inferred by clinical prevalence [11,52–54]. There is a significant gap in accurate detection of pathogenic strains, and it is important to study *V. parahaemolyticus* populations to develop better detection tools for accurate surveillance of the most important pathogenic lineages.

Bacterial Evolution

Bacteria can adapt to their environment under natural selection through the process of genome evolution. This process of changing the content and organization of the bacterial genetic information can occur through four general mechanisms: point mutations and gene conversions, rearrangements (e.g., inversion or translocation), deletions of DNA, and insertions of foreign DNA [e.g., plasmid integration, horizontal gene transfer (HGT), and transposition] [55,56]. Gene loss and recombination of acquired genes are primary forces by which bacteria adapt to new environments, and these are critical forces on the evolution of microbial pathogens [32,55,57–62]. Bacteria can acquire accessory DNA such as plasmids, bacteriophages, transposons, integrative and conjugative elements (ICEs), and genomic islands (GEIs), through the process of HGT [56]. HGT is the transfer of genetic material from one organism to another non-decedent organism [63]. They can also acquire homologous DNA harboring unique nucleotide variation by HGT, which can recombine with and replace shared genetic content. These acquisitions can provide a fitness advantage beneficial to bacteria under certain conditions such as adaptation to changing environmental conditions, enhancing virulence, or immune evasion [32,34,56,57,64–66].

For HGT to occur efficiently, many bacteria can utilize conjugation machinery to either donate or receive DNA directly from other bacteria or receive novel DNA by phage transduction. *Vibrio* spp. have a rare mechanism, termed natural transformation, that enables them to uptake DNA from their surrounding environment instead of direct transfer from other bacteria or phage [67–69]. When grown in the presence of chitin-based substrates, such as those abundant in marine environments, *Vibrio* spp. turn on machinery to take up DNA from their environment [11,34]. Natural transformation occurs in environmental and pathogenic bacteria, including *Acinetobacter* spp., *Bacillus subtilis*, *Streptococcus pneumoniae*, *Neisseria gonorrhoeae*, *Pseudomonas stutzeri*, *Ralstonia solanacearum,* and *Haemophilus influenzae* [56].

Identification of potential virulence determinants of pathogenic *V. parahaemolyticus*

Tdh/Trh

Identification of the virulence effectors important in human gastrointestinal infection of *V. parahaemolyticus* has been an active area of research for years [8,9,70]. Initially, many researchers believed that the hemolysins, the thermostable direct hemolysin (Tdh), and thermostable related hemolysin (Trh), were indicators for the virulence potential of *V. parahaemolyticus* since most enteropathogenic *V. parahaemolyticus* harbored one or both of those genes, but these genes are not essential for human disease [6,8,71–78].

In 1995 during the initial spread of the pandemic strain of *V. parahaemolyticus*, pathogenic strains were identified by ß-hemolysis on specialized Wagatsuma Blood Agar known as Kanagawa Phenomenon (KP) [73]. At that time, 95% clinical isolates of *V. parahaemolyticus* were KP+ (ß-hemolytic), while 98–99% of the environmental samples were KP- (non-hemolytic) [1,6,75,79,80]. This KP+ hemolysis is caused by the hemolysin Tdh [6,7,24,72,81,82].

In addition to hemolysis, Tdh can also induce cytotoxicity, cardiotoxicity, and enterotoxicity [7,82]. Tdh is termed thermostable direct hemolysin due to its stability at 100°C [7,83]. Tdh is a pore-forming protein that alters ion flux in human cells, causing cytotoxicity. Tdh works in three primary steps, it 1) binds to erythrocyte membrane or host cells 2) forms a ~2nm diameter transmembrane pore which 3) disrupts the ion flux of Ca^{2+} and Cl^{-} across the cell membrane, increasing the osmotic pressure of the cell exceeding the upper limit for the cell beyond self-regulation, resulting in cell expansion and death [6,7]. So far, there are seven distinct *tdh* alleles which share a nucleotide sequence identity >96.7% [11,82,83].

Initially during the spread of pathogenic *V. parahaemolyticus* worldwide, the Wagatsuma agar was a good screen for diarrhea-inducing *V. parahaemolyticus* during the original outbreak of the pandemic strain, but then more KP- *V. parahaemolyticus* caused human sickness. The majority of these KP- clinical isolates encode Trh [84–87]. Trh is related to Tdh and plays a similar role as Tdh in tissue pathoglogy [1,88]. Trh is a pore-forming hemolysin and causes similar levels of hemolysis as Tdh *in vitro* [6]. It can form hemolytic zones on usual blood agar plates, but not on Wagatsuma's medium, and thus was used to identify KP- strains [83]. Wagatsuma's medium was specially formulated to produce distinctive ß-hemolysis in *tdh*-harboring *V. parahaemolyticus* due to the specificity of erythrocyte source, pH, salt content, and carbohydrate content [89,90]. Two *trh* alleles have been described so far in various *Vibrio* species [75].

Flaws of *tdh*/*trh* as markers of virulence

Unfortunately, the use of *tdh* and *trh* as pathogen screening markers had major flaws. KP was only effective in identifying strains harboring *tdh*, and some clinical isolates did not contain either of *tdh* or *trh* [8,19,91,92]. Also, deletion of these markers does not ameliorate disease in an animal

model, and environmentally, a high proportion of *tdh* and/or *trh* does not correlate with an increase of outbreaks [20,72,87,93]. This is partly due to the fact that many other species of marine bacteria of unknown virulence potential also harbor homologous genes [94–96]. This suggests that further research is needed in identifying virulence factors important in *V. parahaemolyticus* pathogenicity in humans [20,81,97,98].

Urease

V. parahaemolyticus does not typically produce urease, but some *V. parahaemolyticus* strains carry a *ure* cluster that enables the production of the enzyme [85,99]. Urease is an enzyme that breaks down urea into ammonium carbonate. It is produced by many bacteria, fungi, plants, and some vertebrates [99]. Pathogenic bacteria such as certain strains of *Staphylococcus* spp., *Helicobacter* spp., and *Mycobacterium* spp. typically encode urease [99]. Bacterial ureases are involved in the formation of ammonia during the process of infection, which reduces the acidity of the local environment, beneficial for bacterial survival [85,99]. Urease also contributes to damage of host epithelial cells by activating host monocytes and neutrophils to secrete inflammatory cytokines [85,99]. Urease activity may aid in the colonization of host tissue by bacteria by inhibiting the biosynthesis of mucus and cause intestinal mucus to disassemble at the mucosal surface, allowing for better access of invading bacteria to the epithelial cells [100–103]. Though urease is not essential for *V. parahaemolyticus* infection, it is suggested that it contributes to increased survival of *V. parahaemolyticus* and can be diagnostic for strains harboring Trh, though not those harboring Tdh [85,103].

Vibrio Pathogenicity Islands

Pathogenicity Islands – General

Pathogenicity Islands (PaIs) are a type of genomic island (GI) found on the chromosomes of many bacteria [104]. PaIs are characteristically large chromosomal regions (10 to 200 kb) inserted near tRNA genes and typically harbor functional mobility elements (i.e., direct repeats, integrases, transposases, insertion sequences), and differ in base composition from the core genome. Importantly, they contain virulence genes and are present in pathogens but absent in benign relatives [55]. PaIs typically have a mosaic structure due to their numerous acquisitions and adaptations [55]. Many bacterial pathogens, both gram-negative and gram-positive, have PaIs, including prominent human pathogens such as *V. cholerae, V. parahaemolyticus, Helicobacter pylori, Escherichia coli, Salmonella spp., Shigella spp., Pseudomonas syringae, Listeria spp., Staphylococcus aureus, Streptococcus spp., Enterococcus faecalis* and *Clostridium difficile* [104].

VPaI 1-7

The first fully sequenced *V. parahaemolyticus* genome was completed in 2003 from the pandemic clonal complex strain RIMD2210633 [105]. *V. parahaemolyticus* contains two chromosomes. The pandemic strain genome contained seven Vibrio Pathogenicity Islands (VPaIs), with VPaI-1, -4, -5 and -6 exclusive to the pandemic strain [105,106]. This lineage of *V. parahaemolyticus* gained its human pathogenicity from these VPaIs from unrelated lineages [34,65,98,105–109]. Espejo et al. (2017) reviewed the plausible origins for each VPaI and indicates that the O3:K6 non-pathogenic founder strain acquired a *toxRS* and at least seven novel genomic islands [110]. All seven VPaIs contain the typical characteristics of pathogenicity islands as defined above, such as large size (all were 10-81kb), flanked by direct repeats, G+C content lower than

the rest of the genome (except VPaI-2), and flanked by integrases or transposases for mobility. VPaI-1-4 were inserted adjacent to tRNA genes [106].

VPaIs known to confer virulence

Despite the numerous VPaIs identified from the pandemic clones, almost all clinical *V. parahaemolyticus* isolates contain either one or both hemolysins *tdh* and *trh* alongside a Type Three Secretion System (T3SS) locus on one of three evolutionarily distinct but related VPaI: VPaIα, VPaIß, and VPaIγ [75,98,105,106]. Until recently, researchers named each VPaI based on the strain that harbored them. VPaIα (*tdh*$^+$) was previously reported as Vp-PAI$_{RIMD2210633}$ [98]/Vp-PAI [105]/VPaI-7 [106]/*tdh*VPA [65], whereas VPaIß (*trh*$^+$) was previously reported as Vp-PAI$_{TH3996}$ [98]/*trh*VPA [65,109]. Xu *et al*. designed a non-strain specific naming system for these islands based on the previously designated T3SS they harbored [107]. The T3SS2 on VPaIα is evolutionarily divergent from the T3SS2 on VPaIß/γ; thus these secretion systems are termed T3SS2α on VPaIα, and T3SS2ß on VPaIß [98,107]. VPaIγ contains genetic elements from both VPaIα and VPaIß [107], thus due to this hybrid nature was termed "VPaIγ" [65,98,107,109] . A gene map of all three VPaI from Xu et al. (2017) is shown in Figure 1.1 [107].

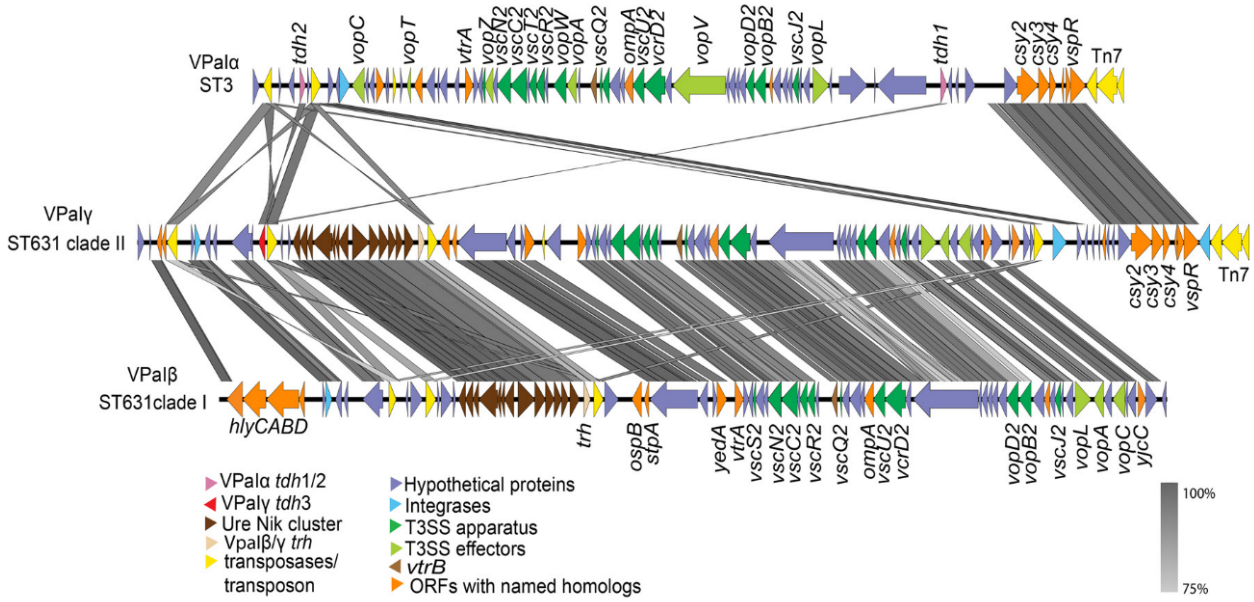


Figure 1.1: Comparison of gene content and conservation in a typical VPaIα, VPaIβ, and VPaIγ. VPaIα was derived from ST3 strain RIMD2210633, VPaIß was derived from ST631 clade II isolate MAVP-Q, and VPaIγ was derived from ST631 clade I isolate MAVP-R. ORFs are depicted in defined colors, and similarities (≥75%) among ORFs are illustrated in gray blocks. Figure from Xu et al. [107]

VPaIα, first described in strain RIMD2210633 as VPaI-7, is currently the most extensively studied VPaI as it is the VPaI harbored by the pandemic clonal complex of *V. parahaemolyticus*, which causes the majority of infections by *V. parahaemolyticus* globally [105]. VPaIα, as described in strain RIMD2210633 (NC_004605 region between VPA1312 and VPA1395), is 81-kb [83 open reading frames (ORFs)] and is located on chromosome 2 in RIMD2210633 [110]. Notably, it contains two copies of *tdh* [*tdh*1 and *tdh*2 alleles (*tdh*S/A respectively) [65]] [75,78,111], cytotoxic necrotizing factor, an exoenzyme T gene, five transposase genes, and T3SS2α apparatus as well as several T3SS secreted toxins [107].

VPaIß is evolutionarily divergent from VPaIα, with limited shared genetic content [65]. VPaIß from ST631 clade I strain MAVP-R (MF066647.2) is 96-kb and contains 88 ORFs [65,107] with ~78 ORFs unique to VPaIß [107] when compared to VPaIα from strain RIMD2210633. VPaIß is located

on either chromosome of *V. parahaemolyticus* dependent on strain [65]. VPaIß contains a T3SS2ß, which is related to the T3SS in non-O1, non-O139 strains of *V. cholerae* [98], and harbors one copy of *trh* [98]. The VPaIß (from strain MAVP-R) contains other interesting gene content including an integrase, transposases, a urease gene cluster, and a peptide/nickel transportation system [65,98].

By 1990, isolates of *V. parahaemolyticus* harboring a VPaI that contained both *tdh* and *trh* caused gastroenteritis [73]. Bioinformatics analysis of quality assemblies identified that all strains harboring *tdh* and *trh* concurrently, harbor these together in VPaIγ [107]. VPaIγ is 102-113-kb (~105-114 ORFs) and is located on chromosome 2 in sequenced genomes. Analysis of this island indicated that the gene content is a hybrid of VPaIα and VPaIß [107]. The 5' end of the island contains a mix of genes from both VPaIα and VPaIß, including a *tdh* gene homologous to that from VPaIα [107]. Interestingly, this *tdh* gene, though apparently from VPaIα, is flanked by gene content from VPaIß [107]. The core of the island is orthologous in content and syntenous with VPaIß, but 3' end of the island is highly conserved with that of VPaIα [97,107].

Secretion Systems and their Effectors

Secretion Systems

Bacteria have developed numerous mechanisms to increase their environmental fitness or pathogenic success. The utilization of a protein secretion system is a method commonly used among bacterial pathogens. Bacterial secretion systems are highly controlled and specific machinery that transfer multiple bacterial encoded proteins across phospholipid membranes [112]. As reviewed in Green and Mecsas (2016), bacterial pathogens often use these systems to successfully invade hosts, damage host tissue, and evade host immune defenses [113].

To date, there are seven bacterial secretion systems. The first six secretion systems are in gram-negative bacteria and the seventh is found in only a few species of *Mycobacteria* and *Corynebacteria*. The six gram-negative secretion systems are depicted in Figure 1.2 as originally published by Gunasinghe (2017) [114]. Type I Secretion Systems (T1SS) resemble ATP-binding cassette (ABC) transporters, and they transport their substrates across both inner and outer bacterial membranes [113,115]. Type II Secretion Systems (T2SS) form a channel in the outer bacterial membrane, and transport folded proteins from the periplasm of bacteria, across the outer membrane, into the extracellular environment [113]. Type III Secretion Systems (T3SS) are discussed below. Type IV Secretion Systems (T4SS) are related to bacterial DNA conjugation systems. They can transport single proteins, protein-protein, and DNA-protein complexes across both its own bacterial inner and outer membranes, but also other bacterial and eukaryotic membranes allowing for direct injection of the substrate into a recipient cell's cytosol [113,116]. Type V Secretion Systems (T5SS) are unique systems of autotransporters that allow secretion of proteins or groups of proteins across the outer membrane of gram-negative bacteria [113,117]. Type VI Secretion Systems (T6SS) are large structures that, like T3SSs and T4SSs, translocate proteins directly from their cytosol into the cytosol of other cells in a contact-dependent manner, though T6SS can translocate proteins into both bacterial and eukaryotic cells [113,118]. Type VII Secretion Systems (T7SS) are the secretion systems specialized for transport of proteins across the heavily lipidated cell wall layer found in gram-positive organisms such as *Mycobacteria* and *Corynebacteria* [113].

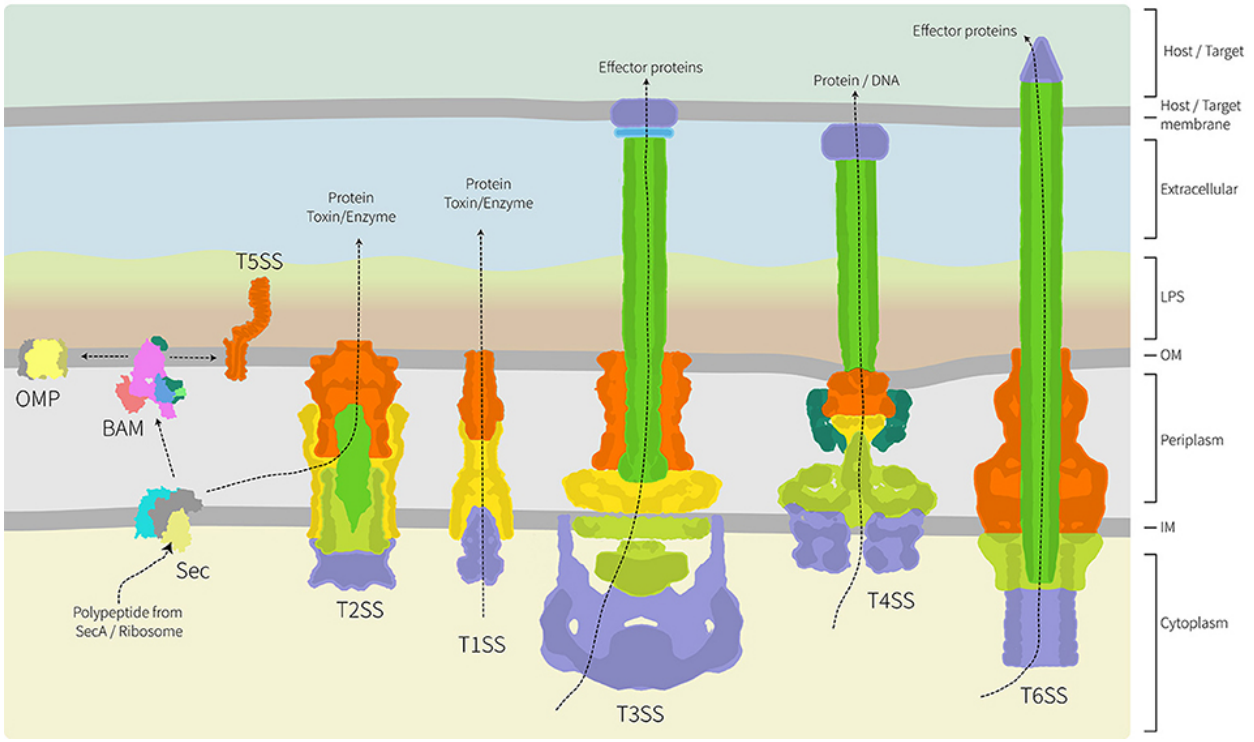


Figure 1.2. Architecture of the major protein secretion systems found in gram-negative bacteria. Figure from Gunasinghe (2017) [114]. Of the six major gram-negative protein secretion systems, five span the inner and outer membranes and the periplasm. These nanomachines are thereby trapped in the peptidoglycan layer. The T1SS, T3SS, T4SS, and T6SS collect substrate proteins directly from the cytoplasm for secretion across both membranes, while the T2SS collects substrate proteins from the periplasm with their delivery there being via the Sec translocon (shown) or Tat transport system (not shown).

T3SS

Discovered in the 1990s [119], the T3SS is a bacterial needle-like apparatus that can detect eukaryotic cells, and inject bacterial protein effectors directly into the plasma membrane and cytoplasm of the eukaryotic cell [1,20,120]. Once in the eukaryotic cell, these bacterial effectors can hijack host cell signaling and can manipulate host cell functions to benefit the bacteria [20,119,120]. This apparatus, termed the injectisome, is evolutionary related to the components of the bacterial flagellum that are involved in flagellar structure assembly [1,119,121].

Though also essential to some symbionts, T3SSs are typically found in gram-negative animal, plant, insect and amoeba pathogens, and tend to contribute to their virulence [119]. Some important human pathogens that use T3SSs for their virulence include *Salmonella enterica* serovars, *Shigella spp., Yersinia spp., Chlamydia spp., Pseudomonas spp., Vibrio spp., Bordetella spp*., and pathogenic strains of *E. coli* [112,119].

A total of 20-30 proteins make up the structure of the injectisome, forming three primary sections: the secretion apparatus, transposon pore, and needle [1,20]. The secretion apparatus traverses both the inner and outer bacterial membranes [1,20]. The transposon pore is composed of secreted proteins called translocators, which form a hole crossing the target eukaryotic membrane [1,20]. The approximately 60-nm-long needle is polymerized and extended into extracellular space to connect the secretion apparatus and transposon pore [1,20,119]. Genes for the injectisome and its translocators are commonly encoded within a single gene cluster located on a mobile genetic element such as a plasmid or pathogenicity island and thus can spread between bacteria through the process of HGT [119].

T3SS Effectors

T3SSs typically deliver multiple effector proteins in a coordinated fashion to modulate a variety of complex cellular processes enabling the bacterium to colonize its host [20,112]. T3SS effectors typically function through some of the following themes: mimicry of host cell proteins, covalent modifications of host cell proteins (e.g. phosphorylation, acetylation, and AMPylation) to transiently alter the activity of the host cellular targets, work in concert with other effector proteins, and work under precise temporal regulation/localization [112]. T3SS effectors have functional redundancy in that they may have different biochemical activity, but target a similar cellular

process for a similar net result [112]. Homologous effectors can be encoded by very dissimilar pathogens, but each homolog may produce a different effect in their recipient hosts [112]. Importantly, T3SS effectors are often encoded outside of the genetic block encoding the injectisome apparatus, allowing for greater nucleotide variation and recombination among effectors while retaining a functional secretion apparatus [119].

V. parahaemolyticus T3SSs and Virulence Effectors

V. parahaemolyticus T3SS

All *V. parahaemolyticus* have an ancestral T3SS (T3SS1). Some pathogenic *V. parahaemolyticus* also have a second horizontally acquired T3SS harbored on VPaIα/ß/γ (T3SS2α or T3SS2ß), which makes them enterotoxic [9,84,98,105,119,122]. T3SS1 is located on Chromosome 1 in all *V. parahaemolyticus* [35]. T3SS1 effectors have a cytotoxic effect in cell culture, but they are not an important component of enterotoxicity [8,9,123,124]. T3SS1 has four main effectors, VopQ, VopS, VopR, and VPA0450 (Table 1.1, adapted from Wang et al. 2015) [6] that initiate a series of events in tissue culture that cause autophagy, membrane blebbing, cell rounding, and cell lysis [1,123]. T3SS1 effectors are thought to be important for systemic *V. parahaemolyticus* infections and may determine the final outcome of disease [125].

T3SS2 is an accessory system on VPaIα/ß/γ, primarily in pathogenic strains of *V. parahaemolyticus* [1,23,35,70,105,122]. The T3SS2 found in *V. parahaemolyticus* has close homology to the T3SS Hrp1 system in *Pseudomonas syringae* [1]. Effectors of T3SS2α have been extensively studied and are discussed in the following section. Some of the effectors with no proven role in human disease are also secreted by T3SS2ß, but there is little homology between the primary T3SS2α effectors VopV and VopZ with those similar regions found on VPaIß/γ [107]. During

gastrointestinal infections, the effectors from both T3SS2α/ß impair epithelial cell function and structure resulting in damage to the intestinal lining, inflammation, and diarrhea [9].

T3SS2 effectors and Gastroenteritis Pathogenesis

The human gut mucosa makes up the primary barrier to the underlying tissues and organs of the human body. The first layer of the gut mucosa are epithelial cells termed enterocytes, with occasional specialized epithelial cells. One mechanism of gut pathogen proliferation and survival is to disrupt this barrier to gain access to the tissues underneath [126].

After consumption, when *V. parahaemolyticus* makes it to the intestinal tract, it first must attach to the enterocytes. A tight attachment to enterocytes is necessary for the bacterial T3SS needle to form a channel and pump effectors into the host cell. Throughout the length of the gastrointestinal tract, there is variation in mucus density, commensal flora composition and a wide diversity of cell types [20]. Due to the highly dynamic environment of the gastrointestinal tract, *V. parahaemolyticus* has developed multiple adhesion mechanisms to allow for efficient attachment to host cells throughout the intestine [20,125]. This attachment is modulated by the adhesion molecules: Mannose Sensitive Hemagglutinin (MSHA) Type IV Pilus and Multivalent Adhesion Molecule 7 (MAM7), which bind to the fibronectin and phosphatidic acid on the enterocyte [6,20,127–129]. Though more studies are needed, it is also suggested that Cell-Associated Hemagglutinin (cHA), GlcNAc Binding Protein A (GbpA), and Capsular Polysaccharide (CPS) also have an effect on adhesion to enterocytes [20,125,130–133].

After attachment, the T3SS2's translocators form a pore in the enterocyte membrane allowing the T3SS to begin pumping virulence effectors into the enterocyte. The T3SS translocation pores forms preferentially in lipid raft microdomains that are rich in cholesterol and

glycosphingolipids in the enterocyte membrane. The cholesterol in these lipid rafts binds the TTSS translocator proteins so they can assemble as pore [134]. After assembly of the translocon pore, the T3SS can pump virulence effectors into the enterocyte.

Characteristic pathological changes during *V. parahaemolyticus* infection include disruption of the intestinal epithelium, modulation of epithelial cell signaling, epithelial cell death, and intense inflammation. These pathological changes are primarily attributed to *V. parahaemolyticus* TTSS2 effectors [20]. *V. parahaemolyticus*'s T3SS2 and its effectors have been extensively characterized in VPaIα (Table 1.1) [6]. The eight characterized VPaIα effectors include VopA/P, VopC, VopL, VopO, VopT, VopV, VopZ, and VPA1380. VopV and VopZ are the two primary effectors necessary for successful *V. parahaemolyticus* infection. VopA, VopC, and VopT directly alter the activity of eukaryotic cell signaling proteins by post-translation modification [125].

V. parahaemolyticus Gastroenteritis Pathogenesis - Epithelial disruption (VopV, VopL, VopC, VopO)

One of the common pathologies of bacterial gastric infections includes disruption of the intestinal epithelium [135]. One function of the gastrointestinal epithelium is to form a selectively permeable barrier. This barrier controls the passage of nutrients, water, and ions into the tissue below [20,136]. By disrupting this primary barrier, gastric pathogens can increase non-selective intestinal permeability, modulate ion and fluid efflux and allow bacteria access to deeper tissues and interfere with host immune responses [20]. This access to deeper tissues is often associated with severe clinical symptoms [126]. During infection, *V. parahaemolyticus* induces significant epithelial damage (Figure 1.3), including epithelial cell detachment and disruption of the villus structure, triggering a substantial inflammatory response that contrasts with *V. cholerae* which does not

cause inflammatory gastroenteritis [125,135,137]. This epithelial damage occurs both on the large scale of villus architecture, but also the small scale of effacement and elongation of microvilli [135,138,139]. *V. parahaemolyticus* colonizes throughout the small intestine (duodenal) and large intestine (rectal) tissues and is highest in the distal small intestine [125]. During infection, *V. parahaemolyticus* exhibits a distinct colonization pattern of discrete microcolonies rather than even distribution across the entire epithelial surface [138].

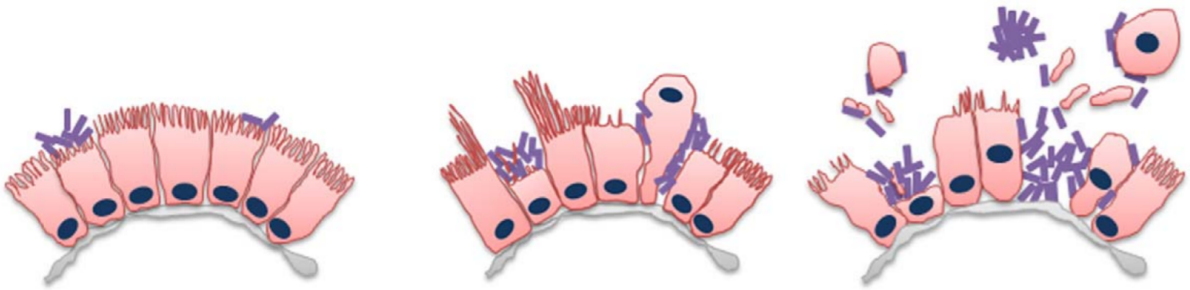


Figure 1.3. Schematic of the kinetics of *V. parahaemolyticus*-induced damage to the intestinal epithelial surface. The purple rods represent *V. parahaemolyticus.* Figure from Ritchie et al. 2012 [138]. Following initial attachment, *V. parahaemolyticus* induces erosion of microvilli and depletion of cytoplasmic contents resulting in the formation of bacterial clusters located just below the level of the surrounding epithelium. Continued depletion of epithelial cell contents either by cytoplasmic 'blebbing,' whole-cell extrusion and microvilli elongation around the edge of the cluster, leaves *V. parahaemolyticus* clusters situated within deeper cavities in the epithelium. Eventually, this leads to the disintegration of normal villus structure and the generation of large amounts of luminal debris. These pathological changes appear to be attributed to T3SS2 as a similar pathology was observed in rabbits infected with mutants lacking TDH or T3SS1.

The nucleation of actin, actin-bundling, and formation of stress fibers/filopodia/actin ruffles serve both a structural and signaling purpose of the host cells. When pathogens disrupt the normal branched architecture of the actin cytoskeleton structure of host epithelial cells, these cells lose their shape and tight junctions to surrounding cells, which facilitates bacterial dissemination [140]. The proper response to infection and damage cannot be induced when the normal cellular signaling network is disrupted [140]. During *V. parahaemolyticus* infection, epithelial disruption is primarily caused by the T3SS effectors VopV, VopL/F, VopC, and VopO which target the actin

cytoskeleton [125,141,142]. VopV binds and bundles the actin in host cells, VopL nucleates actin which forms stress fibers thus altering host cell shape, VopC activates GTPases that lead to the formation of stress fibers, actin ruffles and filipodia, and VopO induces stress fiber formation [125,141–145]. Interference and reorganization of the actin cytoskeleton affects not only cell structure, but it also activates several signal-transduction pathways, including two important for bacterial infection: the mitogen-activated protein kinases (MAPK) signaling and NF-kB inflammatory responses [141,146]

Once secreted into enterocytes, via its F-actin binding-domains, VopV binds to actin filaments and the actin-cross linking protein filamin [139]. Then by bundling the actin, VopV induces the rearrangement of the apical epithelial cell membrane, eliminates the actin-filled microvilli brush border, and elongates a few microvilli [141]. This effacement of the brush border enhances *V. parahaemolyticus* adhesion to the enterocyte [139]. It is proposed that the effacement either provides an altered surface for adhesion or exposes membrane domains and/or a specific host receptor that *V. parahaemolyticus* does not have access to with an intact brush border [139]. It is currently not known how the F-actin-binding activity of VopV affects VopV-dependent inflammation and enterotoxicity [141]. VopV is considered the primary effector in VPaIα and is responsible for T3SS2-dependent enterotoxicity, and colonization [139,141]. In the rabbit ileal loop model, *vopV* deletion mutant *V. parahaemolyticus* induces intestinal damage similar to that of no secretion of any T3SS2 effector and non-infected control, compared to the severe intestinal inflammation (epithelial loss, edema, neutrophil infiltration in the lamina propria and submucosal area, loss of goblet cells, and bleeding) seen with VopV secretion [141]. This suggests that the effacement of the enterocyte brush border is a critical step for *V. parahaemolyticus* infection [139,141].

VopL, VopC, and VopO also affect the actin cytoskeleton, but these effectors are not required for enterotoxicity [141–144,147]. The protein domains of VopL mimic a eukaryotic actin-nucleating

protein that allows it to remodel the actin network by inducing the formation of unbranched actin structures rather than the branched networks [142]. This actin nucleation process forms elongated stress fibers and alters the cell shape [1,142,148]. VopC utilizes a mechanism other than direct contact to manipulate the actin cytoskeleton [143,144]. VopC activates the small GTPases Rac and CDC42 to contribute to stress fibers, actin ruffles, and filipodia formation [143,144,149]. Zhang et al., Okada et al., and de Souza Santos et al. suggest that VopC and VopL help facilitate the entry of *V. parahaemolyticus* into non-phagocytic host cells [143,144,150]. Recently, a germ-free mouse model was developed to assess gastrointestinal pathogenesis of *V. parahaemolyticus* [151]. This model enabled assessment of gastrointestinal pathogenesis in earlier hours of infection to evaluate initial tissue invasion rather than the final result of maximal tissue damage at the end of infection as seen in previous models [138,144,151]. In this model during the early hours of infection VopC promoted epithelial cell invasion and increased gastrointestinal pathogenesis [151]. VopO is an effector that has been shown to activate the RhoA-ROCK signaling pathway, which induces actin stress fiber formation, contributing to the disruption of the host epithelium [145].

V. parahaemolyticus Gastroenteritis Pathogenesis – Cell Death (VopT)

Another characteristic pathological change during *V. parahaemolyticus* infection includes epithelial cell death. During infection, cell death can be a host defense mechanism, to limit infection and colonization by enteric pathogens [126,152]. Some enteric pathogens block apoptosis to preserve their replication sites, while others such as *V. parahaemolyticus*, trigger cell death to induce breaches in the epithelium to gain access to underlying tissue [126,152].

VopT shares the ADP-ribosyltransferase domain of *Pseudomonas aeruginosa* cytotoxins ExoT and ExoS, with biological activity more similar to that of ExoS [124]. VopT is cytotoxic to cultured

intestinal cells but is not enterotoxic [124,141]. VopT can modify Ras, which could affect any number of the Ras signaling pathways such as the mitogen-activated protein kinases (MAPK) and phosphoinositide-3 kinase (PI3K) pathways, which are important in controlling several functions, such as cell growth and survival [124,153,154].

V. parahaemolyticus Gastroenteritis Pathogenesis – Inflammation (VopZ, VopA/P)

Finally, another hallmark of *V. parahaemolyticus* induced gastroenteritis is severe inflammation, which is specifically induced by *V. parahaemolyticus*, not the host in response to the extensive tissue disruption [125]. Pathogens trigger inflammation as a mechanism to outcompete the commensal microbiota, disrupt tight junctions and impair gut barrier integrity [126,155,156]. Some pathogens exhibit higher resistance to host mucosal antimicrobial peptides compared to commensals; thus inflammation alters the microbiota composition in favor of the pathogen [126,157–159]. Certain pro-inflammatory cytokines produced by host cells, such as Tumor Necrosis Factor-α (TNF-α), disrupt the tight junctions between the enterocytes lining the intestinal tract [126,155,156]. This disruption allows bacteria access to deeper tissues [126,155,156]. *V. parahaemolyticus* induces inflammation through recruitment of inflammatory cells, such as PMNs, and induces host cell transcription of the pro-inflammatory cytokines TNF-α, Interleukin-1β (IL-1β), Interleukin-6 (IL-6), and Interleukin-8 (IL-8) [125,135].

Commonly, the effectors of bacterial pathogens target the host intracellular signal transduction cascades [160]. Two of the most frequent effector targets are the inflammatory cell-survival pathways controlled by the MAPK signaling cascade, and the NF-kB signal transduction pathway [160]. The MAPK signaling cascade is essential for normal operation of the eukaryotic cell such as cellular metabolism, growth, and division, as well as inducing inflammation as an

antimicrobial defense mechanism [160]. The NF-kB signal transduction pathway is an important component of innate immunity, with a role in apoptosis, cell survival, and cytokine production [160,161]. Inhibition of these pathways leads to inflammation, distorted epithelial structure, cell death, reduced barrier function, and recruitment of PMNs [147].

The second critical *V. parahaemolyticus* T3SS2 effector is VopZ. VopZ is a multifunctional effector that is essential for *V. parahaemolyticus* colonization, enterotoxicity, and induction of diarrhea [147]. VopZ inhibits activation of the kinase TAK1, which is important for the activation of the two signaling pathways MAPK and NF-κB [147,162]. Previous studies have shown that absence of TAK1, and thus non-activation of MAPK and NF-κB signaling pathways, in the intestinal epithelium can lead to phenotypes such as inflammation, apoptosis, distorted epithelial structure, mislocalization of tight junction proteins, and reduced barrier function [125,147,163]. These phenotypes are advantageous for invading enteric pathogens, allowing for increased colonization and access to deeper tissues [141,146].

Another inflammation-inducing effector is VopA (also referred to as VopP). VopA acetylates MAPK kinases, preventing phosphorylation of the kinase, thus inhibiting the MAPK signaling pathways [1,164,165]. Despite its role in controlling cytokine production, VopA is not a critical effector for infection as there was no obvious reduction in enterotoxicity in rabbit ileal loop [141].

VPA1380 is a T3SS secreted cysteine protease that likely targets a host substrate [166]. Its role in *V. parahaemolyticus* infection is not known. When expressed intercellularly VPA1380 is toxic to yeast, but it is not enterotoxic in the rabbit ileal loop model [141,166].

Table 1.1: Known *V. parahaemolyticus* virulence effectors and their biological activity. Gene names are orthologs found in strain RIMD2210633 VPaIα (NC_004605 region between VPA1312 and VPA1395). "+" = the presence of effector on VPaIα/ß/γ. "-" = the absence of effector and predicted effector on VPaIα/ß/γ. "~" = divergence between orthologs found in VPaIß/γ compared to the characterized effector in VPaIα.

Effector	Gene	Biological Activity	Effects on host cells	Gene presence on VPaI		
				α	ß	γ
Toxins						
Tdh	*tdh*	Forms pores on cells	Cytotoxicity and enterotoxicity	+	-	+
Trh	*trh*	Forms pores on cells	Cytotoxicity and enterotoxicity	-	+	+
Tlh	*tlh*	Hemolysin activity	Cytotoxicity	N/A	N/A	N/A
T3SS1 effectors						
VopQ	VPA1680	Forms pores and binds V-ATPase	Autophagy, cell lysis, MAPK activation, IL-8 secretion	N/A	N/A	N/A
VopS	VPA1686	Inhibition of Rho by AMPlyation	Cell rounding, phagocytes invasion	N/A	N/A	N/A
VPA0450	VPA0450	Phosphatidylinositol phosphatase	Membrane blebbing, destabilization	N/A	N/A	N/A
VopR	VPA1683	Binds PIP2 in membrane	Promoting refolding of T3SS effectors	N/A	N/A	N/A
T3SS2 effectors						
VopA/P	VPA1346	Inhibition of MAPK by acetylation of MKK	Blocking of phosphorylation and ATP binding	+	+	+
VopT	VPA1327	Ras ADP-ribosylation	Cytotoxicity and yeast growth inhibition	+	-	-
VopL	VPA1370	Actin nucleation	Stress fiber formation and cell shape changing	+	+	+
VopC	VPA1321	Activation of Rac and CDC42 by deamination	Invasion of non-phagocytic cells, enterotoxicity	+	+	+
VopO	VPA1329	Activates the RhoA-ROCK pathway	Induces stress fiber formation	+	-	-
VopV	VPA1357	Actin binding and bundling	Enterotoxicity and blunting of villi	+	~	~
VopZ	VPA1336	Inhibition of TAK1 and downstream MAPK and NF-kB	Enterotoxicity and colonization	+	~	~
VPA1380	VPA1380	Cysteine catalysis dependent on IP6	Toxicity in yeast	+	-	-

Epidemiology of *V. parahaemolyticus*

V. parahaemolyticus discovery

V. parahaemolyticus was discovered by Tsunesaburo Fujino of the Research Institute of Microbial Diseases (RIMD), Osaka University in 1950 in Osaka, Japan after an outbreak of acute gastroenteritis from Shirasu in 272 individuals, 20 of whom died [167]. In the decades after this event, outbreaks continued and *V. parahaemolyticus* became recognized as the leading cause of seafood-borne gastroenteritis [168–175].

ST3 (O3:K6)

Prior to 1995, gastroenteritis caused by sporadic *V. parahaemolyticus* infection was associated with many serotypes [45,176,177]. In 1995, an outbreak of *V. parahaemolyticus* occurred in Calcutta (now known as Kolkata), India [2,26,176]. This outbreak was caused by a novel strain of *V. parahaemolyticus* with serovar O3:K6 (with numerous serovariants), later determined by Multi-locus sequence typing (MLST) to be ST3 [26,65,178,179]. This ST3 strain spread throughout Asia[180] [178], Africa[181], Europe[182], and North and South America [18,173,178,180,183,184] and so was called the pandemic clonal complex of strains [26,110,178]. Since then, ST3 strains have caused numerous food-borne outbreaks in Asian countries, the United States and worldwide [1,110]. ST3 strains of *V. parahaemolyticus* have only caused two outbreaks in the United States, one from Texas oysters and the other from New York oysters in 1998 [26,45,46]. Although not prevalent in the United States, the pandemic complex of ST3 strains causes the most disease globally [26,45,173,176,178,185–188]. The T3SS effectors on the VPaIα harbored by the pandemic strain induced the gastroenteritis seen in these outbreaks [65,98,105,106].

Arrival in the United States

The first reported *V. parahaemolyticus* outbreak in the United States was in Maryland in 1971, with three outbreaks [189]. The illnesses were linked to steamed crabs [189]. There was variation in the clinical isolates from these outbreaks with multiple serotypes, and both KP+ (VPaIα) and KP- isolates (no island/VPaIß/VPaIγ) [189]. *V. parahaemolyticus* has continued to cause sporadic infections in the United States [18]. Serotypes with O4 were the most dominant in the United States [45,176,177]. In the United States, the most clinically prevalent STs harbor VPaIß and VPaIγ [19,39,40,190].

Northeast United States

Historically there was low disease incidence of *V. parahaemolyticus* infections in the Northeast United States. This low prevalence of disease in the Northwest Atlantic was attributed to cooler water and unfavorable conditions for pathogen populations to establish compared to warmer southern regions like in the Gulf of Mexico [33,52]. The rare outbreaks that did occur were driven by anomalously warm ocean temperatures, and introduced strains [22,48,50,185]. The recent increase in *V. parahaemolyticus* infections in the North Atlantic are attributed to the invasion and establishment of Pacific-native lineages, in particular hypervirulent ST36, as well as a trend of warmer than usual ocean temperatures allowing more successful movement and establishment of Gulf of Mexico (GOM) native *V. parahaemolyticus* populations [22,26,35,48,191–196]. During this time VPaIs were donated from the invading lineages to the native New England populations [11,22,197].

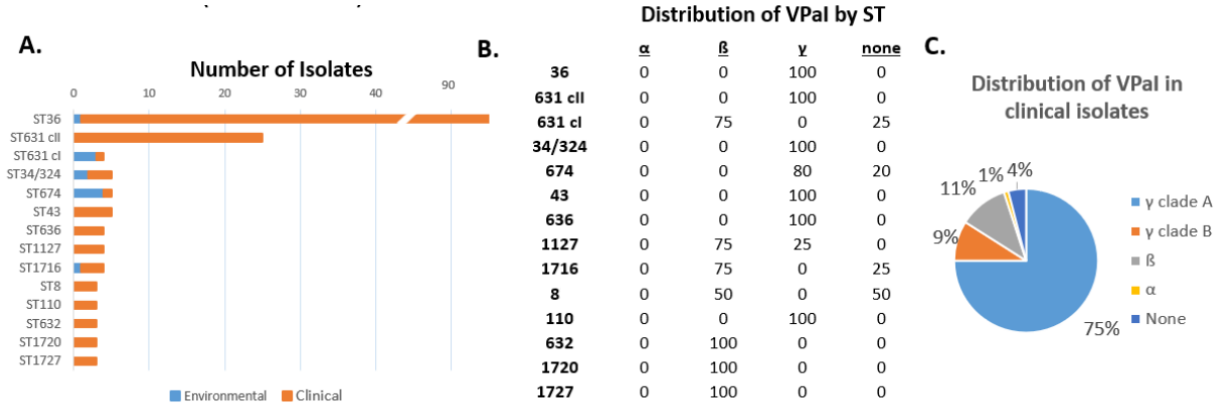


Figure 1.4. Clinically prevalent Sequence Types (ST) and corresponding pathogenicity island content. Adapted from Xu et al. 2017 [11]. A) The total number of clinical and environmental isolates from prevalent *V. parahaemolyticus* lineages reported in four Northeast US States (ME, NH, MA, and CT) from 2010-2016. B) Percentage of pathogenicity island type from prevalent *V. parahaemolyticus* lineages reported in panel A. C) Distribution of VPaI type of all gastric infections reported in ME, NH, MA, and CT.

ST36 (O4:K12) and other Pacific lineages

V. parahaemolyticus ST36 is a long-time resident lineage originating from the Pacific Northwest and is the most clinically prevalent lineage of *V. parahaemolyticus* in the United States and Canada [22,39,40,48,107,177,193,198]. ST36 was first identified in the Pacific Northwest (PNW) in 1979, and until 2006, this PNW resident lineage only contributed to infection on the Northwest US coast [52,177,180,199]. In 1997, ST36 caused the first outbreak linked to US product along the Pacific Coast [200]. Infections from this outbreak caused 207 confirmed cases and one death [200]. The outbreak was precipitated by unusually warm ocean temperatures [200]. ST36 was also a major contributor to multiple outbreaks in 2006 in New York, Oregon, and Washington [201]. Prior to 2012 members of the ST36 clonal complex were introduced into the Atlantic Ocean by an unknown route and established robust local populations [22,48]. In 2012 members of these populations caused a series of outbreaks from Atlantic shellfish originating from New York and Spain [48,49].

Pacific-derived ST36 lineage continues to dominate infections in the Northwest Atlantic [107]. This pattern of invasion of Pacific-native strains into the Atlantic causing outbreaks also included pathogenic sequence types ST636 and ST43 [39,40]. In 2011 and 2013 and in continuing years ST636 and ST43, respectively, caused infections linked to products from along the Northeast US coast [22]. ST36, ST43, and ST636 all harbor VPaIγ.

ST631

In the past decade, a new lineage of pathogenic *V. parahaemolyticus* emerged in the Northwest Atlantic known as ST631. ST631 was first reported in 2007 from a clinical case reported in Maryland that was traced to oysters harvested in Florida [198]. The second reported clinical case arose two years later traced to Prince Edward Island, Canada [40]. Since then, ST631 infections have increased with 35 confirmed cases reported between 2010 and 2015 in just four Atlantic coastal US states [197]. With this increase, ST631 has become the second most predominant pathogenic lineage in the Northwest Atlantic, after ST36 [22,40,197,198]. This emergence of ST631 as a major endemic pathogen [107] also coincided with a trend in warmer ocean temperatures in the Northwest Atlantic and invasion by nonresident pathogen lineages [22,48,50,185]. Non-native pathogen lineages included several Pacific native lineages (ST36, ST43, and ST636), and long-time residents of the Gulf of Mexico (ST34, ST110, and ST308).

Population analysis of the Northwest Atlantic clinical and environmental isolates indicates that two clades diverged from a common ST631 ancestor, each with strains independently acquiring VPaIs from the invading nonresident pathogen lineages [11,107]. Isolates of ST631 clade I harbored either no VPaI or VPaI-ß and were not clinically prevalent [107]. Analysis of the ST631 clade II isolates indicate that it is the clonal clade II that is the major Atlantic endemic pathogenic

lineage [11] that dominates Atlantic-derived ST631 infections. ST631 clade II harbors VPaIγ that was likely donated by the same population as ST36 during its Atlantic invasion [11,107].

Long-time residents of the Gulf of Mexico and their spread

The long-term resident lineages of the Gulf of Mexico (GOM) harbor primarily no island, VPaIß, or VPaIγ. Strains harboring VPaIα are isolated occasionally, but have not established a population since most *tdh*$^+$ isolates in the area are also *trh*$^+$ [53]. STs such as ST34, ST110, and ST308 are native to the GOM that harbor VPaIγ have moved up the coast and are now resident in the Northwest Atlantic [22,53,60,190,192].

GOM native ST34 was first reported in both the GOM and Massachusetts in 1998 [107]. During the past decade, coinciding with a trend in warmer ocean temperatures, ST34, along with other members of the Gulf of Mexico *V. parahaemolyticus* population [53,60,192], has continued its spread north and in 2012 it was also recovered in New Hampshire, where it established residency [22]. GOM strains moving north have also spread their VPaIs via HGT to local populations [11]. Analysis of the VPaIs, show that during its spread north, ST34 donated its VPaIγ to the local North Atlantic resident ST674, transforming it into a pathogen [11]. ST34 and ST674 are the most frequently recovered pathogen lineages among environmental isolates, but despite harboring a VPaIγ, they cause very few infections [107].

Flavors of VPaIγ

Pathogenic lineages of *V. parahaemolyticus* harboring VPaIγ are increasingly associated with disease in North America [19,22,39,40,52,198]. Whereas most Northeast Atlantic *V. parahaemolyticus* pathogens harbor related VPaIγ, there is variation in their clinical prevalence [11].

A recent study by Xu et al. indicated that the geographical location and clinical prevalence of VPaIγ could align with allelic variation and lineage of the *tdh* gene [11]. There are currently seven *tdh* alleles harbored by VPaIs. *tdh*1, *tdh*2, and *tdh*4 alleles are located on VPaIα, whereas VPaIγ contains one of four distinct *tdh* alleles: *tdh*3, *tdh*5, *tdh*6, or *tdh7* [11].

In the United States, the primary populations of VPaIγ-harboring *V. parahaemolyticus* are either in the Pacific Ocean, the Gulf of Mexico, or the Atlantic Ocean. *tdh*3 is exclusively harbored in Pacific-derived lineages, and *V. parahaemolyticus* harboring *tdh*3-VPaIγ are linked to high clinical prevalence [11]. *tdh*5 is closely related to *tdh*3, but is in *V. parahaemolyticus* lineages from the Gulf of Mexico and the Atlantic Ocean [11,22,107,190,198]. All lineages harboring *tdh*5-VPaIγ are of low clinical prevalence [11]. *tdh*7 is divergent from the other *tdh* alleles, and has been isolated in both Pacific-derived and Gulf of Mexico-derived lineages [11,39,40,53]. *tdh*6 has been found in lineages worldwide [11]. *V. parahaemolyticus* harboring all of these VPaIγ lineages have invaded the US Atlantic coast, causing infections, and are spreading the VPaIs to the local *V. parahaemolyticus* residents, transforming them into pathogens [11]. Since pathogen lineages harboring *tdh*3/*tdh*5/*tdh*6-VPaIγ vary so much in clinical and environmental prevalence, this suggests that each VPaIγ could have evolved to confer different degrees of virulence [11].

RESEARCH OBJECTIVES

Chapter II: What T3SS effectors does VPaIγ encode?

V. parahaemolyticus strains harboring VPaIγ cause the majority of infections in the North Atlantic, yet the effectors on this island remain unidentified and uncharacterized. The first objective (Objective I) is to identify the T3SS effectors that are present on VPaIγ. The approach to address this objective was to conduct a bioinformatic prediction of putative effectors VPaIγ based on nucleotide sequence similarity and protein domains to know T3SS effectors. The toxicity of the predicted proteins in two eukaryotic models, as a proxy for enterotoxicity, was used to characterize which toxins are likely most important to human disease.

Chapter III: Which VPaIγ T3SS effector alleles are important for human disease and environmental prevalence?

There are four flavors of VPaIγ in the North Atlantic, reflected in the allelic variation of *tdh* on the island. Each lineage varies in clinical and environmental prevalence, with *tdh3*-VPaIγ predominantly causing infections yet rarely isolated environmentally, and *tdh5*-VPaIγ with the reciprocal relationship. The second objective (Objective II) is to compare the T3SS effectors of the local VPaIγ lineages to examine if sequence variation equates with differences in toxicity in eukaryotic models of clinical and environmental fitness.

CHAPTER II: IDENTIFICATION OF *V. PARAHAEMOLYTICUS* TYPE THREE SECRETION SYSTEM TWO EFFECTORS

Sarah Eggert[1], Randi Foxall[1], and Cheryl Whistler[1,2*]

[1]Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, NH

[2]Northeast Center for Vibrio Disease and Ecology, University of New Hampshire, Durham, NH

[*]Corresponding author email: cheryl.whistler@unh.edu

Abstract

Vibrio parahaemolyticus, the leading cause of seafood-borne bacterial infections, recently increased in clinical prevalence in the United States, particularly from sources in the northwest Atlantic Ocean. Most pathogenic variants of *V. parahaemolyticus* harbor a mobile Vibrio Pathogenicity Island (VPaI) encoding a diagnostic hemolysin gene (*tdh* and/or *trh*) and a Type III Secretion System (T3SS). Three island types have been characterized and include VPaIα (*tdh*$^+$), VPaIβ (*trh*$^+$), and a mosaic VPaIγ (*tdh*$^+$/*trh*$^+$). Although most infections globally are caused by strains harboring VPaIα, for which several major effectors are known, infections in the United States are predominantly caused by strains harboring VPaIγ, which is relatively uncharacterized. Our goal was to identify effectors in VPaIγ using bioinformatics and bioassays. Analysis predicted five putative effectors shared between VPaIβ and VPaIγ but absent in VPaIα. When expressed intracellularly, three predicted effectors impaired *Saccharomyces cerevisiae* growth similarly to the known effector VopA, suggesting they are T3SS2 effectors. However, a third predicted effector did not impair growth, nor did VopZ, one of the major virulence determinants of VPaIα. Four

effectors were cytotoxic to cultured human colon cells, warranting further investigation of their role in human gastroenteritis.

Introduction

Vibrio parahaemolyticus is the leading cause of bacterial-induced seafood poisoning both in the United States and worldwide. Infection with *V. parahaemolyticus* can take on three forms: gastroenteritis, wound infections, and, rarely, lethal septicemia, with the most common form as gastroenteritis [18]. Recent increases in infections in the Northeast US from Atlantic-harvested shellfish coincided with warmer than usual ocean temperatures, and the invasion of a lineage of the "hypervirulent" Pacific Northwest endemic pathogenic strain called Sequence Type (ST) 36 [8,22,48,49]. Increasing illnesses have also been caused by an endemic pathogenic lineage, ST631 clade II, first reported in the North Atlantic in 2007 and which acquired a VPaIγ from the ST36-containing population [11,107]. Currently, ST36 causes the vast majority of infections in the North Atlantic, followed by ST631 clade II [11,107].

The Type Three Secretion System (T3SS) is a needle-like apparatus that can inject effector proteins directly through the plasma membrane and into the cytoplasm of a eukaryotic cell [20,120]. These effectors then can hijack host cell signaling and manipulate host cell functions to benefit the bacteria [20,119,120]. *V. parahaemolyticus* has one ancestral T3SS on chromosome I that is not important for human disease, and some strains, predominantly clinical isolates, have a second accessory T3SS encoded by one of three related Vibrio Pathogenicity Islands (VPaI): VPaIα, VPaIß, and VPaIγ [107]. VPaIγ is a hybrid of genetic elements from both VPaIα and VPaIß [107], with the core content orthologous and syntenous with VPaIß [97,107]. VPaIß and VPaIγ are divergent from VPaIα [107].

Worldwide, *V. parahaemolyticus* strains harboring the VPaIα, known as VPaI-7 [106] in pandemic strain RIMD2210633 [105], dominate infections. VPaIα is the most extensively studied VPaI with eight different T3SS effectors known and characterized: VopA/P, VopC, VopL, VopO, VopT, VopV, VopZ, and VPA1380 [139,141,147,162]. Previous work had only established that VopV and VopZ were critical for enteric disease in an infant rabbit model [139,141,147,162], however, a recent study in germ-free mice has demonstrated that VopC not only promotes epithelial cell invasion during early stages of infection but also contributes to gastrointestinal pathogenesis [151].

Strains harboring VPaIα rarely cause infections in the northeast United States. Instead, strains harboring VPaIγ, or occasionally VPaIß, are isolated from the majority of clinical cases (Figure 1.4) [11]. Though VPaIß and VPaIγ encode a recognizable VopC homolog, the VopV and VopZ effectors encoded in VPaIß and VPaIγ are highly divergent from those characterized in VPaIα and in the case of VopV even lack key domains that promote actin remodeling [11]. This suggests that unidentified effectors on VPaIß and VPaIγ could be primary virulence factors for *V. parahaemolyticus* strains harboring these islands. The goal of this study is to identify T3SS2 effectors that are unique to VPaIγ that is associated with the clinically prevalent STs in the North Atlantic.

T3SS effectors have a staggering level of diversity in structure with more than 75% of characterized effectors having no detectable sequence similarity to other known T3SS effectors [202]. The lack of sequence similarity makes consensus on structural patterns among the effectors challenging to parse out and difficult to predict. Despite this variation, computational analysis of known T3SS effectors indicates a modular composition with multiple functionally distinct domains or motifs [202,203]. Effectors typically have three distinct regions: the amino-terminal secretion signal, followed by the optional chaperone-binding domain, and finally, the region that

harbors host cell effector activities [202–207]. In our study we utilized bioinformatic programs, some of which specialize in T3SS effectors, to analyze open reading frames (ORFs) on VPaIα, VPaIß and VPaIγ to predict T3SS effectors, identify conserved domains, and calculate potential orthologs. We then employed biological assays to evaluate the putative effectors effect on eukaryotic cells to inform their potential significance during infection.

Results and Discussion

VPaIß and VPaIγ contain three putative effectors that are not present on VPaIα

To identify effectors on VPaIs we utilized the bioinformatic program pEffect, which combines homology-based and *de novo* approaches to predict type III effector proteins [208]. This program has a higher performance of identifying effectors over existing tools by identifying effector patterns throughout the protein sequence rather than just the N-terminus allowing for identifying a greater number and rapidly evolving effectors [208]. To first evaluate the utility of this tool we tested it against the well-characterized VPaIα from strain RIMD2210633 that contains eight known effectors among its 83 ORFs, and in parallel analyzed representative VPaIß (from strain MAVP-R, MF066647.2) containing 88 ORFs and VPaIγ (from strain MAVP-Q, MF066646.1) containing 105 ORFs. The correct identification of known effectors from VPaIα indicates this tool's reliability to classify ORFs as virulence effectors correctly. This analysis predicted three known effectors orthologous to VopA/P, VopC, and VopL, all previously characterized in the VPaIα of strain RIMD2210633 in all three VPaI (Table 2.1, Figure 2.1). The analysis also identified the known VPaIα effector VopO, [107]. Similarly, it identified VPaIα effector VopV and also predicted that the highly divergent ORFs encoded in the same location of VPaIγ and VPaIß as effectors. However, pEffect failed to predict the known VPaIα effectors VopT

(VPA1327) and VopZ (VPA1336). Finally, this analysis identified five putative T3SS effectors among the ORFs in VPaIγ, all of which were also present in VPaIß but without identified homologs in VPaIα [107].

Table 2.1: Bioinformatic prediction of T3SS effectors.

Effector name	Locus ID	pEffect Prediction	BLASTx predicted ortholog	Predicted Domains	
				CDD	SMART
Characterized effectors					
VopA/P	**α:** VPA1346	**α:**YES 69 PSI-BLAST	**α:** T3SS YopJ family effector VopA	**α:** PRK15371, Acetyltransf_14	**α:** --
	ß: MAVP-RPIeRC_00081	**ß:** YES 67 PSI-BLAST	**ß:** multidrug DMT transporter permease	**ß:** Acetyltransf_14	**ß:** LCR
	γ: MAVP-QPI_00075	**γ:** YES 69 PSI-BLAST	**γ:** multidrug DMT transporter permease	**γ:** PRK15371, Acetyltransf_14	**γ:** --
VopC	**α:** VPA1321	**α:** YES 70 SVM	**α:** T3SS2 effector GTPase-activating deamidase VopC	**α:** CNF1, CNF1-like	**α:** LCR
	ß: MAVP-RPIeRC_00083	**ß:** YES 83 SVM	**ß:** Cytotoxin	**ß:** CNF1	**ß:** --
	γ: MAVP-QPI_00077	**γ:** YES 70 SVM	**γ:** Cytotoxin	**γ:** CNF1, CNF1-like	**γ:** --
VopL	**α:** VPA1370	**α:** YES 98 SVM	**α:** Hypothetical protein/T3SS2 effector VopL nucleation of actin polymerization	**α:** Drf_FH1	**α:** WH2 (3), LCR (4)
	ß: MAVP-RPIeRC_00080	**ß:** YES 93 SVM	**ß:** Hypothetical protein/putative type III secretion system effector protein Vop	**ß:** --	**ß:** WH2 (2), LCR (4)
	γ: MAVP-QPI_00074	**γ:** YES 98 SVM	hypothetical protein, WH2 motif family protein	**γ:** --	**γ:** WH2 (2), LCR (CT4291 3, MAVP-Q 4)
VopO/ORF76	**α:** VPA1329	**α:** YES 57 SVM	**α:** Exoenzyme T	**α:** --	**α:** --
	ß: MAVP-RPIeRC_00082	**ß:** YES 65 SVM	**ß:** hypothetical protein/conjugal transfer protein TraA	**ß:** --	**ß:** --
	γ: MAVP-QPI_00076	**γ:** YES 68 SVM	**γ:** hypothetical protein/conjugal	**γ:** --	**γ:** --

			transfer protein TraA		
VopT	**α:** VPA1327	**α:** NO	**α:** Conjugal transfer protein TraA	**α:** ADPrib_exo_Tox	**α:** --
VopV	**α:** VPA1357	**α:** YES 87 SVM	**α:** Hypothetical protein	**α:** Glutenin_hmw, PRK07764 (2)	**α:** Transmembrane region, LCR (4)
	ß: MAVP-RPIeRC_00066	**ß:** YES 72 SVM	**ß:** Hypothetical protein	**ß:** --	**ß:** transmembrane region, LCR (15)
	γ: MAVP-QPI_00061	**γ:** YES 87 SVM	**γ:** Hypothetical protein	**γ:** Treacle (2) PRK07764	**γ:** transmembrane region, LCR (22)
VopZ	**α:** VPA1336	**α:** NO	**α:** Hypothetical protein	**α:** --	**α:** LCR
	ß: MAVP-RPIeRC_00049	**ß:** NO	**ß:** Hypothetical protein	**ß:** --	ß: LCR (2)
	γ: MAVP-QPI_00044	**γ:** NO	**γ:** Hypothetical protein	**γ:** --	**γ:** LCR (2)
VPA1380/ORF32	**α:** VPA1380	**α:** YES 70 PSI-BLAST	**α:** OspB protein	**α:** PTZ00440	**α:** --
	ß: MAVP-RPIeRC_00039	**ß:** YES 69 PSI-BLAST	**ß:** hypothetical protein /*OspB homolog	**ß:** --	**ß:** LCR
	γ: MAVP-QPI_00032	**γ:** YES 70 PSI-BLAST	**γ:** hypothetical protein /*OspB homolog	**γ:** --	**γ:** LCR
Putative effectors					
ORF34	**ß:** MAVP-RPIeRC_00041	**ß:** YES 96 SVM	**ß:** SslE/AcfD family lipoprotein zinc metalloprotease	**ß:** DUF4092, Peptidase_M60, M60-like	**ß:** M60-like, LCR
	γ: MAVP-QPI_00034	**γ:** YES 99 SVM	**γ:** DUF4092 domain-containing protein/sugar ABC transporter substrate-binding protein/accessory	**γ:** DUF4092, Peptidase_M60, M60-like	**γ:** M60-like, transmembrane region, LCR

			colonization factor AcfD		
ORF35	**ß:** MAVP-RPIeRC_00042 **γ:** MAVP-QPI_00035	**ß:** 81 YES SVM **γ:** YES 72 SVM	**ß:** hypothetical protein **γ:** hypothetical protein	**ß:** -- **γ:** --	**ß:** -- **γ:** LCR
ORF81	**ß:** MAVP-RPIeRC_00087 **γ:** MAVP-QPI_00081	**ß:** YES 80 SVM **γ:** YES 87 SVM	**ß:** hypothetical protein **γ:** hypothetical protein	**ß:** -- **γ:** --	**ß:** -- **γ:** LCR

Prediction of T3SS2 effectors by pEffect..Gene names provided to the orthologs named strain RIMD2210633 VPaIα (NC_004605 region between VPA1312 and VPA1395), designated as α, strain MAVP-R VPaIß (MF066647.2), designated as ß, and strain MAVP-Q VPaIγ (MF066646.1), designated as γ. "+" = the presence of effector and predicted effector on VPaIα/ß/γ. "-" = the absence of effector and predicted effector on VPaIα/ß/γ. "~" = divergence between orthologs found in VPaIß/γ compared to the characterized effector in VPaIα. Basic Local Alignment Search Tool (BLASTx) and protein domains and motifs predicted from SMART (Simple Modular Architecture Research Tool)[209] and NCBI's Conserved Domain Database (CDD)[210] are detailed. LCR = Low Complexity Region; DUF = Domain of Unknown Function.

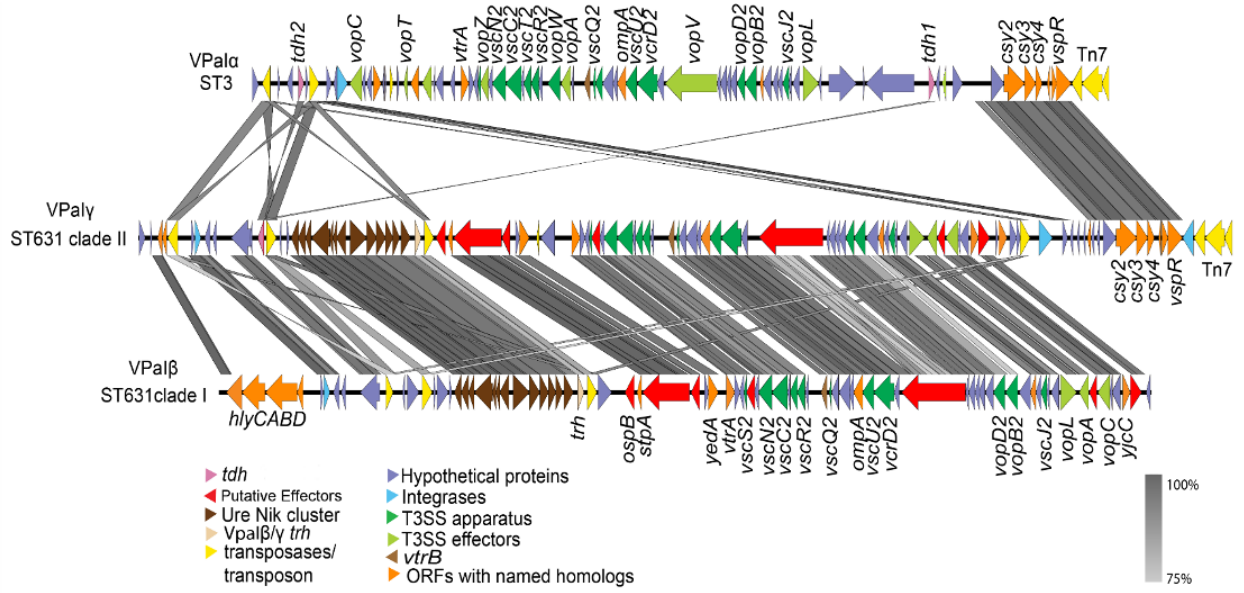


Figure 2.1. Comparison of content and conservation in representative VPaIα, VPaIβ, and VPaIγ. Named effectors from VPaIα (*vopA/C/L/V/Z* and VPA1380) and their homologs are light green. Candidate T3SS2β effectors (including divergent *vopV* and *vopZ* homologs) are red. The program pEffect and BLASTp searches predicted candidate effectors. Figure adapted from Xu et al. 2017 [107].

Protein domains and analogs predict function and support the identity of predicted T3SS effectors

Whereas one predicted effector, ORF32, was identified by homology to a characterized effector, the remaining four effectors, ORF34, 35, 76, and 81 were *de novo* predictions via Support Vector Machines (SVM) machine learning [208,211] (Table 2.1). Attempts to assign potential names to these four by identifying homologous proteins using in the NCBI non-redundant protein database identified the unnamed predicted effectors only as hypothetical proteins. Three predicted effectors (ORF32, 34, and 76) were homologous to named proteins present in other *Vibrio* spp. including *Vibrio cholerae*, *Vibrio harveyi*, *Vibrio diabolicus*, *Vibrio mimicus,* and *Vibrio campbellii*.

Among the predicted effectors, only ORF32 was initially identified based on sequence similarity-based inference by PSI-BLAST (Table 2.1). Though the identifying T3SS effector protein EspS was from *Citrobacter rodentium* (35.7% sequence identity, 94% query coverage, $3x10^{-40}$ E-value) further analysis revealed ORF32 was homologous to the OspB effector encoded by *Shigella spp.* [212] (35.2% sequence identity, 88% query coverage, $1x10^{-31}$ E-value), and more strikingly to a previously identified T3SS2 effector from VPaIα, VPA1380 (43.5%, 93% query coverage, E-value of $9x10^{-76}$) [166]. ORF32 and VPA1380 were not identified as homologs previously[107] based on the low deduced protein identities, but the high E-value supports that these are evolutionarily related. EspS helps maintain physiological balance during mouse infection by repressing colonic crypt hyperplasia [213]. OspB, in *Shigella* pathogens helps repress the host inflammatory response via down-regulation of IL-8 [214], and activates the NFkB pathway [215]. In contrast to OspB, which produced limited, stress-dependent toxicity in *Saccharomyces cerevisiae*, the OspB homolog VPA1380 (with 50% similarity) was toxic in yeast [166]. Thus, researchers predicted that VPA1380 might target different eukaryotic pathway than OspB during normal *S. cerevisiae* growth [166].

Next, to identify these putative effectors' functional domains, the architecture of the deduced proteins were analyzed by two protein domain prediction programs [209,210]. The deduced protein sequence of ORF76 had no conserved domains. The only potential domains in ORFs 32, 35, and 81 were Low Complexity Regions (LCRs), whose biased amino acid composition can lead to more predicted binding partners across different Protein-Protein Interaction (PPI) networks [216]. Some LCRs have identified roles, but the functions of most are not known [216–220]. Previous studies of repetitive sequences, such as those in LCRs, suggests that this genetic pattern enables rapid adaptation to new environments [216,218,221,222]. Although the LCRs in ORFs 32, 35, and 81 have no

further information, perhaps these domains provide a yet to be determined function. Of the five predicted T3SS effectors, pEffect assigned the highest probability score to ORF34. Searches for related proteins suggest ORF34 is most similar to proteins identified as sugar ATP-binding cassette (ABC) transporter substrate-binding protein. Predicted functional domains in ORF34 include a transmembrane helix region, a Pfam DUF (Domain of Unknown Function), and an M60-like domain (Figure 2.2). Bacteria secrete T3SS effectors into either the plasma membrane or directly into the cytosol of the target cell; thus, this transmembrane region of ORF34 could allow for effector binding to eukaryotic membranes [223]. M60-like domains typically possess mucinase activity, which could be beneficial in breaking down the protective mucin layer produced by enterocytes to allow for better bacterial adhesion [224]. Pfam DUF4092 is in other Proteobacteria such as *Escherichia coli*, *Vibrio cholerae*, *Vibrio fischeri*, and *Photobacterium iliopiscarium* [225]. Similarly to ORF34, 18 of the 20 known proteins with Pfam DUF4092 also contain the Peptidase M60-like domain downstream, and most are identified as Accessory colonization factor (AcfD) homologs. During infection, AcfD promotes efficient intestinal colonization in the human enteric pathogens *V. cholerae* [226] and Enterotoxigenic *E. coli* (ETEC) [227]. In *V. cholerae*, AcfD is a lipid anchor at the cell membrane, which also hints at the function of the predicted transmembrane helix region in ORF34. In ETEC, the AcfD homolog, YghJ, is a metalloprotease that influences intestinal colonization of ETEC by degrading the major mucins in the small intestine [227]. The similar functional domains as other enteric pathogens shed some light as to the possible function of ORF34. The close proximity (37bp apart) and directionality of two putative effector genes, ORF34 and ORF35, suggest these are likely co-transcribed under a single promoter's control. No transcriptional terminators were identified [228] within 300nts downstream of ORF35, but one was identified after ORF34.

Figure 2.2: ORF34 contains predicted domains that may influence successful colonization. ORF34 domain prediction from SMART (Simple Modular Architecture Research Tool) [209]. The blue rectangle indicates a transmembrane region. Pfam DUF4092 is in the Protein family "Domain of Unknown Function" 4092. The teal pentagon indicates the M60-like domain. The pink rectangle indicates a Low Complexity Region (LCR).

There were no predicted domains for ORF76, but its possible function may be gathered from its homology to known proteins. ORF76 was homologous to a conjugal transfer protein TraA of *V. parahaemolyticus* (100% sequence identity, 99% query coverage, $7x10^{-164}$ E-value), and several deduced TraA proteins from other members of the *Vibrionaceae* family. TraA is typically involved in pilus biosynthesis and assembly during conjugation [229]. Understandably, pEffect identified ORF76 as a potential T3SS effector since some T3SS apparatus proteins are identified as effectors by pEffect. For example, the translocator VopD2 (VPA1361) [230] and translocon protein VopW (VPA1345) [231] were both predicted by pEffect to be T3SS effectors via SVM (scores of 82 and 94 respectively). These proteins are part of the secretion needle and extend from the bacterial cell into its target, similarly to TraA's role as a pilin precursor protein [229]. Homology to TraA does not suggest that ORF76 has a toxin-related role during infection, but further studies such as secretion assays and animal studies would be necessary to make this determination. Interestingly though ORF76 as a was not previously identified as homologous to to the VPaIα effector VopO (VPA1329) based on sequenced identity (46.05% sequence identity, 99% query coverage), the strong E-value ($1x10^{-75}$) provides statistical support of evolutionary relatedness. During gastrointestinal infection with *V. parahaemolyticus*, VPA1329 contributes to host epithelial

disruption by activating the RhoA-ROCK signaling pathway, which induces actin stress fiber formation [145]. But, VPA1329 was not considered a primary effector during infection of rabbit illeal [145]. Due to the recent findings of the contribution of VopC in a mouse infection model [151], further investigation into ORF76 is necessary to understand its role fully. Unlike with ORF32, 34 and 76, searches for homologs provided little insight into the function of the remaining two effectors, ORF 35 and ORF81.

VPaIß and VPaIγ encode potential homologs of the VPaIα-encoded effectors critical for the enterotoxicity of pandemic strains, VopV (VPA1357) and VopZ (VPA1336). Previous analysis using Roary [232] indicated VopV and VopZ did not meet the 50% threshold used for identifying homologs between VPaI even though divergent ORFs located in the same location and gene neighborhoods in VPaIß and VPaIγ were still considered potentially analogous [11]. A closer examination and comparison of the deduced VopV protein in VPaIα (VPA1357) to the potential analog in VPaIγ (ORF61) indicates that even though they share only 40.3% protein sequence identity with 60% query coverage, the accompanying E-value of $6x10^{-95}$ provides strong statistical support of homology. Although the lower query coverage could be related to size differences of the deduced VopV proteins (VPA1357 at 4869bp and ORF61 at 7209bp), size alone does not account for this poor coverage. VopV encoded by VPA1357 possesses multiple domains, the interspersed or long repeat domains (rep1/2/3) and C-terminal domain, that are crucial to VopV induced enterotoxicity [139,141]. These domains bind to actin filaments and the actin-cross linking protein filamin [139]. Thereby bundling actin in the host cell leading to rearrangement of the apical epithelial cell membrane and elimination of the actin-filled microvilli brush border, leading to enhanced adhesion of *V. parahaemolyticus* to the host cell [139,141]. Alignment of VPA1357 and ORF61 proteins indicate that these important domains in VPA1557 are either not present in ORF61

or contain numerous amino acid composition changes. The large discrepancy in these important domains could indicate that ORF61 does not have the same enterotoxic capabilities as its characterized homolog VPA1357.

The VopZ analogs are relatively similar in size, with VopZ from VPaIα (VPA1336) 753bp and the potential VopZ analog in VPaIγ (ORF44) 768bp. Despite their similarity in size, VPA1336 and ORF44 have only 27% protein percent identity (71% query coverage) though again, the strong E-value ($1x10^{-15}$) indicates evolutionary relatedness. *V. parahaemolyticus* induced diarrhea and other intestinal pathology, but not colonization, *in vivo* relies on amino acids 38–62 in VPA1336 [147]. When the proteins are aligned, only 32% of this important region in the VPA1336 sequence is retained in ORF44. There is also a deletion of one amino acid and an insertion of four amino acids. This divergence in sequence in this important domain in ORF44 from the characterized VPA1336 might contribute to changes in intestinal pathology during *V. parahaemolyticus* infection. Previous studies indicate that truncation of VPA1336 after amino acids 38–62 resulted in the loss of *V. parahaemolyticus* colonization of the small intestine [147]. Since this region is important for *V. parahaemolyticus* colonization perhaps the large divergence in sequence from VPA1336 effectively removed this capability of ORF44. Since ORF61 and ORF44 on VPaIγ are so divergent from their enterotoxic orthologs on VPaIα, this suggests that further investigation is needed to assess which effectors contribute to *V. parahaemolyticus* pathogenicity.

Predicted effectors are toxic to eukaryotic cells

To examine putative effector toxicity towards eukaryotic cells, we used heterologous overexpression in the yeast *S. cerevisiae* and quantified growth inhibition following protein induction. Yeast is often used as part of T3SS effector characterization since they utilize conserved

eukaryotic processes that are typical targets of T3SS effectors, such as cell signaling and cytoskeletal dynamics, without immunity mechanisms that counteract T3SS effectors [233–235]. Due to this, while effector toxicity as inferred by growth attenuation by heterologous effector expression in yeast, does not directly recapitulate all aspects of enterotoxicity or disease, the model is an excellent initial screen to evaluate proposed effectors and gain initial insight into the role an effector might have on their eukaryotic target.

During intracellular overexpression of each putative effector, relative toxicity was quantified by their ability to inhibit yeast growth compared to negative (empty vector) and positive (VopA) controls. When grown under inducing conditions, the ORF32 construct did not reduce yeast growth, similarly to the yeast harboring the empty vector (Figure 2.3). In contrast, induction of ORF35 and ORF81 dramatically reduced yeast growth. Expression of ORF34 and ORF76 moderately attenuated yeast growth. In congruence with previous studies [124], the VopA homolog inhibited yeast growth, the VopC homolog did not attenuate yeast growth. Finally, VopL was not previously evaluated for yeast toxicity, and its homolog from VPaIγ did not inhibit yeast in our study.

Even though a key VPaIα effector, VopZ is highly divergent from gene products in the same location in VPaIß/γ, these were both identified as putitive effectors and may be analogous toxins, and thus both were evaluated for toxicity. Neither VopZ analog from VPaIα (VPA1336) nor VPaIγ (ORF44) inhibited yeast growth. During *V. parahaemolyticus* infection, VopZ from VPaIα promotes colonization by inhibiting the activation of the kinase TAK1, which disrupts cell signaling cascades leading to phenotypes such as inflammation, distorted epithelial structure, and mislocalization of tight junction proteins [125,147,163]. Yeast do not exhibit these phenotypes, thus this model cannot appropriately evaluate the effect of these proteins. While this does limit this model's

effectiveness for evaluating some characteristics of epithelial disease, the wide breadth of effectors that the model can evaluate, such as those that alter the cytoskeleton and organelles, target MAPK-signaling cascades, or perturb vesicle trafficking warrants this as an effective preliminary assay [234]. Other assays useful in evaluating effectors whose effect cannot be visualized in yeast include more laborious assays such as Western Blot analysis, cell-culture, and animal models. Despite the minimal growth reduction conferred by ORFs 32 and 34, these effectors could still contribute to disease if their eukaryotic target does not confer a similar phenotype in yeast, similar to the finding of VopZ. Other models including cytotoxicity [122] and brush border remodeling[139] of cultured human colon cells, and animal infection models such as the rabbit ileal loop model [98,138] or a mouse infection model[151] are needed to address this issue. We did not evaluate VopV in this model due to its large size and the repetitive nature of the sequence that thwarted our attempts to clone the gene.

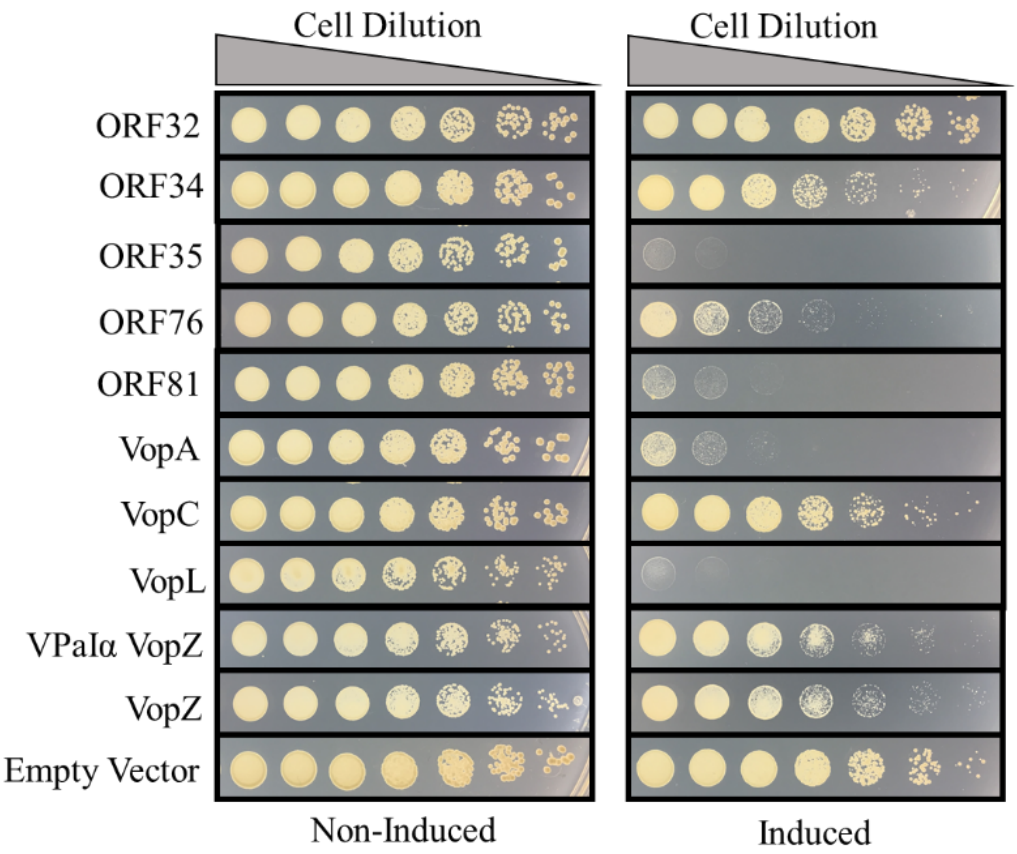


Figure 2.3: Heterologous expression of putative effectors inhibit yeast growth. Serial dilutions of yeast transformed with pYES2.1 plasmid harboring putative toxin (from VPaIγ or VPaIα) were grown under non-inducing (glucose) and inducing (galactose) conditions. Data is representative of three experiments performed in triplicate. All effectors except VPaIα VopZ were constructed from ST631 clade II MAVP-Q harboring VPaIγ genomic DNA. VPaIα VopZ template DNA was ST3 MAVP-C harboring VPaIα genomic DNA.

Predicted effectors induce cytotoxicity in cultured human colon cells

Next, we assessed T3SS2-dependent cytotoxicity of each putative effector against cultured Caco-2 colorectal tumor cells. Caco-2 cells were co-cultured with wild-type (wt) *V. parahaemolyticus* and with mutant derivatives altered for T3SS ability. After co-culture, the cytotoxicity of the Caco-2 cells was quantified (Figure 2.4). Effectors secreted by T3SS1 contribute to cytotoxicity towards Caco-2 cultured cells, but are unimportant for enterotoxicity and

can mask the effect of T3SS2 effectors. For this reason, the *vscN*1 gene encoding the ATPase powering secretion through the T3SS1 apparatus was deleted to disable T3SS1 effector secretion prior to the generation of individual effector gene deletions. We included this derivative to eliminate background cytotoxicity, allowing us to evaluate each putative effector's contribution to cytotoxicity. All *V. parahaemolyticus* strains were cultured in 0.04% bile before co-incubation with Caco-2 cells to simulate gastrointestinal conditions and induce differential regulation of virulence genes [236–238]. However, a recent study indicates that genes on VPaIγ in strain 04-1290 were not induced by bile, though the assertion that strain 04-1290 containing VPaIγ did not contain bile-inducible VtrABC was not correct [107,238,239]. In our study, *V. parahaemolyticus* growth in bile increased cytotoxicity towards Caco-2 cells (data not shown), and thus bile was used for all experiments. Elimination of VscN1 (Δ*vscN*1) did not significantly reduce cytotoxicity under the conditions of our assays (Figure 2.4), as expected from previous data [9]. This importantly alludes to the substantial impact of toxic effectors from T3SS2 contributing to the cytotoxicity of Caco-2 cells, and the utility of the model. This is further borne out by the elimination of T3SS2 in the T3SS1 mutant background, which significantly impaired cytotoxicity, reflective of an inability of these bacteria to secrete T3SS2 toxins, again replicating previously published work [9]. Following this validation, we subsequently eliminated predicted T3SS2 effectors, as well as ORF61 (VopV) and ORF44 (VopZ), and evaluated their impact on cytotoxicity. Notably, we did not generate individual ORF34 or ORF35 deletions because we predicted that these ORFs are cotranscribed under the control of a shared promoter, thus both ORFs were eliminated simultaneously generating the mutant designated ΔORF34-35 .

Elimination of the OspB/VPA1380 homolog ORF32 did not attenuate cytotoxicity, however, elimination of any of the remaining putative toxins, or the combined elimination of

ORF34 and ORF35, significantly reduced cytotoxicity ($P > 0.01$) compared with either wild type or the parental T3SS1 mutant. Taken together, these results along with the ability of these toxins to inhibit yeast growth suggest that putative effectors ORFs 34-35, 76, and 81 are four of the genes responsible for T3SS2-dependent cytotoxicity, along with the *vopV* and *vopZ* analogs (ORFs 61 and 44 respectively). Interestingly eliminating any singular putative toxin appears to reduce the cytotoxicity to similar levels as preventing the secretion of all T3SS2 effectors. Since VPaIα effectors work in concert to achieve their enterotoxic effect, we believe that our cytotoxicity model, as a surrogate for enterotoxicity, could work similarly, in that deletion of one effector affects the ability of other effectors to carry out their function.

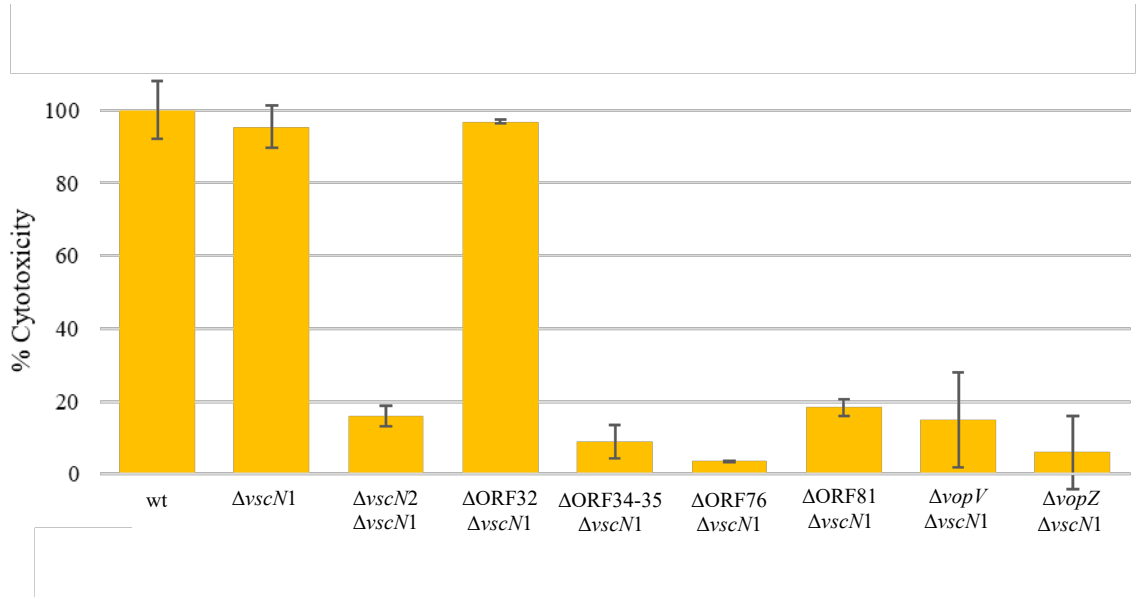


Figure 2.4. Reduced cytotoxicity towards Caco-2 cells mirrors the difference in toxicity towards yeast. The human colorectal tumor cell line, Caco-2 cells, were co-incubated with *V. parahaemolyticus* strains defective in T3SS1, and harboring indicated ORF deletions. We predict ORFs 34 and 35 are co-transcribed and thus deleted both simultaneously. Percent cytotoxicity, estimated by the quantity of lactate dehydrogenase (LDH) in the well after co-culture, was calculated by the following equation: [optical density at 490 nm (OD_{490}) of experimental release - OD_{490} of spontaneous release]/(OD_{490} of maximum release - OD_{490} of spontaneous release)*100%. Spontaneous release is the amount of LDH released from the cytoplasm of uninfected cells. The maximum release is the total amount of LDH released after the complete lysis of uninfected cells. * = $p < 0.001$ Tukey Kramer HSD for pair-wise comparisons. Data represents the mean ± standard deviation from one experiment using five individual bacterial colonies per strain, performed in triplicate. The experiment was repeated twice with similar results.

Conclusion

Critical forces on microbial pathogen's evolution are gene gain or loss, and recombination of acquired genes [32,55,57–62]. Acquisitions of genetic material, such as virulence effectors, can provide a fitness advantage beneficial to bacteria under certain conditions such as adaptation to changing environmental conditions, enhancing virulence, or immune evasion [32,34,56,57,64–66]. Virulence effectors can also undergo strong selection for either increasing allelic variation, maintaining a beneficial allele, and removing deleterious alleles [240]. Diversity in effector content can be beneficial for bacteria, because each toxin has a particular function. While some may be more important than others in a given environment, they all have a specific role that may benefit the bacteria, especially when they work in concert [112,240,241].

In addition to confirming the presence of homologs of four named effectors in VPaIα (VopV, VopZ, VopO and VPA1380) that eluded prior discovery, we also identified three T3SS2β effectors encoded in VPaIγ and β without homologs in VPaIα, all of which may contribute to disease caused by the most clinically prevalent strains in North America, including ST631 and ST36. The protein domains and cytotoxicity induced by ORF34 suggest it could be an important effector and promote enterotoxicity. The transmembrane anchoring, mucinase activity, and enhanced colonization associated with these domains characterized in other enteric pathogens suggest that ORF34 may contribute these same characteristics to *V. parahaemolyticus* during infection. Mucus and mucins are the first line of defense for enterocytes [242,243]; thus, mucin-degrading capabilities by *V. parahaemolyticus* would be extremely beneficial to infection success. These predicted activities may also explain the lack of toxicity in yeast. The likely co-transcribed protein ORF35 was toxic to yeast and likely contributed to Caco-2 cytotoxicity; however, individual gene deletions are needed to identify if cytotoxicity was due to the synergistic effects

of ORF34 and ORF35, or just ORF35. Caco-2 cells do not secrete mucins [244,245], thus if mucin-degradation is the primary role of ORF34, then the cytotoxic effect seen can be attributed to ORF35. The quantification of mucus degradation after co-culture with a mucus-producing cell line such as HT29-MTX could assess this hypothesis [244–246]. The cytotoxicity of ORF76 and ORF81 towards both yeast and human colon cells indicates that these effectors are toxic toward cultured human colon cells, suggesting a potential role during infection in humans. ORF32 was the only effector predicted by pEffect based on nucleotide sequence similarity to other known T3SS effectors, but it did not inhibit yeast growth and is not cytotoxic towards cultured human colon cells. Thus, we believe other models for ORF32's characterization are more appropriate, and further studies are needed. These effector's impact in the models utilized lays important foundational work towards virulence effector identification and characterization in VPaIγ. To elucidate the contribution these effectors have towards disease, infection analyses such as human colon cell brush border effacement [139], or gastrointestinal colonization and pathogenesis studies in infant rabbits[138,139] or mice [151] are necessary.

Identification of the virulence effectors in the prominent pathogenic *V. parahaemolyticus* lineages in North America could aid in monitoring local populations. This, combined with *tdh* and *trh* monitoring, can be translated into sensitive surveillance tools to identify the most important emergent lineages, which can prevent negative publicity and reactive harvest closure of local businesses [247–249]. These findings can also inform therapeutic interventions and novel vaccines beneficial for individuals with an increased risk for infection [250]. This study identifies *V. parahaemolyticus* T3SS effectors on VPaIβ and VPaIγ, but it does not fully characterize each effector's role during gastrointestinal infection. Though we do identify effectors encoded by VPaIβ and VPaIγ, there are different lineages of VPaIγ that have evolved in geographically distinct

populations, and strains that harbor these variants differ in clinical prevalence [11]. T3SS effectors on VPaIα are essential for disease, thus since there is variation in clinical prevalence among VPaIγ, perhaps this is due to variation in their T3SS effectors. Further studies are needed to examine whether the variation in these five T3SS effectors on VPaIβ and VPaIγ correlate with clinical prevalence of their harboring strains.

Materials and Methods

Bacterial and Yeast Strains, Media and Culture Conditions

The bacterial strains and plasmids used in this study are described in Table 2.2 and Table 2.3, respectively. *V. parahaemolyticus* was routinely grown at 37°C in Luria-Bertani medium supplemented with 2% NaCl (LBS) [251] or in defined minimal media [252] supplemented with 0.0058% K_2HPO_4, 0.1% NH_4Cl, 0.01mM $FeSO_4$, 0.1 mM Tris pH 7.4, 0.6% glycerol, 1x Artificial Sea Water, 0.625µg/mL chloramphenicol, and tap water for trace minerals for transformation. *E. coli* strains were grown at 37°C in Luria-Bertani medium (LB) as previously described [251]. Antibiotics were supplemented for selection of mutations and plasmids at the following concentrations for *V. parahaemolyticus*: 12 µg/ml gentamicin (Gen), 2.5 µg/ml chloramphenicol (Chl), and 5 µg/ml erythromycin (Erm) and at the following concentrations for *E. coli*: 50 µg/ml kanamycin (Kan), 100 µg/ml ampicillin (Amp), and 25 µg/ml chloramphenicol (Chl).

S. cerevisiae BY4742 (obtained from Clyde Denis, [253]) was maintained in yeast extract (1%)-peptone (2%)-and dextrose (2%) medium (YPD) [253], and grown on Synthetic Complete Minimal Media (SC-MM) made with Yeast Synthetic Drop-out Medium Supplements without uracil (Sigma-Aldrich), supplemented with glucose at 2% once transformed with pYES2.1/V5-His-TOPO®. *S. cerevisiae* was grown at 28°C.

Bioinformatic Analyses

The annotated VPaIs, including VPaIα (from strain RIMD2210633, NC_004605 region between VPA1312 and VPA1395), VPaIß (from strain MAVP-R, MF066647.2) and VPaIγ (from strain MAVP-Q, MF066646.1), were analyzed by the pEffect online algorithm [208]. The predicted T3SS ORFs from pEffect were then submitted individually to the following programs according to their

normal operating instructions: Basic Local Alignment Search Tool (BLASTx) [254], SMART (Simple Modular Architecture Research Tool) [209], and NCBI's Conserved Domain Database (CDD) [210]. To find potential terminators, the 300 nucleotide region following the stop codon of ORF35 and ORF34 was submitted to the FindTerm (Softberry Inc.) [228] webpage application set to standard operating instructions with the "All putative terminators" box checked.

Molecular genetic methodologies and plasmid construction

Strain Construction Overview

Targeted deletion mutants of *V. parahaemolyticus* strains were constructed through the process of PCR Splicing by Overlap Extension (SOE) [255] wherein genes of interest were eliminated and replaced with an antibiotic resistance cassette, the resulting constructs were cloned into plasmid pCR2.1 TOPO (TOPO™ TA Cloning™ Kit, Thermo Fisher), which was subsequently used for natural transformation. PCR SOE primers have 3' complementary ends to the products intended to fuse together (Table 2.2). During PCR cycling, the ends overlap and are extended into a single molecule.

PCR SOE and cloning

Primers for generating constructs used in marker exchange mutagenesis were designed as recommended [255]. Oligonucleotide primers were synthesized by © Integrated DNA Technologies (Coralville, IA). PCR cycling conditions were optimized using AccuStart II PCR SuperMix (Quantabio, Beverly, MA). Genomic DNA from *V. parahaemolyticus* strain MAVP-Q, a ST631 clade II member containing VPaIγ, from VPaIγ, and strain MAVP-C, a pandemic ST3 strain containing VPaIα,was used as a template to generate amplicons upstream and downstream of the

gene of interest, whereas amplicons of the antibiotic cassettes were generated using plasmids pBD4 and pEVS170 (Table 2.4). Genomic DNA was prepared by organic DNA extraction [256]. Oligonucleotide primers for gene knockouts were designed to contain at least 400bp regions flanking the gene of interest to enable efficient recombination (Table 2.3). Oligonucleotide primers for the antibiotic cassette to replace the gene of interest were designed to contain the native antibiotic resistance cassette promoter from the plasmid (Table 2.3). Analysis of the secondary structure of primers was performed by NetPrimer (©PREMIER Biosoft). Gene knockouts and antibiotic resistance genes were amplified from target DNA with Phusion® High-Fidelity DNA Polymerase or Phusion® High-Fidelity PCR Master Mix (©New England Biolabs, Ipswich, MA) according to ©New England Biolabs Inc. protocol. PCR SOE was performed as described [255]. PCR SOE product was cloned into pCR2.1TOPO® and transformed into chemically competent *E. coli* 10-beta cells (New England Biolabs®, Ipswich, MA) following Invitrogen pCR2.1TOPO® (Invitrogen K202020) protocol. Plasmid DNA was purified using ©QIAGEN Plasmid Kit protocol with house-made ©QIAGEN reagents. Confirmation that the constructs contained no PCR generated errors was determined by sequencing using the Sanger method by ©GENEWIZ LLC (South Plainfield, NJ, USA).

Competency Plasmid Construction

To enhance the competency of *V. parahaemolyticus* enabling the use of natural transformation for strain construction, we designed and constructed a plasmid to express the native *tfoX* gene from *V. parahaemolyticus* strain MAVP-Q under the control of the arabinose-inducible and glucose-repressible promoter P_{BAD} along with the gene encoding the arabinose operon regulatory protein *araC* gene (*araC*-P_{araBAD}-*tfoX*) into shuttle vector pEVS79, generating pSEE1. Briefly, primers

were designed to clone the *araC* and its native promoter and a downstream P_{araBAD} promoter from pSW7848, and in parallel to amplify the native *V. parahaemolyticus tfoX* from MAVP-Q, and the two amplicons were fused using PCR SOE [255]. The resulting construct was cloned into pCR2.1TOPO plasmid, which was propagated in *E. coli* One-Shot TOP10. Purified plasmids were sequenced to confirm the constructs were correct and error-free by Sanger sequencing at ©GENEWIZ LLC (South Plainfield, NJ, USA). The *araC*-P_{araBAD}-*tfoX* cassette was subsequently cloned from the resulting plasmid (pSEE21) into pEVS79 [257] by restriction digest with SpeI and XhoI (©New England Biolabs, Ipswich, MA) to move the cassette from pSEE21 to pEVS79 [257], resulting in the final plasmid pSEE1. pSEE1 was conjugated into *V. parahaemolyticus* using Tri-Parental Mating with the helper plasmid pEVS104 as described [257]. Briefly equimolar concentrations of mid-log phase *E. coli* harboring pSEE1, *E. coli* harboring pEVS104, and recipient *V. parahaemolyticus* were individually washed in LBS and resuspended together in 20 µL LBS. The entire suspension was spotted onto LBS agar and grown overnight at 37°C. The following day entire spot was collected, diluted, plated onto LBS+Chl2.5, and grown at 28°C overnight to select for *V. parahaemolyticus*.

Transformation of *V. parahaemolyticus*

V. parahaemolyticus strains harboring pSEE1 were grown to turbidity in LBS and subsequently a 2% inoculum was used to grow an overnight culture at RT in minimal medium supplemented with 0.2% arabinose. Once cultures reached $OD_{600} > 0.5$, 0.5 mL culture was mixed with 2 µg of either plasmid DNA or ~10 µg of genomic DNA that was sheared by freeze-thaw three times at -20°C, and incubated at 28°C statically for 30 minutes. 0.5 mL LBS was added to the culture and was incubated at 37°C for 20-60 minutes. Cells were pelleted and suspended in ~50 µl LBS broth.

Equal volumes of transformant were plated onto three concentrations of antibiotic flanking estimated optimal concentration (10/12/15 μg/ml gentamycin and 2.5/5/8 μg/ml erythromycin) and grown at 37°C overnight. Colonies were re-streaked to a plate with a higher concentration of antibiotic, grown at 37°C, then the colony screened by PCR to identify the presence of the mutation. Mutants harboring pSEE1 were passaged at a 2% inoculum in LBS twice per day for two days to lose plasmid.

Cytotoxicity Assay

Eukaryotic cell culture

Caco-2 cells were regularly maintained in Dulbecco's modified Eagle's medium (DMEM, ATCC®) containing 10% (v/v) Fetal Bovine Serum (FBS, Sigma®) and Corning® Antibiotic-Antimycotic Solution at 37°C in 5% CO_2.

Cytotoxicity measurements

Caco-2 cells at 80% confluency were washed in DMEM without phenol red + 10% FBS and seeded 25μL of cells at 315 cells/μl in 384-well plates and cultured for 48hr at 37°C in 5% CO_2 to confluency. Unmodified *V. parahaemolyticus* strains were streaked for isolation onto LBS agar, Δ*vscN1*:Gen *V. parahaemolyticus* mutants were streaked for isolation onto LBS + 5 μg/ml Gen, and Δ*vscN1*:Gen/Δ(gene of interest):Erm *V. parahaemolyticus* mutants were streaked for isolation onto LBS + 5 μg/ml Erm. All *V. parahaemolyticus* strains were grown at 37°C until robust colony growth. Five colonies of *V. parahaemolyticus* were grown overnight in LBS + 0.04% bile salts at 37°C in a roller drum. Cultures were washed twice in PBS and diluted to add 5μL of culture to 48hr Caco-2 cells at an MOI of 10. Three replicates per sample were added to Caco-2 cells. After

6hrs of co-culture, the release of lactate dehydrogenase (LDH) into the medium was quantified by using a CytoTox96® non-radioactive cytotoxicity kit (Promega®) according to the manufacturer's instructions. To quantify percent cytotoxicity, the LDH release was calculated with the following equation: [optical density at 490 nm (OD_{490}) of experimental release - OD_{490} of spontaneous release]/(OD_{490} of maximum release - OD_{490} of spontaneous release)*100%. Spontaneous release is the amount of LDH released from the cytoplasm of uninfected cells, whereas the maximum release is the total amount of LDH released after the complete lysis of uninfected cells.

Statistical Analysis

Data from Caco-2 cytotoxicity assay represents the mean ± standard deviation from one experiment using five individual bacterial colonies per strain, performed in triplicate. Pairwise comparison across strains was performed utilizing Tukey Kramer HSD.

Table 2.2: Strains used in this study

Strain	Description	Source of Reference
E. coli strains		
DH5α	pEVS79	[257]
CC118l*pir*	pEVS104	[257]
TOP10	pSEE1	This study
One-Shot TOP10F'		Thermo Fisher®
10-beta cells		New England Biolabs®
V. parahaemolyticus strains		
MAVP-Q	WT	[22]
TBD	Δ*vscN1*:Gen	This study
TBD	Δ*vscN1*:Gen Δ*vscN2*:Erm	This study
TBD	Δ*vscN1*:Gen ΔORF32:Erm	This study
TBD	Δ*vscN1*:Gen ΔORF34-35:Erm	This study
TBD	Δ*vscN1*:Gen ΔORF85:Erm	This study
TBD	Δ*vscN1*:Gen ΔORF90:Erm	This study
TBD	Δ*vscN1*:Gen Δ*vopV*:Erm	This study
TBD	Δ*vscN1*:Gen Δ*vopZ*:Erm	This study
S. cerevisiae strains		
BY4742	his3-Δ1 leu2-Δ0 ura3-Δ0 lys2-V0	[253]

Gen = gentamycin; Erm = erythromycin

Table 2.3: Oligonucleotides used in this study.

Primer name	Target Gene	DNA template	Sequence
Yeast Toxicity Assay			
32YES FRBS2	ORF32	MAVP-Q gDNA	AAAATGGCTATCTCTTTAACCGGATGTTT
32YES R	ORF32	MAVP-Q gDNA	GGTCAGAGCCTCATTGTTGATATTAGTCC
PrSEE52 34YES FRBS2 631_674	ORF34	MAVP-Q gDNA	TTGATGGTTTGTTTTTCAGGTTTTTTC
PrSEE53 34YES R2 631_674	ORF34	MAVP-Q gDNA	TACAACTAGTCTCCAGCCTAATGTACCGT
35YES FRBS	ORF35	MAVP-Q gDNA	GAAATGGGTATTATGATTCCTACATTCAACAGAG
35YES R	ORF35	MAVP-Q gDNA	CTAAAATCAAATCCTTTCTATCAATGTTCACT
85YES FRBS	ORF76	MAVP-Q gDNA	ACTATGGCTTTAAATAAAATAAACCCTATTCAGT
85YES R	ORF76	MAVP-Q gDNA	CAGGAAACTCTACTAACACAAAAGAAGCC
90YES FRBS	ORF81	MAVP-Q gDNA	AAAATGGTAATACAAAGCCAAAGACTGAGA
90YES R	ORF81	MAVP-Q gDNA	TTACCATATAATACCATGTTCACGAAGTATC
VopAYES FRBS	*vopA*	MAVP-Q gDNA	GACATGGATGTCGATAGTAAAGCAGGACC
VopAYES R	*vopA*	MAVP-Q gDNA	AGAGAAGTCACTCACTATTCACACCGCA
VopCYES FRBS	*vopC*	MAVP-Q gDNA	AACATGGTAGAATTAAAGCAAATATTTA
VopCYES R	*vopC*	MAVP-Q gDNA	CTACGCTAATTTGACTACTTTACTTTG
VopLYES FRBS	*vopL*	MAVP-Q gDNA	AGACGATGGTTAAGTCAACCTTTA
VopLYES R	*vopL*	MAVP-Q gDNA	TTAAGACAATTTTGCTGCCA
VopZYES FRBS	*vopZ* - VPaIα	MAVP-C gDNA	ATTATGGCTGGGTATACTGATGTAAAACCGT
VopZYES R	*vopZ* - VPaIα	MAVP-C gDNA	ATCAAGAACTGTCATGGCTTTCCTCTA
VopZYES FRBS	*vopZ* - VPaIγ	MAVP-Q gDNA	ATTATGGCTGGGTATACTGATGTAAAACCGT
VopZYES R	*vopZ* - VPaIγ	MAVP-Q gDNA	ATCAAGAACTGTCATGGCTTTCCTCTA
VscNYES FRBS	*vscN2*	MAVP-Q gDNA	GTGATGGTACAGCACTATTTTAAGGTCAAGGA
VscNYES R	*vscN2*	MAVP-Q gDNA	TATATCAATAAGCTGACCAAATTCTCTCCAT
Competency Plasmid			
PrSEE34 araC SOE AF	*araC*	pSW7848	TGCTCTGCGAGGCTG
PrSEE35 araC SOE AR	*araC*	pSW7848	TTGACCTCTTTAATCTGAATAACGGGTATGGAGAAACA
PrSEE36 tfox SOE BF	*tfoX*	MAVP-Q gDNA	TGTTTCTCCATACCCGTTATTCAGATTAAAGAGGTCAA
PrSEE37 tfoX SOE BR	*tfoX*	MAVP-Q gDNA	AACGATGACGATTGGACA

Genetic Knockouts			
Puskar's vscN1 AF	*vscN1*	MAVP-26 gDNA	ATTTTCTGTTCTATTGCCACC
PrSEE10 vscN1-GmAR2	*vscN1*	MAVP-26 gDNA	AGGCGACAAGGTGCTGAATAGAAAGGATGG
PrSEE13 GmF pBD4	Gen^R cassette	pBD4 plasmid prep	TCAGCACCTTGTCGCCT
PrSEE14 GmR pBD4	Gen^R cassette	pBD4 plasmid prep	AATTGTTAGGTGGCGGTA
PrSEE9 VSCN1-GmBF2	Gen^R cassette	PCR product	CCATCCTTTCTATTCAGCACCTTGTCGCCT
PrSEE11 VscN1-GmBR2	Gen^R cassette	PCR product	TGGGCACGAGCAAATTGTTAGGTGGCGGTA
PrSEE12 VscN1-GmCF2	*vscN1*	MAVP-26 gDNA	TACCGCCACCTAACAATTTGCTCGTGCCCA
Puskar's vscN1 CR	*vscN1*	MAVP-26 gDNA	CTTTGGTTTTATGGGCTTG
PrSEE69 631_vscN2_Erm AF	*vscN2*	MAVP-Q gDNA	GAACGCCCGATGAAGT
PrSEE71 vscN2_Erm AR	*vscN2*	MAVP-Q gDNA	GTTTCCGCCATTCTTTGGTGTCTTGATATGACTCCTT
PrSEE72 vscN2_Erm BF	Erm^R cassette	MAVP-Q gDNA	AAGGAGTCATATCAAGACACCAAAGAATGGCGGAAAC
PrSEE73 vscN2_Erm BR	Erm^R cassette	MAVP-Q gDNA	AGACAGCCGAACGTACTCCTTTACAAAAGCGACTCATAGA
PrSEE74 vscN2_Erm CF	*vscN2*	MAVP-Q gDNA	TCTATGAGTCGCTTTTGTAAAGGAGTACGTTCGGCTGTCT
PrSEE75 vscN2_Erm CR	*vscN2*	MAVP-Q gDNA	TGATACCCAGTCATAATAAATG
PrSEE76 32_Erm AF	ORF32	MAVP-Q gDNA	AGGTATTACTCAAGAAGGACTGT
PrSEE77 631 32_Erm AR	ORF32	MAVP-Q gDNA	GTTTCCGCCATTCTTTGAGTGTTAGTTGTGTTATGTTGTT
PrSEE78 631 32_Erm BF	Erm^R cassette	pEVS170	AACAACATAACACAACTAACACTCAAAGAATGGCGGAAAC
PrSEE79 631 32_Erm BR	Erm^R cassette	pEVS170	GTCCACTTCAGAATACATTGTTTACAAAAGCGACTCATAGA
PrSEE80 631 32_Erm CF	ORF32	MAVP-Q gDNA	TCTATGAGTCGCTTTTGTAAACAATGTATTCTGAAGTGGAC
PrSEE81 631 32_Erm CR	ORF32	MAVP-Q gDNA	AGACACCAGCACTCACGA
ORF34.5SoAF	ORF34-35	MAVP-Q gDNA	ATAACAACGGTTGATTTAATACTCTT
ORF34.5SoAR	ORF34-35	MAVP-Q gDNA	GTTTCCGCCATTCTTTGCTAACTCTGTTGAATGTAGG
ORF34.5SoBF	Erm^R cassette	pEVS170	CCTACATTCAACAGAGTTAGCAAAGAATGGCGGAAAC
ORF34.5SoBR	Erm^R cassette	pEVS170	GTTCTAACCATAGGACTGATTTACAAAAGCGACTCATAGA
ORF34.5SoCF	ORF34-35	MAVP-Q gDNA	TCTATGAGTCGCTTTTGTAAATCAGTCCTATGGTTAGAAC
ORF34.5SoCR	ORF34-35	MAVP-Q gDNA	ACCTTGCTCCTTAATTTGCCT
PrSEE41 ORF85 AF	ORF76	MAVP-Q gDNA	TGCCATTTACCTGTGACAT

PrSEE42 ORF85 AR	ORF76	MAVP-Q gDNA	GTTTCCGCCATTCTTTGACTACGCTAATTTGACTACTT
PrSEE43 ORF85 BF	Erm^R cassette	pEVS170	AAGTAGTCAAATTAGCGTAGTCAAAGAATGGCGGAAAC
PrSEE44 ORF85 BR	Erm^R cassette	pEVS170	TTCGCATCTGATTCAACTTTACAAAAGCGACTCATAGA
PrSEE45 ORF85 CF	ORF76	MAVP-Q gDNA	TCTATGAGTCGCTTTTGTAAAGTTGAATCAGATGCGAA
PrSEE46 ORF85 CR	ORF76	MAVP-Q gDNA	CATACAATGAATACCTCCACA
G88SoAF2	ORF81	MAVP-Q gDNA	TGCCCATTCAAATACGCT
G88SoAR	ORF81	MAVP-Q gDNA	GTTTCCGCCATTCTTTGTATTAACATTTTTATCCG
G88SoBF	Erm^R cassette	pEVS170	CGGATAAAAATGTTAATACAAAGAATGGCGGAAAC
G88SoBR	Erm^R cassette	pEVS170	GGATGTATTTCGGGTTGTTTACAAAAGCGACTCATAGA
G88SoCF	ORF81	MAVP-Q gDNA	TCTATGAGTCGCTTTTGTAAACAACCCGAAATACATCC
G88SoCR2	ORF81	MAVP-Q gDNA	GTTATTGACACTGATAGTGGCTT
QVopVSoAF	*vopV*	MAVP-Q gDNA	CGATGACTCCCGATTTCCA
QVopVSoAR	*vopV*	MAVP-Q gDNA	GTTTCCGCCATTCTTTGCTATTTATCATATTAAAACCTCA
QVopVSoBF	Erm^R cassette	pEVS170	TGAGGTTTTAATATGATAAATAGCAAAGAATGGCGGAAAC
QVopVSoBR	Erm^R cassette	pEVS170	CGGAATCTCATAATCAGACTTTACAAAAGCGACTCATAGA
QVopVSoCF	*vopV*	MAVP-Q gDNA	TCTATGAGTCGCTTTTGTAAAGTCTGATTATGAGATTCCG
QVopVSCR	*vopV*	MAVP-Q gDNA	ATCATATAGACTCCAAGACACCTC
Q44SoAF	*vopZ* VPaIγ	MAVP-Q gDNA	AGGAGTACGTTCGGCTGTCT
Q44SoAR	*vopZ* VPaIγ	MAVP-Q gDNA	GTTTCCGCCATTCTTTGATACCCAGTCATAATAAATG
Q44SoBF	Erm^R cassette	pEVS170	CATTTATTATGACTGGGTATCAAAGAATGGCGGAAAC
Q44SoBR	Erm^R cassette	pEVS170	GAAATACTCACCATCGAGTTTTACAAAAGCGACTCATAGA
Q44SoCF	*vopZ* VPaIγ	MAVP-Q gDNA	TCTATGAGTCGCTTTTGTAAAACTCGATGGTGAGTATTTC
Q44SoCR	*vopZ* VPaIγ	MAVP-Q gDNA	AGTGCTAATCCTGTTCAAGTGG

Erm = erythromycin; Gen = gentamycin

Table 2.4: Plasmids used in this study

Plasmid	Description	Relevant characteristics	Reference
Competency Plasmid			
pEVS79	moderate suicide vector in *V. parahaemolyticus*	*V. fischeri* suicide cloning vector, Chl^R, Tet^R	[257]
pSW7848	pBAD promoter	Chl^R, R6K γ-*ori*-based suicide vector, ccdB toxin gene under the control of an arabinose-inducible and glucose-repressible promoter PBAD.	[258]
pSEE1	competency plasmid, moderate suicide vector in *V. parahaemolyticus*	Chl^R	This Study
pSEE21	pCR2.1-TOPO: *araC-*P_{araBAD}*-tfoX*	Kan^R	This Study
Genetic Knockouts			
pBD4	Gen cassette	Erm^R	[259]
pEVS170	Erm cassette	Chl^R, mini-Tn5-Erm, oriVR6K oriTRP4 Kan^R	[260]
pCR2.1-TOPO	cloning vector	Kan^R	Thermo Fisher Scientific
pEVS104	helper plasmid for TPM	Conjugative plasmid with *tra* and *trb* genes	[257]
pSEE2	pCR2.1-TOPO:ST631vscN1::Gen *vscN1* gene knockout	Kan^R	This Study
pSEE3	pEVS79:ST631vscN1::Gen *vscN1* gene knockout on moderate suicide vector	Chl^R, Gen^R	This Study
pSEE4	pCR2.1-TOPO:ST631vscN2::Erm ST631 vscN2 knockout	Chl^R, Erm^R	This Study
pSEE5	pCR2.1-TOPO:ST631ORF32::Erm ST631 ORF32 knockout	Chl^R, Erm^R	This Study
pSEE6	pCR2.1-TOPO:ST631ORF34-35::Erm ST631 ORF34/ORF35 knockout	Chl^R, Erm^R	This Study
pSEE7	pCR2.1-TOPO:ST631ORF76::Erm ST631 ORF76 knockout	Chl^R, Erm^R	This Study
pSEE8	pCR2.1-TOPO:ST631ORF81::Erm ST631 ORF81 knockout	Chl^R, Erm^R	This Study
pSEE9	pCR2.1-TOPO:ST631ORFVopV::Erm ST631 VopV knockout	Chl^R, Erm^R	This Study

pSEE10	pCR2.1-TOPO:ST631ORFVopZ::Erm	ST631 VopZ knockout	Chl^R, Erm^R	This Study
Yeast Toxicity Assay				
pYES2.1-V5-His-TOPO	cloning vector		GAL1 promoter	Thermo Fisher Scientific
pSEE11	pYES2.1-V5-His-TOPO:ST631ORF32	ST631 ORF32 expression vector	GAL1 promoter	This Study
pSEE12	pYES2.1-V5-His-TOPO:ST631ORF34	ST631 ORF34 expression vector	GAL1 promoter	This Study
pSEE13	pYES2.1-V5-His-TOPO:ST631ORF35	ST631 ORF35 expression vector	GAL1 promoter	This Study
pSEE14	pYES2.1-V5-His-TOPO:ST631ORF76	ST631 ORF76 expression vector	GAL1 promoter	This Study
pSEE15	pYES2.1-V5-His-TOPO:ST631ORF81	ST631 ORF81 expression vector	GAL1 promoter	This Study
pSEE16	pYES2.1-V5-His-TOPO:ST631VopA	ST631 VopA expression vector	GAL1 promoter	This Study
pSEE17	pYES2.1-V5-His-TOPO:ST631VopC	ST631 VopC expression vector	GAL1 promoter	This Study
pSEE18	pYES2.1-V5-His-TOPO:ST631VopL	ST631 VopL expression vector	GAL1 promoter	This Study
pSEE19	pYES2.1-V5-His-TOPO:ST631VopZ	ST631 VopZ expression vector	GAL1 promoter	This Study
pSEE20	pYES2.1-V5-His-TOPO:ST3VopZ	ST3 VopZ expression vector	GAL1 promoter	This Study

Gen = gentamycin; Erm = erythromycin; Chl = chloramphenicol; Kan = kanamycin; Tet = tetracycline

CHAPTER III: ANALYSIS OF *VIBRIO* PATHOGENICITY ISLAND VARIATION THAT IS LINKED TO CLINICAL AND ENVIRONMENTAL PREVALENCE

Sarah Eggert[1], Randi Foxall[1], Kara Rzasa[1], and Cheryl Whistler[1,2*]

[1]Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, NH

[2]Northeast Center for Vibrio Disease and Ecology, University of New Hampshire, Durham, NH

[*]Corresponding author email: cheryl.whistler@unh.edu

Abstract

Vibrio parahaemolyticus is an emergent human pathogen that is the primary cause of seafood-borne bacterial disease in the United States (US) and worldwide. Cases of *V. parahaemolyticus* infection have increased recently, especially from the North Atlantic Ocean. Most *V. parahaemolyticus* clinical cases harbor one of three evolutionary related, yet distinct Vibrio Pathogenicity Islands (VPaIs): VPaIα (*tdh*$^+$), VPaIβ (*trh*$^+$), or VPaIγ (*tdh*$^+$/*trh*$^+$). Strains harboring VPaIα cause the most infections globally, but VPaIγ-harboring strains cause most US infections. Of the different VPaIγ lineages identified thus far, *tdh*3-VPaIγ is most frequently associated with disease whereas *tdh*5-VPaIγ are more often associated with strains from the environment. In this study, we provide evidence that effectors from *tdh*3-VPaIγ and *tdh*5-VPaIγ are evolutionarily divergent, and variation impacts effector toxicity, which may enhance environmental fitness. This work lays the foundation for understanding how the unique

environmental and population context of each VPaIγ lineage shaped the evolution of differing degrees of toxicity that promote host fitness.

Introduction

Vibrio parahaemolyticus is an important human pathogen that is the primary cause of seafood-borne bacterial infections in the United States and worldwide [10,18,261]. *V. parahaemolyticus* populations are diverse, and the majority of strains are harmless to humans [10,18]. Human infection with *V. parahaemolyticus* causes inflammatory gastroenteritis through the production of secreted toxins encoded by one of three evolutionarily distinct mobile Vibrio Pathogenicity Islands (VPaIs) identified as VPaIα, VPaIß, and VPaIγ [65] containing a Type Three Secretion System (T3SS) (see Chapter 1 Figure 1.1 and Xu et al. 2017 [107]) [75,98,105–107]. These VPaI each confer production of an array of toxins; however, only a few of these toxins are known to contribute to enterotoxicity. Though the VPaIs also encode two hemolysin genes, the thermostable direct hemolysin (Tdh) and thermostable related hemolysin (Trh), both of which are helpful in identifying presence and identity of different VPaIs, neither are key virulence factor for gastric disease [6,8,71–78].

Populations of *V. parahaemolyticus* harboring any one of the three VPaI are typically in distinct, but non-uniform, geographic locations worldwide. Strains harboring VPaIα (*tdh*$^+$), first identified as VPaI-7 [106] in strain RIMD2210633, and especially pandemic *V. parahaemolyticus* known as Sequence Type (ST) 3 or serovar O3:K6 [105], are prevalent in Asia and cause the majority of infections globally [105]. Besides causing two outbreaks and sporadic infections, ST3 has not established long-term populations along the Atlantic coastal areas of North America [26,45,173,176,178,185–188]. The most clinically prevalent STs causing infections in the United States, harbor VPaIγ (*tdh*$^+$/*trh*$^+$) or less frequently VPaIß (*trh*$^+$) [19,39,40,65,190]. These related VPaI are

functionally analogous to but divergent from VPaIα, with limited shared content [65]. VPaIγ [11,65,98,107,109] contains genetic elements from both VPaIα and VPaIß [107], but displays a core content orthologous and syntenous with VPaIß [97,107] and shares the same array of predicted T3SS2 effectors, only some of which are conserved in VPaIα (see Chapter 2 Figure 2.1 and Xu et al. 2017 [107]). Thus far, experimental evidence links only three of the eight toxins encoded by VPaIα, VopV, VopZ, and most recently VopC, to gastrointestinal disease. Both VPaIß and VPaIγ encode VopC. Since the genes that are analogous to the VopV and VopZ toxins are so divergent from the characterized toxins in VPaIα, toxins other than VopV and VopZ, conserved between VPaIß and VPaIγ, may be critical for disease.

Historically, *V. parahaemolyticus* rarely caused infections from sources in the Northeast United States; thus, there was a long-held belief that the North Atlantic's environmental conditions did not sustain pathogenic populations [33,44]. However, recently, infections associated with North Atlantic harvested shellfish have increased. This trend coincided with warmer than usual ocean temperatures in the region and the incursion and establishment of non-native lineages from the Pacific Northwest and the Gulf of Mexico [8,22,48,49]. Non-natives harboring VPaIγ include several Pacific derived pathogenic lineages (ST36, ST43, and ST636) [39,40,107], and long-time residents of the Gulf of Mexico or Atlantic (ST34, ST110, and ST308) [22,53,60,190,192]. However, not all of these strains cause human infections, and those that have evolved among the Pacific population are more clinically prevalent in the Northeast US, whereas those arising in the Atlantic and Gulf of Mexico are less clinically prevalent [11]. It is noteworthy that the originating population corresponds more strongly with clinical prevalence than even VPaI type, as strains harboring VPaIγ reside in both populations, and have invaded the Northwest Atlantic.

Of the seven identified *tdh* alleles, four, including *tdh*3, *tdh*5, *tdh*6, and *tdh*7, are harbored by VPaIγ suggesting the potential that these are distinct VPaI lineages evolving in parallel that, due to the mobility of strains and horizontal nature of VPaI movement, are now globally distributed (Figure 3.1). VPaIγ that harbor the *tdh*3 allele predominate among the North Pacific population of *V. parahaemolyticus* (blue population block in Figure 3.1). These most notably include all members of the ST36 clonal complex [11] (ST36, ST38, ST39, and ST59) and several unrelated lineages isolated in the Pacific, including ST50 and ST636. In contrast, VPaIγs that harbor a *tdh*5 allele predominate among the longstanding Gulf of Mexico and the Atlantic Ocean populations (red and purple populations in Figure 3.1) including ST34, ST110, and ST674 [11,22,107,190,198] . Importantly, *V. parahaemolyticus* harboring *tdh*3-VPaIγ, even those that acquired this VPaI relatively recently (e.g., ST631), are clinically prevalent, whereas no lineage harboring *tdh*5-VPaIγ is clinically prevalent even though they are frequently cultured from the environment [11]. Strain lineages harboring *tdh*6- and *tdh7*-VPaIγ are among several of the global populations, including ST308, ST23, ST1123, and ST676 (Figure 3.1), but, notably, very few of these caused repeated infections. Specifically ST43, a lineage that likely arose in the Gulf of Mexico (red population in Figure 3.1) based on population structure analysis, but that translocated and evolved among the *tdh*3-VPaIγ strains of the Pacific population (blue population in Figure 3.1), harbors the *tdh*7-VPaIγ whereas ST23, a Gulf of Mexico residential lineages also harbors a *tdh*7-VPaIγ [11,39,40,53].

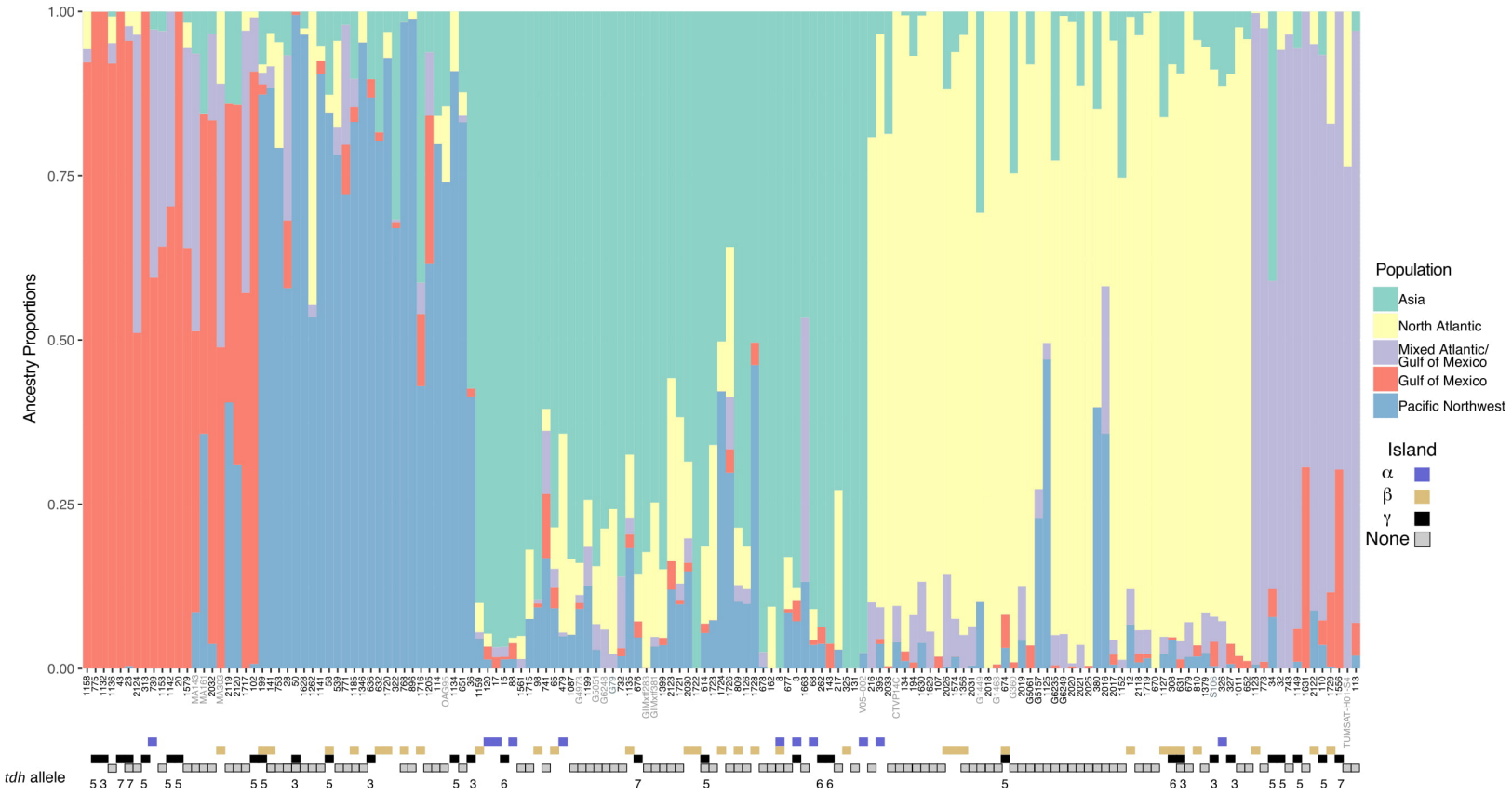


Figure 3.1: Population origin and global distribution of VPaIs. Global population structure of *V. parahaemolyticus* from whole genomes evaluated in the R statistical package LEA[262]. *tdh* alleles harbored on VPaIγ designated below VPaI blocks.

V. parahaemolyticus harboring any of the four VPaIγ types (based on *tdh* allele) have invaded, and some of these are causing infections from US Atlantic coastal waters [11]. Furthermore, genomics analyses suggest these non-indigenous lineages have donated their VPaIs to the local *V. parahaemolyticus* residents [11]. There is a pattern of variation in clinical and environmental prevalence in pathogen lineages correlated with the VPaIγ lineage they harbor. Since each VPaIγ variant originated from different geographic locations, this suggests a possibility that the population in which these VPaIγ variants evolved could have influenced virulence potential they confer [11]. Perhaps, then since the T3SS2 toxins confer virulence, perhaps T3SS2 effector variation underlies the differences in clinical prevalence.

Opportunistic pathogens, such as *V. parahaemolyticus*, evolve in the environment, and the "coincidental evolution hypothesis" suggests that maintenance of virulence factors is a response to selective environmental pressures instead of for virulence against humans [263]. One of the leading mortality factors of bacteria in marine and freshwater systems is grazing by phagotrophic protists [264–267]. In response to protozoan predation, virulence mechanisms as defense strategies have evolved in bacteria to combat this selective pressure [268]. Extracellular pathogens such as *V. parahaemolyticus*, utilize toxin secretion and biofilm formation to avoid internalization by protozoan predators and phagocytes [263]. Perhaps variation in the novel effectors encoded by VPaIγ and linked to population in which the lineages harboring these VPaI evolved might be due to differences in environmental conditions such as protozoan predation.

This study aims to compare the T3SS2 effectors from VPaIγ through analysis of population and lineage phylogenetics, assess recombination patterns and effector selection, and utilize functional models to elucidate the drivers of the variance in clinical and environmental prevalence of these islands in the North Atlantic.

Results

VPaIγ variation arose in geographically distinct populations

Analysis of the distribution of VPaIs among clinical isolates (2010-2016), updated from work in Xu 2017 [11], exemplifies that not all VPaIγ-harboring strains are equally prevalent (Table 3.1). *tdh*5-VPaIγ is rare among clinically prevalent strains, but it is in environmental isolates fairly commonly. *tdh*3-VPaIγ is in two unrelated sequence types (ST36 and ST631) that cause most infections in the Northeast United States.

Table 3.1: Demographics of Northeast United States pathogenic *V. parahaemolyticus*. Infections reported in four Northeast US States (ME, NH, MA, and CT) between 2010 and 2016, sorted by Vibrio Pathogenicity Island (VPaI) and hemolysin content. Note that not all infections were traced to local sources. The environmental prevalence of strains harboring VPaIs is based on routine surveillance.

ST	Clinical isolates	Environmental Isolates	VPaI
3	2	0	α
36	91	1	*tdh*3-γ
631	24	0	*tdh*3-γ
631	1	2	ß
631	0	1	none
43	5	0	*tdh*7-γ
636	4	0	*tdh*3-γ
1127	4	0	ß
110	3	0	*tdh*5-γ
34/324	2	2	*tdh*5-γ
674	0	4	*tdh*5-γ
674	1	0	none
308	2	0	*tdh*6-γ
12	2	0	ß
162	2	0	none
194	2	0	none
809	2	0	ß
1716	2	0	ß
1123	1	1	ß
8	1	0	ß
23	1	0	*tdh*7-γ
749	1	0	*tdh*6-γ
1295	1	0	none
134	1	0	none
741	1	0	none
98	1	0	ß
1205	1	0	none
1561	1	0	none
1717	1	0	none
1725	1	0	α

To understand this relationship between *tdh* allele and prevalence, we built phylogenetic trees of the *tdh* genes to glean the evolutionary relationships of the different alleles (Figure 3.2A). This analysis exemplifies the distinctive phylogenetic signatures of *tdh* alleles encoded by VPaIγ that differ from *tdh*1 and *tdh*2 harbored by VPaIα (Figure 3.2A). Though *tdh*3 and *tdh*5 clearly represent divergent lineages, *tdh*6 most recently evolved from a common ancestor with *tdh*3, whereas *tdh*7 and *tdh*4 share ancestry with *tdh5*. Though the phylogenetic signatures of *tdh* alleles could mean that there are four distinct lineages of VPaIγ (Figure 3.2A), when the entire VPaIγ sequences are used for constructing phylogenies and compared with the *tdh* gene phylogenies, *tdh*3-VPaIγ and *tdh*5-VPaIγ are again phylogenetically distinct. Curiously *tdh*7-VPaIγ do not form a single clade; rather these VPaIγ group with other VPaIγ by population, where the *tdh*7-VPaIγ harbored by Pacific residential ST43 groups most closely with the Pacific dominant *tdh*3-VPaIγ, and the *tdh*7-VPaIγ harbored by Gulf of Mexico resident strain ST23 groups with the *tdh*5-VPaIγ, also from Atlantic or Gulf of Mexico residential strains (Figure 3.2B). This pattern of whole island phylogenies that differs somewhat from *tdh* allele phylogenies underscores the evolutionary influence of local populations.

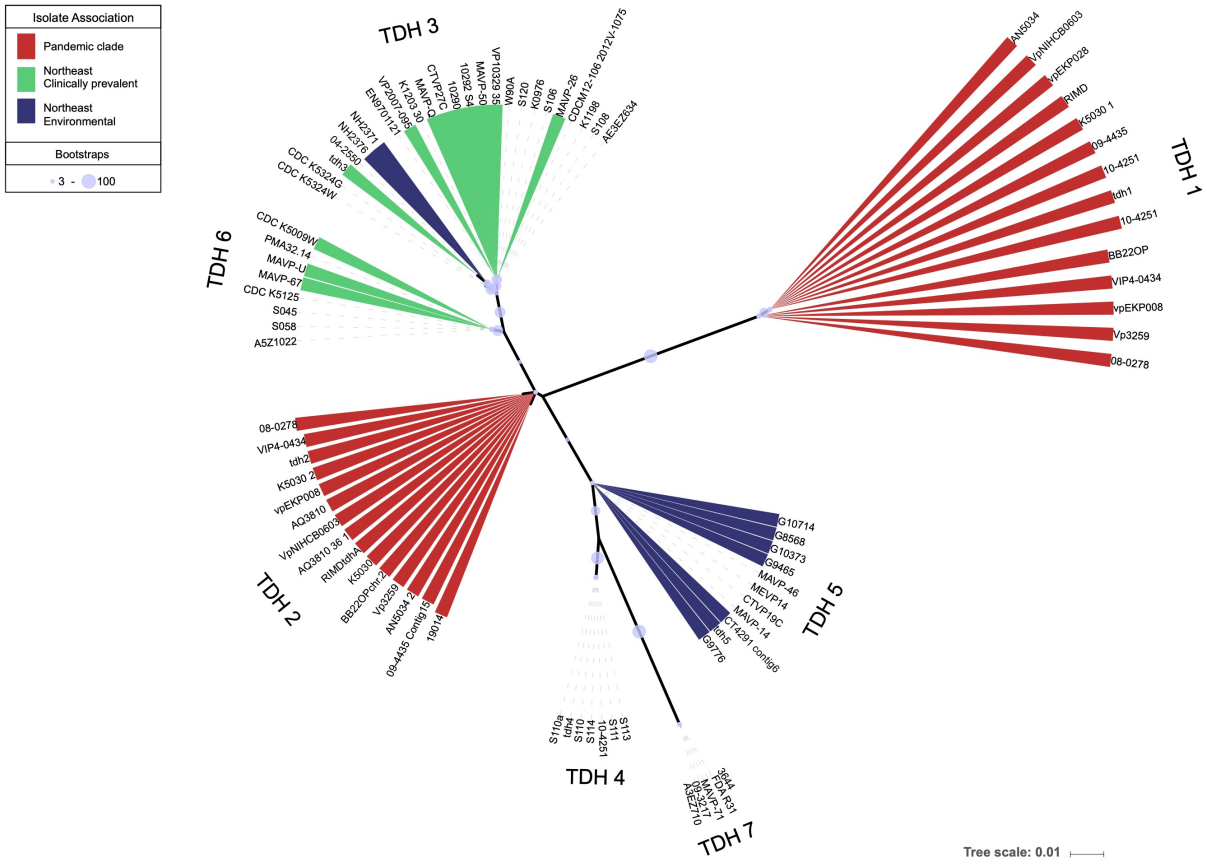


Figure 3.2A. Unique *tdh* alleles associate with each VPaIγ lineage; clinical prevalence correlates with *tdh*3-containing VPaIγ whereas environmental prevalence correlates with *tdh*5-containing VPaIγ. Maximum likelihood (ML) phylogenetic tree of *tdh* alleles from high-quality genomes on the National Center for Biotechnology Information (NCBI) database. Alleles are colored red to indicate those encoded in the pandemic island, VPaIα. *tdh* alleles prevalent in either clinical or environmental isolates from the Northeast United States coastal areas are colored green and blue.

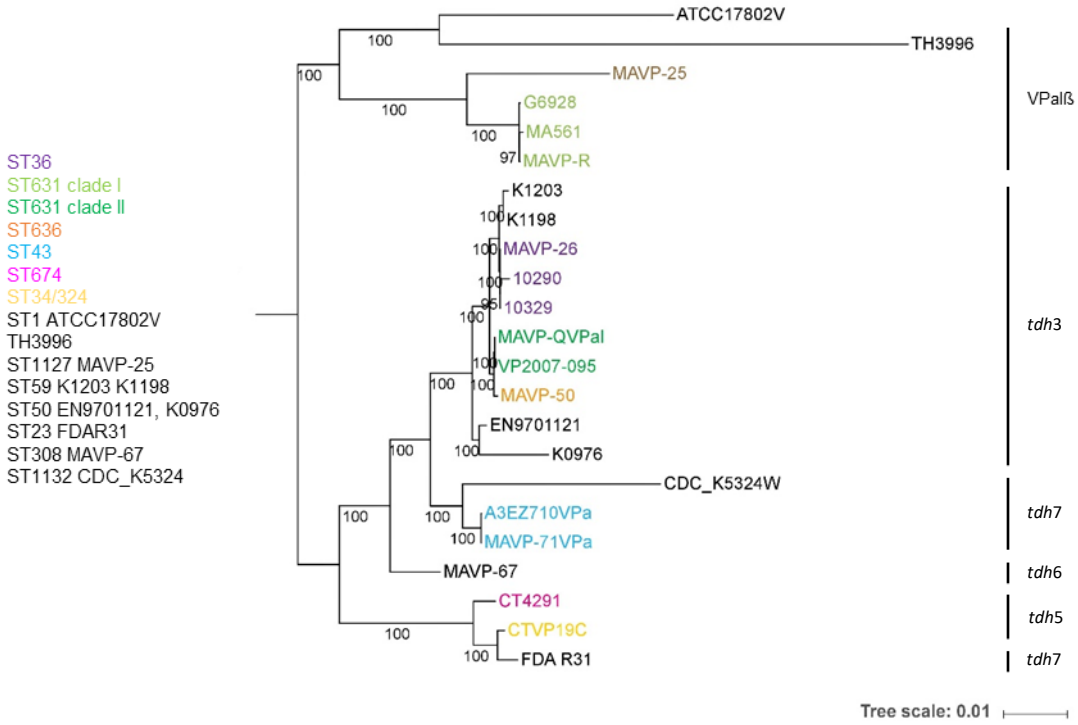


Figure 3.2B: *tdh*3-VPaIγ and *tdh*5-VPaIγ are distinct lineages. Maximum Likelihood (ML) tree constructed using the Jukes-Cantor model with 1000 bootstrap resamplings of VPaIβ and VPaIγ from select genomes constructed on sequences corresponding to VPaIγ of ST631 strain MAVP-Q.

Geographic signatures in the patterns of recombination among and between different lineages of VPaIγ suggest they evolved differently in different populations

Xu et al. examined the short-term evolution of the four types of VPaIγ, identified loci with convergent variation, and determined island relatedness based on the non-recombining regions using the bioinformatics program Gubbins [11,68]. This analysis identified that the Northwest Atlantic native *V. parahaemolyticus* ST631 harbored a *tdh*3-VPaIγ related to and derived from the same population as the "hypervirulent" Pacific-native ST36, and not the *tdh5*-VPaIγ that is more typical of Atlantic or Gulf of Mexico *V. parahaemolyticus* populations [11]. Their results also identified patterns of clustered SNPs between the four VPaIγ types [11]. Further analysis of this data indicates several regions of the VPaIγ exhibit population-specific patterns of homoplasy indicative

of either recombination or convergence (Figure 3.3). This could help explain the unresolved common ancestry *tdh7*-VPaIγ that has been blurred by the geographic separation of host strains (ST23 and ST43).

The VPaIß/VPaIγ maximum likelihood (ML)-tree built with all SNPs (Figure 3.2B), or built only with non-recombining SNPs (Figure 3.3) indicate the common ancestry of *tdh*3-VPaIγ, regardless of the strain lineage or population (Atlantic or Pacific). However, phylogenies of *tdh*7-VPaIγ constructed using variation only in non-recombining regions produced unexpected topology differences for strains that evolved in the Pacific compared to those that evolved in the Atlantic. Specifically, when variation excluded recombining regions (Figure 3.3) *tdh*7-VPaIγs branch by geographic population in which they evolved, perhaps reflective of localized adaptation tied to variation in non-recombining regions. Furthermore, the *tdh*7-VPaIγ from ST43 that evolved in the Pacific displays recombination patterns similar to that of the Pacific-native *tdh*3-VPaIγ potentially resulting from between island recombination. The *tdh*7-VPaIγ that evolved in the Gulf of Mexico displays recombination patterns similar to that of the Gulf of Mexico population *tdh*5-VPaIγ. Interestingly, one of the four novel T3SS2 toxins (see Ch. 2), Open Reading Frame (ORF) 76 which is nearby the VopC toxin that is linked to gastric disease caused by ST3, is located in a region of elevated SNP density in the *tdh*5-VPaIγ, indicative of recombination within the *tdh*5-VPaIγ clade. The first half of a second T3SS toxin, ORF35, is in a recombination region within all four VPaIγ clades.

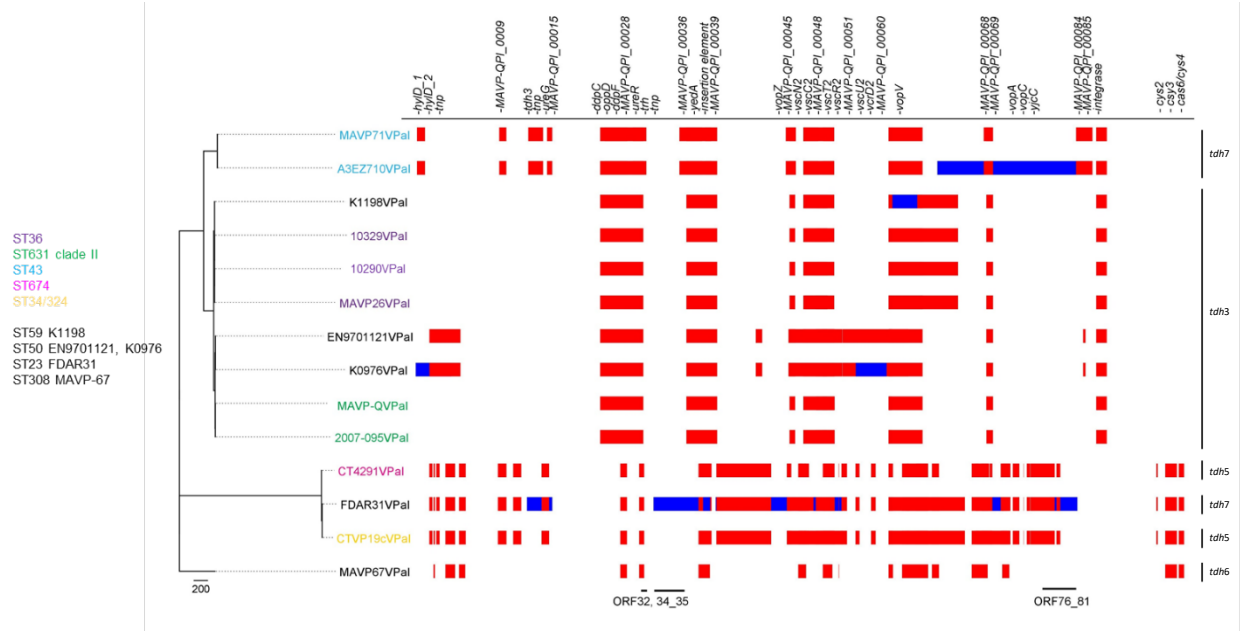


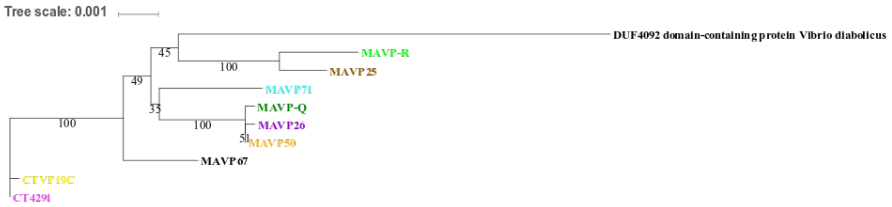
Figure 3.3: Geographic signatures of recombining regions of VPaIγ and predicted toxin genes constructed with Gubbins. ML phylogeny was built on SNPs identified in non-recombining regions (non-colored regions) and excluded regions of recombination exhibiting a higher SNP density (colored blocks) [68]. Red blocks indicate recombination within a clade of related isolates, whereas blue blocks indicate recombination with isolates that were absent from the analysis. Strains are colored by Sequence Type (ST). "PaI" designation after strain name indicates that just the VPaI sequence was used in the analysis. Isolates are colored by sequence type using a uniform coloring scheme as in Figure 3.2B and Figure 3.4.

Next, we investigated in more detail whether the phylogeny of putative effectors exhibited the same topology as the entire island to evaluate whether the tree topology group the genes by their historical lineages or local environments. We compared effector sequences from clinical isolates harboring, *tdh*3-VPaIγ, *tdh*5-VPaIγ, *tdh*7-VPaIγ, and VPaIß isolated in the northeast United States between 2010 and 2016. ML trees were constructed for each putative effector from isolates representing STs and VPaIs that either commonly cause infections or prevail environmentally in the Northeast United States (Table 3.1, Figure 3.2B). Effectors from the clinically prevalent/environmentally infrequent STs (ST36, ST631 clade II, and ST636) harboring *tdh*3-VPaIγ diverge from the orthologous effectors present in environmentally prevalent/clinically

infrequent STs (ST674 and ST34) harboring *tdh*5-VPaIγ. The divergence between *tdh*3-VPaIγ and *tdh*5-VPaIγ effector orthologs is most notable in the ORF76 phylogeny. Effectors from *tdh*6- and *tdh*7-VPaIγs did not show as strong of patterns in their evolutionary relationships like the other VPaIγ effectors.

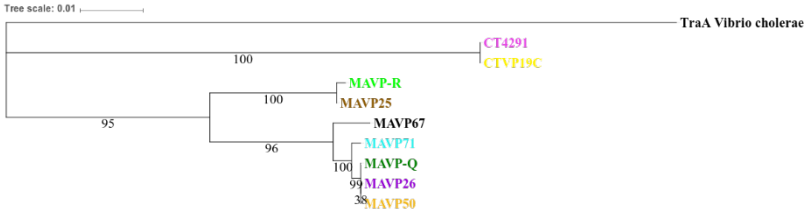
A. ORF32

B. ORF34



C. ORF35

D. ORF76



E. ORF81

volution of some T3SS2 effectors differs from the evolution of their whole VPaI.

Figure 3.4: The e A-E) Maximum-likelihood gene trees of putative effectors ORF32, ORF34, ORF35, ORF76, and ORF81. *tdh*3-VPaIγ: MAVP-Q=ST631 clade II, MAVP-26=ST36, MAVP-50=ST636, *tdh5*-VPaIγ strains include CT4291=ST674 and CTVP19C=ST34/324, *tdh*6/7-VPaIγ strains include MAVP67=ST308 and MAVP71=ST43, and VPaIβ strains include MAVP25=ST1127, MAVP-R=ST631 clade I. 100 bootstraps were run and bootstrap value shown for each branch. The homologs used as outgroups for all five effectors were chosen based on high

sequence identity and 99% query coverage when using the effectors from strain MAVP-Q as a query in BLASTn. Outgroup percent sequence identity corresponding to each ORF in MAVP-Q for ORF 32/34/35/76/81 were 68.28%, 98.49%, 89.6%, 86.03%, and 68.28% respectively.

*tdh*3-VPaIγ and *tdh*5-VPaIγ are very similar in structure, with a few striking differences (Figure 3.5). *tdh*5-VPaIγ (from strain CT4291) contains ten more ORFs than *tdh*3-VPaIγ (from strain MAVP-Q), and several of these additional genes encode RTX toxin components that are conserved in VPaIβ (data not shown) suggesting they were potentially lost from *tdh*3-VPaIγ. Homologous effectors are slightly divergent, as indicated by a decreased BLASTn percent identity. In the gene map, regions that flank ORF76 are divergent, and these regions correspond to areas identified by the prior analysis as having undergone recombination in *tdh*5-VPaIγ lineages.

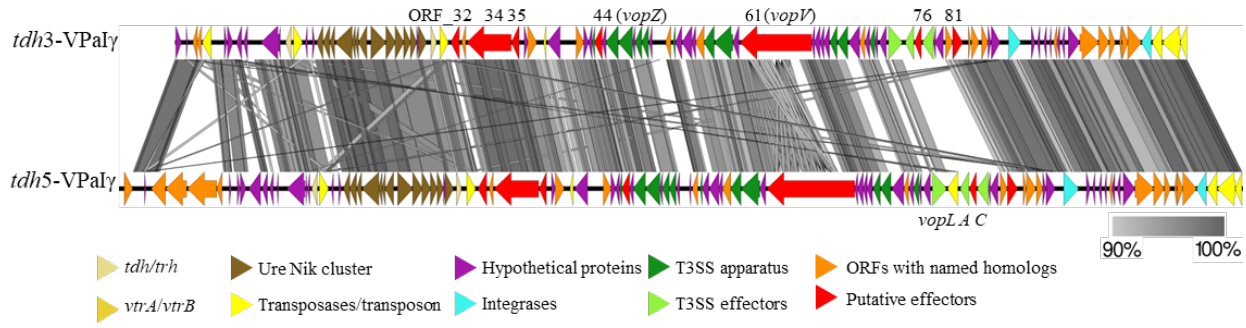


Figure 3.5. Comparison of content and conservation in *tdh*3-VPaIγ and *tdh*5-VPaIγ. Candidate effectors (including divergent *vopV* and *vopZ* homologs) are colored red. Arrows designate open reading frames. Grey bars indicate the percent similarity between homologs. *tdh*3-VPaIγ representative strain MAVP-Q (MF066646). *tdh*5-VPaIγ representative strain CT4291 (SRX7416425).

VPaIγ effector orthologs contain amino acid variation

Next, we compared the effectors' amino acid composition to examine whether the nucleotide divergence between VPaIγ effectors led to nonsynonymous mutations that could impact effector function. Alignments of effectors used to construct the ML trees in Figure 3.4, were

translated and color-coded by amino acid grouping to aid in the visualization of differences (Figure 3.6) [269]. After collapsing the conserved sites, the effectors contained unequal percentages of variable amino acid sites. The percentage of variable amino acid sites in ORF32, ORF35, ORF76, and ORF81 was 15.9%, 4.4%, 16.6%, and 10.8% respectively. Due to the deletions and/or duplications and the repetitive nature of the ORF34 sequence that makes alignment difficult, the percentage of variable amino acid sites was not calculated for ORF34.

A direct comparison of the effectors' amino acid composition from the *tdh*3-VPaIγ and *tdh*5-VPaIγ revealed three of the novel effectors, ORFs 32/76/81, differ in amino acid composition in their N-terminal region. This region of T3SS effectors contain structural patterns that include a targeting signal in the first ~15-30 amino acid residues, an optional chaperone binding domain in the following ~50-150 amino acids, and the remaining protein generally consists of effector binding domains. ORF76 has a dramatic change in the N-terminal region, with a cysteine at position 15 in *tdh*5-VPaIγ and a bulky tyrosine in *tdh*3-VPaIγ. This results not only in different size residues but also potentially alters the formation of a disulfide bridge. There is no recognizable universal secretion signal of T3SS effectors, and thus advanced machine learning approaches are necessary to predict these effectors [202,207,208,270–272]. Despite the amino acid alterations in protein's secretion signal region, we cannot determine if these changes alter secretion. The amino acid composition of the chaperone binding and effector domain regions of all five effectors differed between *tdh*3-VPaIγ and *tdh*5-VPaIγ. Notably, ORF34 orthologs from *tdh*3-VPaIγ and *tdh*5-VPaIγ contained two amino acid differences in the effector domain region, by lineage, that could alter each protein structure. The two orthologs have amino acid residues, either isoleucine or a tyrosine, that vary in size at position 1348. At position 1412 in the protein, the two orthologs have either a negatively charged glutamic acid residue or a positively charged lysine residue. These dramatic

changes suggest that protein structure would change, thus potentially altering function. However, since there is no data on the function of these effectors, we cannot predict how these changes actually alter the functionality of this protein.

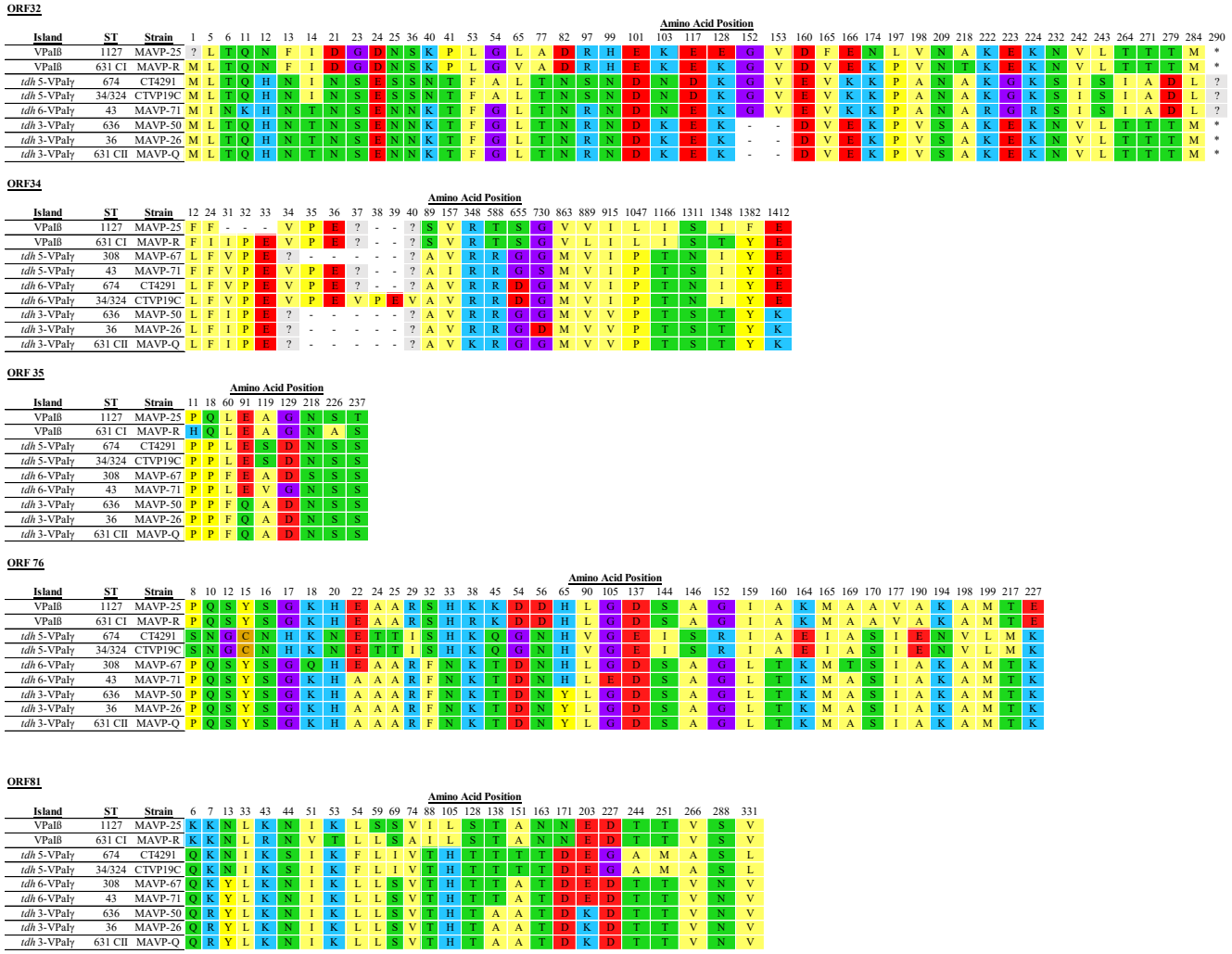


Figure 3.6: Variable amino acid composition associated with VPaIγ lineage. *tdh*5-VPaIγ strains include CT4291=ST674 and CTVP19C=ST34/324, VPaIβ strains include MAVP25=ST1127, MAVP-R=ST631 clade I, *tdh*6/7-VPaIγ strains include MAVP67=ST308 and MAVP71=ST43, and *tdh*3-VPaIγ MAVP-Q=ST631 clade II, MAVP-26=ST36, MAVP-50=ST636. Amino Acid coloring scheme: blue = positively charged, red = negatively charged, green = polar uncharged, light yellow = hydrophobic, yellow = proline/tyrosine, orange = cysteine, purple = glycine.

When comparing just the effector amino acid composition of *tdh*3-VPaIγ and *tdh*5-VPaIγ effector orthologs, ORF32 and ORF76 have the lowest percent amino acid similarity at 94.8% and 89.9%, respectively. In contrast, ORF34 and ORF35 orthologs are more highly conserved, with amino acid identity at 99.3% and 99.2% (Table 3.2).

Table 3.2: VPaIγ gene variation. BLASTx percent sequence identity between *tdh*5-VPaIγ strain CT4291 and *tdh*3-VPaIγ strain MAVP-Q.

Gene	Percent Sequence Identity
ORF32	94.8
ORF34	99.2
ORF35	99.3
ORF76	89.9
ORF81	97.4
VopA	91.1
VopC	87.9
VopL	94.3
VopV analog	96.2 (146 partial hits)
VopZ analog	95.5

***tdh*5-VPaIγ-harboring *V. parahaemolyticus* require T3SS2 effectors for survival against amoeba predation.**

Since *tdh*3-VPaIγ and *tdh*5-VPaIγ confer reciprocal patterns of clinical and environmental prevalence, we next examined whether T3SS2 secretion of effectors promote environmental fitness. We hypothesize that differences in the effectors could have resulted from environmental selection of predation, and in so doing deter protist grazing. We analyzed the growth of *V. parahaemolyticus* strains in the presence of the amoeba predator *Acanthamoeba castellanii* [273] to test whether the T3SS toxins secreted by *tdh*3-VPaIγ and *tdh*5-VPaIγ deters protist grazing (Figure 3.7). Loss of T3SS1 and T3SS2 effectors secretion did not alter survival under grazing pressure for *V. parahaemolyticus* harboring *tdh*3-VPaIγ. Conversely, loss of the T3SS effector secretion in

V. parahaemolyticus strains harboring *tdh*5-VPaIγ did impair survival under grazing pressure from *A. castellanii*.

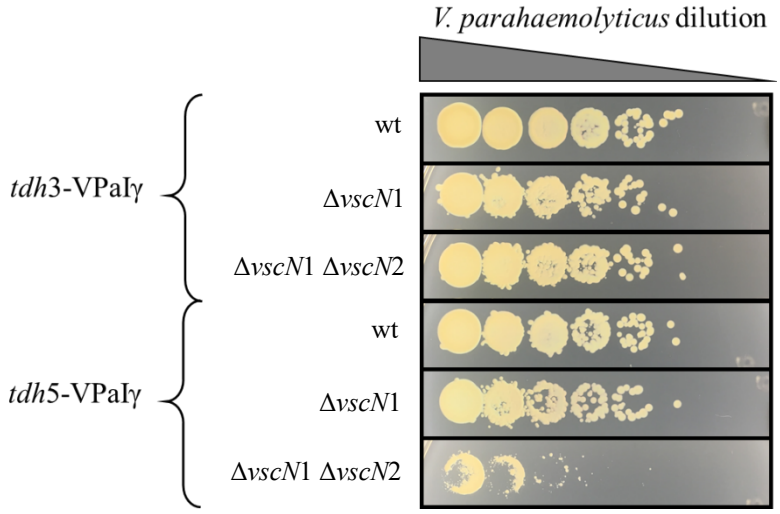


Figure 3.7: *tdh*5-VPaIγ T3SS effectors protect against protozoan predation. 10-fold dilutions of *V. parahaemolyticus* were co-cultured with amoeba predator *A. castellanii* and grown on LB agar.

T3SS effectors from *tdh*3-VPaIγ promote killing of *Galleria mellonella*

Next, we used the well-established virulence model, *Galleria mellonella*, to assess the relative contribution of T3SS2 effectors from *tdh*3-VPaIγ and *tdh*5-VPaIγ to pathogenesis. Larvae of the greater wax moth *G. mellonella* have been used as an infection model since 1987 [274] because they provide a useful insight into the pathogenesis of a wide range of microbial infections, including fungal and bacterial pathogens [275]. Similar to mammals, insects such as *G. mellonella* possess a complex innate immune system that can provide relevant information towards the mammalian infection process [275]. In this study, *G. mellonella* were challenged with *V. parahaemolyticus* strains with and without functional T3SSs to assess the pathogenicity contributed by each system (Figure 3.8).

Post-infection, *G. mellonella* were assigned a Health Index score as reported in Loh et al. [276,277] based on survival, activity, cocoon formation, and melanization. *G. mellonella* had similar health index scores post-infection with wild type *tdh*3-VPaIγ and *tdh*5-VPaIγ harboring strains. Elimination of secretion by the ancestral and conserved T3SS1 (Δ*vscN*1) ameliorated larvae damage by an environmentally prevalent *tdh*5-VPaIγ strain (ST674 strain, CT4291) to similar levels of PBS buffer injection control. Although elimination of T3SS1 (Δ*vscN*1) qualitatively ameliorated larval damage by a clinically prevalent *tdh3*-VPaIγ strain (ST631 strain MAVP-Q), it did not do so significantly, likely due to substantial intra-animal variability in response to injection with the wild type. Additionally, eliminating the T3SS2 (Δ*vscN*2) system's function from the T3SS1 mutants (a T3SS1/T3SS2 (Δ*vscN*1/Δ*vscN*2) double mutant) did not significantly increase larval health by either strain compared to the single T3SS1 (Δ*vscN*1) mutant. But, importantly, level or larval damage by the *tdh*3-VPaIγ-harboring T3SS1/T3SS2 (Δ*vscN*1/Δ*vscN*2) double mutant was significantly different from levels of damage achieved by wild type bacteria.

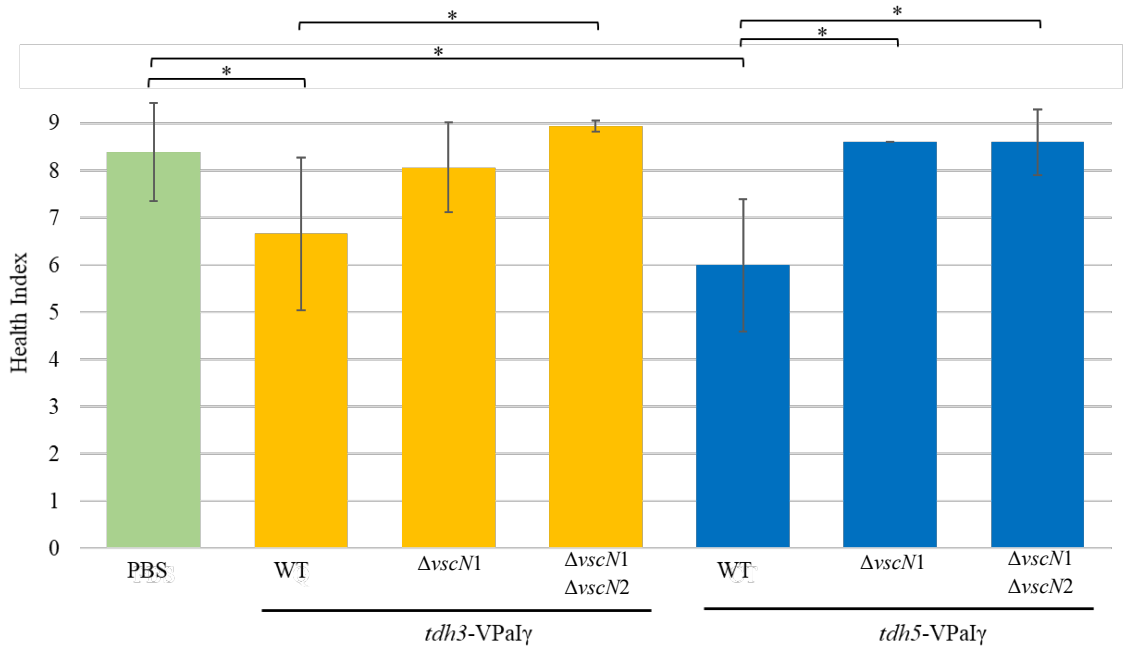


Figure 3.8: *V. parahaemolyticus* kills *Galleria mellonella* using a T3SS. Health index score of *G. mellonella* following infection with *V. parahaemolyticus* strains harboring *tdh*3-VPaIγ and *tdh*5-VPaIγ, and corresponding TTSS1 and T3SS2 mutants after infection for 24 hours. A higher health index rating indicates better health of *G. mellonella* post-infection. * = $p < 0.05$ each pair Student's t for pair-wise comparisons.

VPaIγ effectors confer differential toxicity towards eukaryotic cells

To evaluate each effector's impact on conserved eukaryotic processes, we used a heterologous expression system for T3SS effector toxicity in *Saccharomyces cerevisiae*. The proteins were expressed under the control of a galactose-inducible promoter to determine if their toxicity impaired yeast growth. When yeasts expressed ORF34 from *tdh*3-VPaIγ or *tdh*5-VPaIγ under inducing conditions, they grew nearly as well as they did when grown under the same conditions with control empty vector, suggesting neither ortholog was individually toxic (Figure 3.9A). In contrast, both *tdh*3-VPaIγ and *tdh*5-VPaIγ ORF35 orthologs impaired yeast growth. Additionally, *tdh*3-VPaIγ and *tdh*5-VPaIγ effector orthologs for ORF32, ORF76, and ORF81

impaired growth to differing degrees, based on visual assessment. Yeast harboring ORF32 from *tdh*3-VPaIγ grew the same under both inducing and non-inducing conditions. ORF32 from *tdh*5-VPaIγ expression reduced yeast growth substantially. While both ORF76 and ORF81 from *tdh*3-VPaIγ reduced yeast growth more than the orthologous ORFs from *tdh*5-VPaIγ, only ORF81-induced growth reduction in yeast was significant (Figure 3.9B). In congruence with previous studies [124], known effectors VopA was toxic to yeast, and VopC was not.

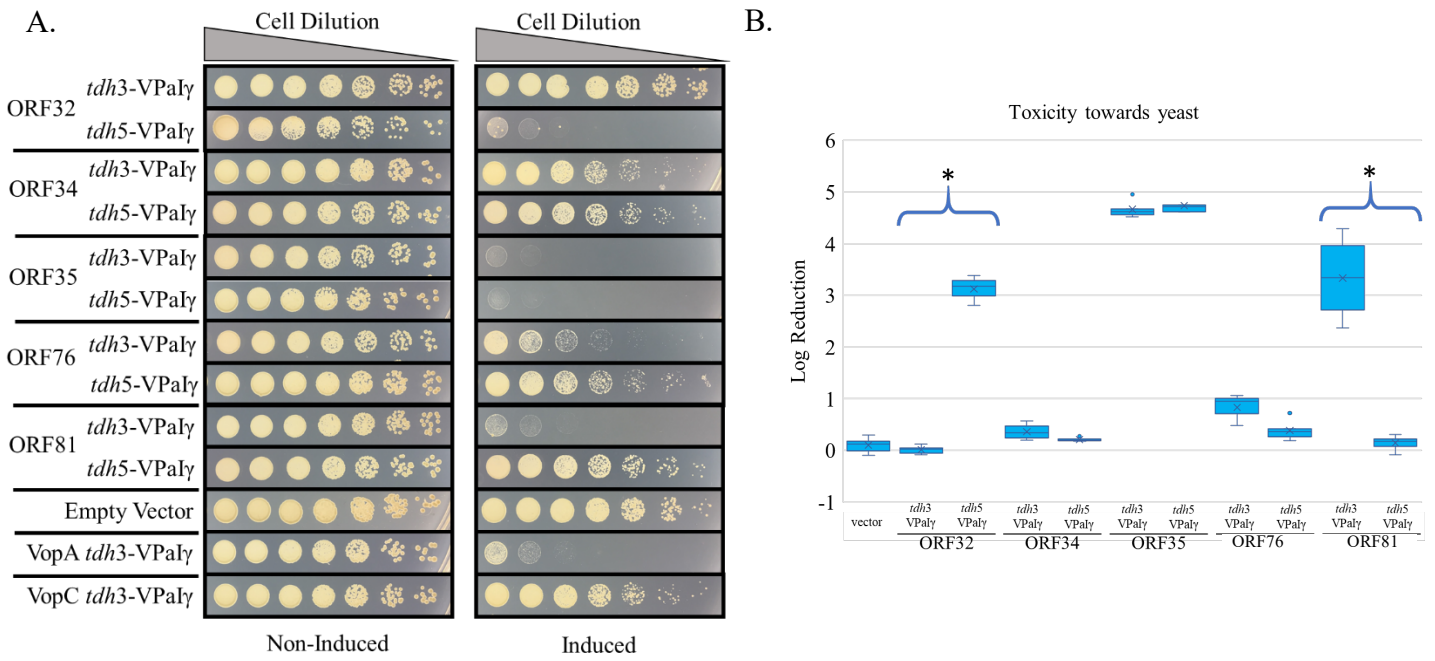


Figure 3.9: Orthologous effectors from *tdh*3-VPaIγ and *tdh*5-VPaIγ lineages exhibit differences in toxicity towards yeast. A) Four-fold dilutions of yeast transformed with pYES2.1:toxin (from *tdh*3-VPaIγ of ST631 clade II and *tdh*5-VPaIγ ST674) were grown under non-inducing (glucose) and inducing (galactose) conditions. B) Toxicity toward yeast was measured by growth on agar plates. CFU/mL was estimated from colony growth under non-inducing and inducing conditions and reported as a reduction in growth upon induction. * = $p<0.001$ Tukey Kramer HSD in pairwise comparisons.

T3SS2 effectors from *tdh*3-VPaIγ and *tdh*5-VPaIγ differ in the cytotoxicity of human colon cells.

Next, to compare the degree to which *tdh*3-VPaIγ and *tdh*5-VPaIγ effector orthologs were involved in T3SS2 dependent cytotoxicity against human colon cells, we co-cultured effector deletion mutants of *V. parahaemolyticus* strains with cultured Caco-2 colorectal tumor cells. The strain's cytotoxicity was evaluated by the release of cytosolic lactate dehydrogenase (LDH) from Caco-2 cells (Figure 3.10). Because the effectors secreted by the ancestral T3SS1 of *V. parahaemolyticus* do not contribute to enterotoxicity during infection but are cytotoxic to cultured human colon cells, we eliminated the secretion of T3SS1 effectors by deleting the T3SS1 ATPase (*vscN1*) prior to generating individual T3SS2 effector knockouts. Therefore, all strains other than the wild type carry a Δ*vscN1* mutation.

V. parahaemolyticus strains harboring T3SS1 were cytotoxic towards Caco-2 cells, regardless of which VPaI they carried (*tdh*3-VPaIγ or *tdh*5-VPaIγ) with *tdh*5-VPaIγ strain causing slightly less cell damage. This similar cytotoxicity is likely due to the presence of the ancestral T3SS1. When this system is removed, the strain harboring *tdh*3-VPaIγ was more cytotoxic than the strain harboring *tdh*5-VPaIγ. Since Caco-2 cytotoxicity assays are a proxy for *V. parahaemolyticus* virulence, this result agrees with our expectation that the more clinically prevalent strains harboring *tdh*3-VPaIγ are more toxic than strains harboring *tdh*5-VPaIγ.

Next, we evaluated the contribution of T3SS2 and its effectors to cytotoxicity by T3SS2. Deleting the accessory T3SSs ATPases (Δ*vscN*2) in *tdh*3-VPaIγ and *tdh*5-VPaIγ harboring strains reduced cytotoxicity towards Caco-2 cells to similar levels. Deleting ORF32 from *tdh*3-VPaIγ did not reduce the cytotoxicity of *V. parahaemolyticus*, whereas knocking out ORF32 from *tdh*5-VPaIγ dramatically reduced its cytotoxicity. ORFs 34 and 35 are likely co-transcribed and thus

were deleted together, and designated ORF34-35. The ΔORF34-35 mutation reduced cytotoxicity of the *tdh*3-VPaIγ significantly more than the same mutation in the *tdh*5-VPaIγ strain. ΔORF76 and ΔORF81 homolog deletions also followed a similar pattern of cytotoxicity as ΔORF34-5, but only the ΔORF81 derivatives significantly differed from each other. The analogs to the VPaIα primary effectors VopV and VopZ were also deleted in VPaIγ strains and assessed for cytotoxicity. Though divergent, the *tdh*3-VPaIγ VopV and VopZ homologs are cytotoxic.

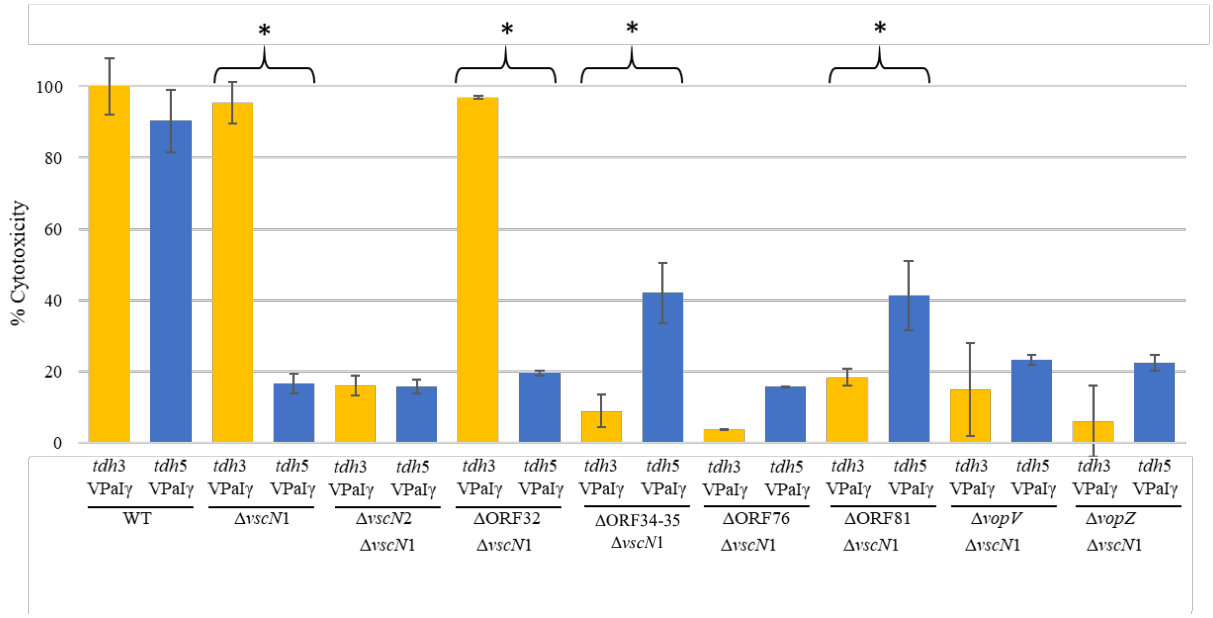


Figure 3.10. Reduced cytotoxicity towards Caco-2 cells mirrors the difference in toxicity towards yeast. The human colorectal cancer cell line, Caco-2 cells, were co-incubated with *V. parahaemolyticus* strains defective in T3SS1 and harboring indicated ORF deletions. We predict ORFs 34 and 35 are co-transcribed and thus deleted both simultaneously. Percent cytotoxicity, estimated by the quantity of lactate dehydrogenase (LDH) in the well after co-culture, was calculated by the following equation: [optical density at 490 nm (OD_{490}) of experimental release - OD_{490} of spontaneous release]/(OD_{490} of maximum release - OD_{490} of spontaneous release)*100%. Spontaneous release is the amount of LDH released from the cytoplasm of uninfected cells. The maximum release is the total amount of LDH released after the complete lysis of uninfected cells. * = $p < 0.001$ Tukey Kramer HSD for pair-wise comparisons. Data represents the mean ± standard deviation from one experiment using five individual bacterial colonies per strain, performed in triplicate. The experiment was repeated twice with similar results

Discussion

The continuous global spread of *V. parahaemolyticus* has led to population mixing and horizontal gene transfer of VPaIs [11,22,26,48,197]. The pathogenicity island's additional genetic content expands the flexible gene pool and may enhance fitness under specific environmental conditions [278]. For example, toxins secreted by the T3SS2 on VPaIα enhance *V. parahaemolyticus* persistence while in the presence of bacterivorous protists [273] and enable growth and invasion in mammalian gastrointestinal tract [139,141,147]. Here we provide evidence that strains with different VPaIγ lineages recently spread into the North Atlantic population [11,22,197] contain evolutionary divergent toxins that may provide a fitness benefit. The blocks of homoplasy in the VPaIγ lineages likely indicate recombination in each of the geographic populations. This is exemplified by *tdh*6-VPaIγs containing different recombination regions that correlate with the population they evolved with. The Pacific-evolving *tdh*7-VPaIγ has blocks of recombination similar to that of the fellow Pacific-evolving *tdh*3-VPaIγ clade, and the Gulf of Mexico-evolving *tdh*7-VPaIγ has some blocks of recombination similar to that of the fellow Gulf of Mexico-evolving *tdh*5-VPaIγ clade. This recombination pattern indicates that the surrounding population is a major contributor to the content of these VPaIs. This pattern could have arisen by an ancestor VPaIγ-harboring strain that began evolving with their local areas during global spread. Continued analysis of these populations that recently relocated to the Northwest Atlantic will demonstrate local population's role in shaping the evolution of these islands.

In the North Atlantic, *tdh*3-VPaIγ containing strains are clinically prevalent and environmentally infrequent, while the reciprocal pattern occurs with *tdh*5-VPaIγ containing strains [11]. This reciprocal clinical/environmental relationship between *tdh*3-VPaIγ and *tdh*5-VPaIγ containing strains suggests that variation in gene content between the islands could promote

infection or environmental success, respectively. Since geographic signatures in the patterns of recombination differ between the Pacific and the Gulf of Mexico/Atlantic evolving populations, we evaluated whether the toxins from the *tdh*3-VPaIγ and *tdh*5-VPaIγ lineages originating out of these regions contribute to the VPaIs' reciprocal environmental and clinical prevalence in the North Atlantic. We assessed the contribution of the *tdh*3-VPaIγ and *tdh*5-VPaIγ towards clinical and environmental success using three functional models systems: *G. mellonella* infection [279], human colon cell culture cytotoxicity [9], and *A. castellanii* predation [280].

During infection, *V. parahaemolyticus* faces defenses from the human innate immune system. To assess whether the effectors secreted by the accessory T3SS on VPaIγ promote successful invasion, we challenged *G. mellonella* with *V. parahaemolyticus*. The *G. mellonella* infection model has been used to assess bacterial virulence against innate immune defenses for many gram-negative pathogens [277] such as *Pseudomonas aeruginosa* [281], *Escherichia coli* [282], *Legionella pneumophila* [283], and *Klebsiella pneumonia* [284]. The insect innate immune system is complex and mirrors responses seen in mammals such as cellular response, humoral response, and antibacterial peptide production [285,286]. Our results suggest that the T3SS effectors from *tdh*3-VPaIγ contribute towards the successful invasion of *V. parahaemolyticus* by overcoming innate immune system defenses [275,277]. This response displays the virulence fitness benefit conferred by the presence of VPaIγ in *V. parahaemolyticus*.

V. parahaemolyticus colonizes primarily the distal small intestine during infection, where it forms microcolonies and induces substantial epithelial damage [125,135,137,138]. In this study the Caco-2 were representative of differentiated enterocytes that occur in the small intestine [245], previously shown as an effective proxy for evaluating gastric pathogen virulence [124,141,165]. We identified which T3SS effectors promoted *V. parahaemolyticus* virulence, without other

confounding variables such as mucus, microbiota, and the innate immune system [245]. Co-culture of Caco-2 cells with *tdh*3-VPaIγ and *tdh*5-VPaIγ harboring strains of *V. parahaemolyticus* indicate that the *tdh*3-VPaIγ T3SS2 effectors are greater contributors to the cytotoxicity of these human colon cells than those from *tdh*5-VPaIγ.

Like other opportunistic pathogens such as *P. aeruginosa* [287], *V. parahaemolyticus* evolves in the environment where toxin secretion is an important defense tactic against protozoan predation [263,288]. Protozoan predators are considered the "training grounds" for bacterial pathogens, allowing refinement of defense tactics, which are coincidentally beneficial to pathogenesis in humans [263,289,290]. *V. parahaemolyticus* is a member of the marine community [33,291,292] and thus commonly interacts with protozoan predators [273,293]. The T3SS effectors on VPaIα promote survival of *V. parahaemolyticus* against diverse protist predators [273], similarly to the protective effects shown by T3SS effectors from *tdh*5-VPaIγ against the predation pressure from *A. castellanii*. Intriguingly, the *tdh*3-VPaIγ-harboring strain was not affected by *A. castellanii* predation. To be a significant human pathogen, *V. parahaemolyticus* needs to survive in the environment and grow to high enough concentrations to be ingested by humans, perhaps these *V. parahaemolyticus* lineages have a mechanism outside of the VPaIγ that allows for increased environmental survival.

Our findings suggest that the difference between the clinical prevalence of *tdh*3-VPaIγ and *tdh*5-VPaIγ may in part be attributed to their T3SS effectors. To determine which effectors may underlie potential differences in virulence, we utilized a bioinformatic and experimental approach to compare the effectors in these islands and assess their virulence contribution. Three of the five effectors, ORFs 34/35/81 were not in blocks of homoplasy, most likely indicating that they were either not in recombining regions or convergence was not occurring on these loci. It appears that ORF35 could be undergoing convergence. Convergence at this loci could be supported since

ORF35 had the shortest branch lengths of all ortholog phylogenies and had the lowest percentage of variable amino acid content between all orthologs examined. The most divergent lineages of ORF35, *tdh*3-VPaIγ and *tdh*5-VPaIγ, still maintain 99.336% sequence identity at the nucleotide level, suggesting maintenance at this locus is important, likely indicting positive selection [294]. Two of the five toxins, ORF76 and the first half of ORF32, were in blocks of recombination. The presence of ORF76 in this region is indicative of recombination at this loci within the *tdh*5-VPaIγ clade. ORF76 orthologs contain the highest percentage of variable amino acid content of all five T3SS effectors, and they have the most substantial divergence in the ortholog phylogeny between *tdh*3-VPaIγ and *tdh*5-VPaIγ as indicated by the longest branch lengths. ORF32 is in a region of recombination within all four VPaIγ lineages, and this recombination shines through in its high percentage of variable amino acid content between the ORF32 orthologs. The ortholog phylogenies of all five toxins show substantial divergence between the *tdh*3-VPaIγ and *tdh*5-VPaIγ toxin sequences. The modifications in toxins may contribute to their effectiveness in conferring pathogenesis or environmental fitness.

In comparing just effector amino acid composition from *tdh*3-VPaIγ and *tdh*5-VPaIγ orthologs, there are major amino acid changes between the two islands in both the chaperone binding and effector domain regions of all orthologs, though the way in which these novel effectors function is not yet known. Even so, these changes could alter virulence if the chaperone cannot bind as effectively, or the functional domains are not working as well. One of the most interesting differences is two amino acid changes in the M60-like domain of ORF34. This domain typically possesses mucinase activity, which could be beneficial in breaking down the protective mucin layer produced by enterocytes to allow for better bacterial adhesion [224,227,295,296]. This domain is conserved in both pathogenic and commensal bacteria in Proteobacteria, Firmicutes, and

Bacteroidetes [224,295]. For example, the Enterotoxigenic *Escherichia coli* (ETEC) effector Yghl, also contains an M60-like protease domain, which degrades mucus proteins in mouse small intestine epithelium [227,296]. The two changes that could affect the protein structure are a size change at position 1348 from isoleucine to tyrosine, and a side chain charge switch at position 1412 from a negatively charged glutamic acid to a positively charged lysine. Three of the predicted effectors, ORFs 32/76/81, have a possible amino acid change in their N-terminal secretion signal region. Changes in sequence here could mistarget an effector, causing less secretion, and thus less biological effect on its target. ORF76 has a dramatic change in the N-terminal region, with a cysteine at position 15 in *tdh*5-VPaIγ and a bulky tyrosine in *tdh*3-VPaIγ. There is a size difference between these residues, and the possible formation of a disulfide bridge can dramatically alter the protein structure. Nonsynonymous substitutions can affect numerous protein features such as expression, folding, and binding affinity, which in turn may influence fitness [297–299]. Amino acid changes seen in these toxins may in part influence the clinical prevalence of strains harboring these VPaIγs.

Finally, to begin efforts to elucidate each effector's function, we utilized growth inhibition phenotyping of yeast expressing the toxins from *tdh*3-VPaIγ and *tdh*5-VPaIγ. This yeast growth inhibition model assesses if a protein negatively affects conserved eukaryotic processes such as those associated with the cytoskeleton, Rho G and MAPK signaling pathways, apoptosis, and vesicle trafficking [233,235]. Of the surveyed effectors, only ORF32 and ORF81 homologs varied significantly in their toxicity towards yeast, mirroring the results from human colon cell infection. The *tdh*5-VPaIγ ORF32 was significantly more toxic towards yeast than its homolog from *tdh*3-VPaIγ. Since *V. parahaemolyticus* evolves in the environment, we hypothesize that the ORF32 ortholog in *tdh*5-VPaIγ has best evolved for survival against eukaryotic predators in the

environment, and Caco-2 cells happen to be susceptible as well. Perhaps ORF32 is the T3SS effector from *tdh*5-VPaIγ that promoted survival under grazing pressure from *A. castellanii*. Our results could follow the coincidental evolution hypothesis, which suggests that bacterial virulence factors arise as a protective trait due to the selective pressure of bacterial-protozoan interactions, and their virulence in humans is an added benefit [263,300]. T3SS effectors on VPaIα have shown to promote survival in the presence of predatory protists [273]. In the yeast model, both ORF34 homologs had almost identical levels of minor growth reduction, and both ORF35 homologs both killed yeast equally well. When knocked out together, their combined effect truly parses out the difference between *tdh*3-VPaIγ and *tdh*5-VPaIγ homologs. The *tdh*3-VPaIγ ΔORF34-35 strain showed significantly more reduction in cytotoxicity than *tdh*5-VPaIγ ΔORF34-35, indicating that the structural differences between these two genes may affect infection success. ORF76 and ORF81 from *tdh*3-VPaIγ both might contribute to this island's clinical prevalence since they are more toxic to yeast and are more cytotoxic to human colon cells than their *tdh*5-VPaIγ homologs. Even though further studies are needed to assess each effector's contribution to gastric infection [138,139,301], we can see preliminary patterns of toxicity associated with prevalence.

All four VPaIγ lineages contain recombination patterns indicative of population-based evolution. Our study infers that the T3SS effector variation between *tdh*3-VPaIγ and *tdh*5-VPaIγ could in part contribute to the clinical prevalence of strains harboring *tdh*3-VPaIγ and environmental prevalence of strains harboring *tdh*5-VPaIγ. However further studies are needed to evaluate the impact these effectors have on *V. parahaemolyticus* fitness. Based on our preliminary results, we believe that the variation in ORFs 34/35, ORF76, and ORF81 between *tdh*3-VPaIγ and *tdh*5-VPaIγ orthologs could contribute to the *tdh*3-VPaIγ clinical prevalence. With all four VPaIγ

lineages now present and mixing in the North Atlantic, it will be interesting to see how these islands change over the next decade.

Materials and Methods

Bacterial, Yeast, and Protist Strains and Culture Conditions

The bacterial strains and plasmids used in this study are described in Table 2.2 and Table 2.3, respectively. *V. parahaemolyticus* was routinely grown at 37°C in Luria-Bertani medium supplemented with 2% NaCl (LBS) [251] or in defined minimal media [252] supplemented with 0.0058% K_2HPO_4, 0.1% NH_4Cl, 0.01mM $FeSO_4$, 0.1 mM Tris pH 7.4, 0.6% glycerol, 1x Artificial Sea Water, 0.625µg/mL chloramphenicol, and tap water for trace minerals for transformation. *E. coli* strains were grown at 37°C in Luria-Bertani medium (LB) as previously described [251]. Antibiotics were supplemented for selection of mutations and plasmids at the following concentrations for *V. parahaemolyticus*: 12 µg/ml gentamicin (Gen), 2.5 µg/ml chloramphenicol (Chl), and 5 µg/ml erythromycin (Erm) and at the following concentrations for *E. coli*: 50 µg/ml kanamycin (Kan), 100 µg/ml ampicillin (Amp), and 25 µg/ml chloramphenicol (Chl).

S. cerevisiae BY4742 (obtained from Clyde Denis, [253]) was maintained in Yeast extract (1%)-peptone (2%)-and dextrose (2%) medium (YPD)[253]. Yeast were grown on Synthetic Complete Minimal Media (SC-MM) made with Yeast Synthetic Drop-out Medium Supplements without uracil (Sigma-Aldrich), supplemented with glucose at 2% once transformed with pYES2.1/V5-His-TOPO®. *S. cerevisiae* was grown at 28°C.

A. castellanii was routinely maintained in proteose peptone, yeast extract, and glucose (PYG) medium (ATCC 712 PYG). *A. castellanii* were maintained by subculturing 4.2% culture into fresh PYG every 10-14 days.

Bioinformatic Analyses

To investigate the phylogeny of VPaI elements, the sequences of pathogenicity islands were extracted from assembled draft genomes using as a reference the island sequences of MAVP-Q (MF066646), allowing the identification of complementary regions in draft genomes using Mauve[302]. Coordinates were recorded for the pathogenicity island region from the contigs of each genome which were then manually extracted. The sequences from each VPaI were then aligned by MAFFT[303]. The aligned sequences file was then analyzed and an ML tree constructed as described for *tdh* genes. Subsequently, phylogeny and recombination among islands from select genomes was also conducted using the aligned FASTA file in Gubbins[68] with the default setting. The identities of genes within recombining regions were determined based on the coordinates of the reference sequence. To determine the phylogeny of effector orthologs, all ortholog gene sequences were extracted from both draft and closed genomes using BLASTn[254]. Orthologs identified using BLASTx[254] with 99% sequence identity and ranging from 61.28% -98.49% query coverage were used as outgroups. Alignments were performed by MAFFT[303] and maximum-likelihood (ML) tree was constructed using RAxML [304]. Phylogenetic trees were viewed and color-coded in Interactive Tree Of Life (iTOL) [305]. Alignments were viewed in MEGA[269] to visualize amino acid variability. Gene content and order of the VPaI elements in MAVP-Q and CT4291 were illustrated by Easyfig [306].

Molecular genetic methodologies and plasmid construction

PCR SOE and cloning

Primers for generating constructs used in marker exchange mutagenesis were designed as recommended [255]. Oligonucleotide primers were synthesized by © Integrated DNA Technologies

(Coralville, IA). PCR cycling conditions were optimized using AccuStart II PCR SuperMix (Quantabio, Beverly, MA). Genomic DNA from *V. parahaemolyticus* strain MAVP-Q, a ST631 clade II member containing VPaIγ, from VPaIγ, and strain MAVP-C, a pandemic ST3 strain containing VPaIα,was used as a template to generate amplicons upstream and downstream of the gene of interest, whereas amplicons of the antibiotic cassettes were generated using plasmids pBD4 and pEVS170 (Table 2.4). Genomic DNA was prepared by organic DNA extraction [256]. Oligonucleotide primers for gene knockouts were designed to contain at least 400bp regions flanking the gene of interest to enable efficient recombination (Table 2.3). Oligonucleotide primers for the antibiotic cassette to replace the gene of interest were designed to contain the native antibiotic resistance cassette promoter from the plasmid (Table 2.3). Analysis of the secondary structure of primers was performed by NetPrimer (©PREMIER Biosoft). Gene knockouts and antibiotic resistance genes were amplified from target DNA with Phusion® High-Fidelity DNA Polymerase or Phusion® High-Fidelity PCR Master Mix (©New England Biolabs, Ipswich, MA) according to ©New England Biolabs Inc. protocol. PCR SOE was performed as described [255]. PCR SOE product was cloned into pCR2.1TOPO® and transformed into chemically competent *E. coli* 10-beta cells (New England Biolabs®, Ipswich, MA) following Invitrogen pCR2.1TOPO® (Invitrogen K202020) protocol. Plasmid DNA was purified using ©QIAGEN Plasmid Kit protocol with house-made ©QIAGEN reagents. Confirmation that the constructs contained no PCR generated errors was determined by sequencing using the Sanger method by ©GENEWIZ LLC (South Plainfield, NJ, USA).

Transformation of *V. parahaemolyticus*

V. parahaemolyticus strains harboring pSEE1 were grown to turbidity in LBS and subsequently a 2% inoculum was used to grow an overnight culture at RT in minimal medium supplemented with 0.2% arabinose. Once cultures reached OD_{600} > 0.5, 0.5 mL culture was mixed with 2 μg of either plasmid DNA or ~10 μg of genomic DNA that was sheared by freeze-thaw three times at -20°C, and incubated at 28°C statically for 30 minutes. 0.5 mL LBS was added to the culture and was incubated at 37°C for 20-60 minutes. Cells were pelleted and suspended in ~50 μl LBS broth. Equal volumes of transformant were plated onto three concentrations of antibiotic flanking estimated optimal concentration (10/12/15 μg/ml gentamycin and 2.5/5/8 μg/ml erythromycin) and grown at 37°C overnight. Colonies were re-streaked to a plate with a higher concentration of antibiotic, grown at 37°C, then the colony screened by PCR to identify the presence of the mutation. Mutants harboring pSEE1 were passaged at a 2% inoculum in LBS twice per day for two days to lose plasmid.

***Acanthamoeba castellanii* predation assay**

Overnight cultures of *V. parahaemolyticus* were grown in LBS + 0.04% bile salts was subcultured 1% into fresh LBS + 0.04% bile salts and grown at 37°C until OD_{600} = ~0.5-1.0. Cultures were washed once in PBS and adjusted to an OD_{600} = 0.125. Eight 10-fold dilutions of culture were performed in 96 well plate. Six-day old *A. castellanii* culture was pelleted at 500g for 4 minutes and resuspended in PBS. *A. castellanii* were adjusted to $1x10^5$ cells/mL, and an equal volume of *A. castellanii* was gently mixed with the *V. parahaemolyticus* dilutions. After 4 minutes of co-culture, 10μL dilutions of each well were spotted onto an LB agar plate and grown at room temperature for two days.

Galleria mellonella

Overnight cultures of *V. parahaemolyticus* were subcultured 1% into fresh LBS + 0.04% bile salts to induce T3SS activity, or with no addition, and grown to mid-log phase. Pelleted cells were washed with Phosphate's Buffered Saline (PBS) twice, and the OD_{600} used to estimate cell numbers taken and cells diluted appropriately. *G. mellonella* were separated into treatment groups and two controls: one no-injection and one injected with PBS. All individuals in the treatment groups were inoculated with 10ul diluted *V. parahaemolyticus* in PBS, and their survival and health were monitored and recorded at 24 hours, 48 hours and 72 hours post-inoculation. The Colony Forming Units (CFU) were determined by the direct plating method for each inoculum dose. *G. mellonella* were scored for survival, activity, cocoon formation and melanization and a health index score was calculated for each treatment as per Tsai 2016 [277].

Toxin expression vector construction

To express putative toxins in yeast, genes were cloned from designated *V. parahaemolyticus* strains, cloned into a pYES2.1/V5-His-TOPO® (Invitrogen) vector shuttle plasmid that can be generated in *E. coli*, and then expressed in yeast. The forward oligonucleotide primers to amplify toxins contained a synthetic altered Kozak sequence (FRBS) overlapping the start codon of the gene of interest as recommended by manufacturer (Invitrogen). The forward primer contained the native sequence, except for the bolded nucleotides surrounding the underlined start codon (**G/A**)NNATG**G** for optimal expression in yeast. The reverse primer was designed for amplification near or at the stop codon of the gene of interest. Both primers were designed with a melting temperature between 63°- 67°C, and within 1°C of each other. Melting temperatures and

the secondary structure was analyzed by NetPrimer (©PREMIER Biosoft). Oligonucleotide primers were synthesized by ©Integrated DNA Technologies (Coralville, IA). PCR cycling conditions were optimized using AccuStart II PCR SuperMix (Quantabio, Beverly, MA). Toxin genes were amplified with Phusion® High-Fidelity DNA Polymerase or Phusion® High-Fidelity PCR Master Mix (©New England Biolabs, Ipswich, MA) according to ©New England Biolabs Inc. protocol. PCR products were cloned into a pYES2.1/V5-His-TOPO vector (See Table 2.4) and transformed into One-Shot TOP10F' *E. coli* ® following pYES2.1 TOPO™ TA Yeast Expression Kit (Invitrogen) protocol. Plasmid DNA was prepared using ©QIAGEN Plasmid Kit protocol with house-made ©QIAGEN reagents and absence of errors confirmed by sequencing and analysis ©GENEWIZ LLC (South Plainfield, NJ, USA).

Yeast Transformation

To generate competent *S. cerevisiae*, five colonies of *S. cerevisiae* BY4742 were grown overnight with shaking at 28°C. Cultures were diluted to an OD_{600} of 0.4 in 25 ml of YPD medium and grown an additional three hours at 28° while shaking. *S. cerevisiae* was washed twice in 1X LiAc/0.5X TE for a final volume of 1 mL. *S. cerevisiae* was incubated for 10 minutes at room temp. 50µL of *S. cerevisiae* cells were incubated with 10µL of sheared salmon sperm DNA and 1µg of plasmid prep of toxin:pYES2.1/V5-His-TOPO® for 1hr at 30°C. The sample was mixed with 250µL 40% PEG 3350 and incubated at 30°C for at least 1hr. Samples were heat-shocked for 15min at 42°C, pelleted, and resuspended in 80µL Milli-Q water. Samples were plated on SC-MM medium containing 2% glucose and incubated for two days at 28°C/30°C.

Assessment of effector toxicity by heterologous expression in yeast

To assess the effect of each putative toxin on yeast, *S. cerevisiae* harboring pYES2.1/V5-His-TOPO® containing individual toxins were streaked for isolation onto SC-MM + glucose 2% and grown for two days at 28°C. Since glucose can mildly repress the GAL1 promoter driving each toxin gene, individual colonies were streaked onto SC-MM + raffinose 2% and grown for two days at 28°C. *S. cerevisiae* grown in raffinose was inoculated into SC-MM + raffinose 2% broth and grown overnight at 28°C with shaking. Overnight cultures were washed in Milli-Q water, adjusted to an OD_{600} of 2.0, and diluted serially four-fold. 10µL of each dilution was spotted onto SC-MM + glucose 2% agar plates to assess normal yeast growth and SC-MM + galactose 1%/raffinose 1% agar plates to assess yeast growth when galactose induces production of each putative toxin and grown for two days at 28°C.

Yeast Enumeration

To calculate the effect of each toxin on the yeast, CFU/mL was estimated for *S. cerevisiae* growth on SC-MM +glucose 2% agar plates (inducing conditions) and SC-MM + galactose 1%/raffinose 1% agar plates (non-inducing conditions). After two days, log reduction in growth was calculated by the following formula: Log Reduction = $\log_{10}$(CFU/mL *S. cerevisiae* from SC-MM +glucose 2% agar plates / CFU/mL *S. cerevisiae* from SC-MM + galactose 1%/raffinose 1% agar plates).

Cytotoxicity Assay

Eukaryotic cell culture

Caco-2 cells were regularly maintained in Dulbecco's modified Eagle's medium (DMEM, ATCC®) containing 10% (v/v) Fetal Bovine Serum (FBS, Sigma®) and Corning® Antibiotic-Antimycotic Solution at 37°C in 5% CO_2.

Cytotoxicity measurements

Caco-2 cells at 80% confluency were washed in DMEM without phenol red + 10% FBS and seeded 25µL of cells at 315 cells/µl in 384-well plates and cultured for 48hr at 37°C in 5% CO_2 to confluency. Unmodified *V. parahaemolyticus* strains were streaked for isolation onto LBS agar, Δ*vscN1*:Gen *V. parahaemolyticus* mutants were streaked for isolation onto LBS + 5 µg/ml Gen, and Δ*vscN1*:Gen/Δ(gene of interest):Erm *V. parahaemolyticus* mutants were streaked for isolation onto LBS + 5 µg/ml Erm. All *V. parahaemolyticus* strains were grown at 37°C until robust colony growth. Five colonies of *V. parahaemolyticus* were grown overnight in LBS + 0.04% bile salts at 37°C in a roller drum. Cultures were washed twice in PBS and diluted to add 5µL of culture to 48hr Caco-2 cells at an MOI of 10. Three replicates per sample were added to Caco-2 cells. After 6hrs of co-culture, the release of lactate dehydrogenase (LDH) into the medium was quantified by using a CytoTox96® non-radioactive cytotoxicity kit (Promega®) according to the manufacturer's instructions. To quantify percent cytotoxicity, the LDH release was calculated with the following equation: [optical density at 490 nm (OD_{490}) of experimental release - OD_{490} of spontaneous release]/(OD_{490} of maximum release - OD_{490} of spontaneous release)*100%. Spontaneous release is the amount of LDH released from the cytoplasm of uninfected cells, whereas the maximum release is the total amount of LDH released after the complete lysis of uninfected cells.

Statistical Analysis

For *G. mellonella* killing assay health index score was calculated for each treatment as per Tsai 2016 [277], and Student's t pairwise comparisons performed for each pair. For yeast, CFU/mL was estimated from colony growth under non-inducing and inducing conditions. Reduction in growth upon induction was compared across samples with Tukey Kramer HSD in pairwise comparisons. Data from Caco-2 cytotoxicity assay represents the mean ± standard deviation from one experiment using five individual bacterial colonies per strain, performed in triplicate. Pairwise comparison across strains was performed utilizing Tukey Kramer HSD.

Table 3.3 Strains and plasmids used in this study

Strain or plasmid	Description	Source of Reference
E. coli strains		
One-Shot TOP10F'		Thermo Fisher®
10-beta cells		New England Biolabs®
V. parahaemolyticus strains		
MAVP-Q	WT	GCA_002209725.2
TBD	MAVP-Q derivative Δ*vscN1*:Gen	This study
TBD	MAVP-Q derivative Δ*vscN1*:Gen Δ*vscN2*:Erm	This study
TBD	MAVP-Q derivative Δ*vscN1*:Gen ΔORF32:Erm	This study
TBD	MAVP-Q derivative Δ*vscN1*:Gen ΔORF34_ORF35:Erm	This study
TBD	MAVP-Q derivative Δ*vscN1*:Gen ΔORF85:Erm	This study
TBD	MAVP-Q derivative Δ*vscN1*:Gen ΔORF90:Erm	This study
TBD	MAVP-Q derivative Δ*vscN1*:Gen Δ*vopV*:Erm	This study
TBD	MAVP-Q derivative Δ*vscN1*:Gen Δ*vopZ*:Erm	This study
CT4287	WT	This study
TBD	CT4287 derivative Δ*vscN1*:Gen	This study
TBD	CT4287 derivative Δ*vscN1*:Gen Δ*vscN2*:Erm	This study
TBD	CT4287 derivative Δ*vscN1*:Gen ΔORF32:Erm	This study
TBD	CT4287 derivative Δ*vscN1*:Gen ΔORF34_ORF35:Erm	This study
TBD	CT4287 derivative Δ*vscN1*:Gen ΔORF85:Erm	This study
TBD	CT4287 derivative Δ*vscN1*:Gen ΔORF90:Erm	This study
TBD	CT4287 derivative Δ*vscN1*:Gen Δ*vopV*:Erm	This study
TBD	CT4287 derivative Δ*vscN1*:Gen Δ*vopZ*:Erm	This study
S. cerevisiae strains		
BY4742	his3-Δ1 leu2-Δ0 ura3-Δ0 lys2-V0	Pearce 1999
A. castellanii		
30234™		ATCC®
G. mellonella		
N/A	N/A	Grubco
Plasmids		
pSEE1	Competency plasmid; Chl^R	This Study
pBD4	Gm^R cassette, Erm^R	Duerkop 2007

pEVS170	Erm^R cassette, Chl^R, mini-Tn5-Erm, oriVR6K oriTRP4 Kan^R	Adin 2008
pCR2.1-TOPO	cloning vector, Kan^R	Thermo Fisher Scientific
pSEE21	pCR2.1-TOPO:ST674ORF32::Erm ST674 ORF32 knockout, Chl^R, Erm^R	This Study
pSEE22	pCR2.1-TOPO:ST674ORF34.5::Erm ST674 ORF34/ORF35 knockout, Chl^R, Erm^R	This Study
pSEE23	pCR2.1-TOPO:ST674ORF76::Erm ST674 ORF76 knockout, Chl^R, Erm^R	This Study
pSEE24	pCR2.1-TOPO:ST674ORF81::Erm ST674 ORF81 knockout, Chl^R, Erm^R	This Study
pSEE25	pCR2.1-TOPO:ST674ORFVopV::Erm ST674 VopV knockout, Chl^R, Erm^R	This Study
pSEE26	pCR2.1-TOPO:ST674ORFVopZ::Erm ST674 VopZ knockout, Chl^R, Erm^R	This Study
pSEE27	pYES2.1-V5-His-TOPO::ST631ORF32	This study
pSEE28	pYES2.1-V5-His-TOPO::ST631ORF34	This study
pSEE29	pYES2.1-V5-His-TOPO::ST631ORF35	This study
pSEE30	pYES2.1-V5-His-TOPO::ST631ORF76	This study
pSEE31	pYES2.1-V5-His-TOPO::ST631ORF81	This study
pSEE32	pYES2.1-V5-His-TOPO::ST631VopA	This study
pSEE33	pYES2.1-V5-His-TOPO::ST631VopC	This study
pSEE34	pYES2.1-V5-His-TOPO::ST674ORF32	This study
pSEE35	pYES2.1-V5-His-TOPO::ST674ORF34	This study
pSEE36	pYES2.1-V5-His-TOPO::ST674ORF35	This study
pSEE37	pYES2.1-V5-His-TOPO::ST674ORF76	This study
pSEE38	pYES2.1-V5-His-TOPO::ST674ORF81	This study

Gen = gentamycin; Erm = erythromycin; Chl = chloramphenicol; Kan = kanamycin

Table 3.4: Oligonucleotides used in this study.

Primer name	Target Gene	DNA template	Sequence
Genetic Knockouts			
PrSEE70 674_vscN2_Erm AF	*vscN2*	CT4291 gDNA	TGAAATTGGCAAAGCG
PrSEE71 vscN2_Erm AR	*vscN2*	CT4291 gDNA	GTTTCCGCCATTCTTTGGTGTCTTGATATGACTCCTT
PrSEE72 vscN2_Erm BF	Erm^R cassette	pEVS170	AAGGAGTCATATCAAGACACCAAAGAATGGCGGAAAC
PrSEE73 vscN2_Erm BR	Erm^R cassette	pEVS170	AGACAGCCGAACGTACTCCTTTACAAAAGCGACTCATAGA
PrSEE74 vscN2_Erm CF	*vscN2*	CT4291 gDNA	TCTATGAGTCGCTTTTGTAAAGGAGTACGTTCGGCTGTCT
PrSEE75 vscN2_Erm CR	*vscN2*	CT4291 gDNA	TGATACCCAGTCATAATAAATG
PrSEE76 32_Erm AF	ORF32	CT4291 gDNA	AGGTATTACTCAAGAAGGACTGT
PrSEE82 674 32_Erm AR	ORF32	CT4291 gDNA	GTTTCCGCCATTCTTTGAGTGTTAGTTATGTTATGTTGTTT
PrSEE83 674 32_Erm BF	Erm^R cassette	pEVS170	AAACAACATAACATAACTAACACTCAAAGAATGGCGGAAAC
PrSEE84 674 32_Erm BR	Erm^R cassette	pEVS170	AGAACAAATACTAATCGTTGATATTTACAAAAGCGACTCATAGA
PrSEE85 674 32_Erm CF	ORF32	CT4291 gDNA	TCTATGAGTCGCTTTTGTAAATATCAACGATTAGTATTTGTTCT
PrSEE86 674 32_Erm CR	ORF32	CT4291 gDNA	TAGCGGTAGACACCAGCA
ORF34.5SoAF	ORF34-35	CT4291 gDNA	ATAACAACGGTTGATTTAATACTCTT
ORF34.5SoAR	ORF34-35	CT4291 gDNA	GTTTCCGCCATTCTTTGCTAACTCTGTTGAATGTAGG
ORF34.5SoBF	Erm^R cassette	pEVS170	CCTACATTCAACAGAGTTAGCAAAGAATGGCGGAAAC
ORF34.5SoBR	Erm^R cassette	pEVS170	GTTCTAACCATAGGACTGATTTACAAAAGCGACTCATAGA
ORF34.5SoCF	ORF34-35	CT4291 gDNA	TCTATGAGTCGCTTTTGTAAATCAGTCCTATGGTTAGAAC
ORF34.5SoCR	ORF34-35	CT4291 gDNA	ACCTTGCTCCTTAATTTGCCT

PrSEE63 674 85 SoAF	ORF76	CT4291 gDNA	AAGGGCGTTATTGGAGG
PrSEE64 674 85 SoAR	ORF76	CT4291 gDNA	CGTTTCCGCCATTCTTTAATGATTTCCATACAGTTAGAGTC
PrSEE65 674 85 SoBF	Erm^R cassette	pEVS170	GACTCTAACTGTATGGAAATCATTAAAGAATGGCGGAAACG
PrSEE66 674 85 SoBR	Erm^R cassette	pEVS170	TTTCTCGTCTGACTCAACTTTACAAAAGCGACTCATAGA
PrSEE67 674 85 SoCF	ORF76	CT4291 gDNA	TCTATGAGTCGCTTTTGTAAAGTTGAGTCAGACGAGAAA
PrSEE68 674 85 SoCR	ORF76	CT4291 gDNA	TCACCATAGCTGCGTTG
PrSEE55 674 90SoAF	ORF81	CT4291 gDNA	ATAGGCAGTTGAATGTTAAGTT
PrSEE56 674 90SoAR	ORF81	CT4291 gDNA	GTTTCCGCCATTCTTTGTATTAACATTTTTACCCA
PrSEE57 674 90SoBF	Erm^R cassette	pEVS170	TGGGTAAAAATGTTAATACAAAGAATGGCGGAAAC
G88SoBR	Erm^R cassette	pEVS170	GGATGTATTTCGGGTTGTTTACAAAAGCGACTCATAGA
G88SoCF	ORF81	CT4291 gDNA	TCTATGAGTCGCTTTTGTAAACAACCCGAAATACATCC
G88SoCR2	ORF81	CT4291 gDNA	GTTATTGACACTGATAGTGGCTT
Yeast Toxicity Assay			
674_32YES FRBS	ORF32	CT4291 gDNA	AAAATGCCTATATCTTTAACCGGATGTTT
674_32YES R	ORF32	CT4291 gDNA	TCAGTCCACTTCTGAATACAGTGTTTCTCTG
PrSEE52 34YES FRBS2 631_674	ORF34	CT4291 gDNA	TTGATGGTTTGTTTTTCAGGTTTTTTC
PrSEE53 34YES R2 631_674	ORF34	CT4291 gDNA	TACAACTAGTCTCCAGCCTAATGTACCGT
35YES F	ORF35	CT4291 gDNA	GAAATGAGTATTATGATTCCTACATTCAACAGAG
35YES R	ORF35	CT4291 gDNA	CTAAAATCAAATCCTTTCTATCAATGTTCACT
674_85 YES FRBS	ORF76	CT4291 gDNA	GCTATGGCTTTAAATAAAATTAACTCTATTAACT

674_85YES R	ORF76	CT4291 gDNA	CAGAAAATCTCTGCTAATGTAAAATTAACC
674_90YES FRBS	ORF81	CT4291 gDNA	AAAATGGTAATACAAAGCCAAAAACTGAGA
674_90YES R	ORF81	CT4291 gDNA	TTACCATATAATACCATGTTCACGAAGTATA

Erm = erythromycin

CHAPTER IV: CONCLUSION

This study identified four effectors unique to VPaIß/γ and absent from the pandemic VPaIα, and suggests that three alleles could contribute to the clinical success of *tdh3*-VPaIγ. The most useful tool generated from this work is the construction of pSEElostfoX and the corresponding rapid *Vibrio* natural transformation protocol. When in the presence of arabinose, pSEElostfoX induces natural competence in *V. parahaemolyticus* and *V. fischeri*. This allows for rapid transformation and genetic deletions in these species. This tool is to rapidly generate mutations can hopefully be utilized in other *Vibrio* labs.

This work provides a series of eukaryotic models that can be used to evaluate the role of T3SS effectors in *V. parahaemolyticus*. The *S. cerevisiae* toxicity model optimized in this study can easily be utilized for other toxin identification for other bacteria and initial assessment due to its ease of use, and rapid results. As exemplified by expression of the known effector VopZ, this model does have limitations of evaluating toxins that elicit a singular cell effect. After multiple lab members had trouble reproducing data from previous work [273], the *A. castellanii* co-culture model optimized in this study can be used to evaluate individual effectors. *G. mellonella* is a popular model to study bacterial virulence in the presence of an innate immune system such as pathogens would experience during human infection. This model was developed for our *V. parahaemolyticus* collection by Kara Rzasa as an undergraduate research project, before her graduation. We expect to achieve a higher resolution of this model with further experiments and perhaps knockout of hemolysin genes *tdh* and *trh*. The final eukaryotic model developed in this study was the use of Caco-2 cells to assess the cytotoxic effect of T3SS toxins on human colon cells. Other Whistler lab researchers have already used the protocol developed for this model. Future researchers can

follow the progression through these eukaryotic models for their *V. parahaemolyticus* work, or quickly adapt these to other bacteria as has already been done for *V. fischeri* work in the Whistler lab.

Further characterization of the effectors is warranted to understand their contribution towards infection and environmental persistence. The first step will be to assess if each of the effectors is secreted by the T3SS [166]. A secretion assay will also assess if differences in cytotoxicity between *tdh3*-VPaIγ and *tdh5*-VPaIγ are related to differences in secretion rates of the effectors. Scanning Electron Microscopy images of *V. parahaemolyticus*-induced effacement of the enterocyte brush border with Caco-2BBe cells will visualize the *V. parahaemolyticus* infection and evaluate if an effector contributes to the infection process. Further characterization of the mucin-degradation potential of ORF34 is warranted. Co-culture with a mucus-producing cell line such as HT29-MTX followed by an enzyme-linked immunosorbent assay (ELISA) could be used to characterize this protein further. Finally, the use of the premier model for gastric infection, the infant rabbit infection model, can truly assess whether the effectors play a role in gastric disease. A better representation of environmental fitness would be the use of microcosms. Microcosms with natural seawater would provide a more appropriate setting for assessment of *V. parahaemolyticus* environmental fitness than the controlled settings in the *A. castellanii* model.

The blocks of homoplasy in the VPaIs (Figure 3.3) indicates that these islands are evolving in geographic populations. Now present in the North Atlantic, it is likely that these four lineages of VPaI will continue this geographic population evolution and begin to recombine with each other. This work is important because if we can better understand the effector(s) that influence environmental or clinical prevalence, we can identify not only if *V. parahaemolyticus* harbors

tdh3-, tdh5-, tdh6-, or *tdh7*-VPaIγ, but as these islands evolve if they contain the effector allele associated with enhanced virulence.

LIST OF REFERENCES

1. Letchumanan, V., Chan, K.-G. & Lee, L.-H. Vibrio parahaemolyticus: a review on the pathogenesis, prevalence, and advance molecular identification techniques. *Front. Microbiol.* **5**, 453–460 (2014).
2. Su, Y. C. & Liu, C. Vibrio parahaemolyticus: A concern of seafood safety. *Food Microbiol.* **24**, 549–558 (2007).
3. Takemura, A. F., Chien, D. M. & Polz, M. F. Associations and dynamics of vibrionaceae in the environment, from the genus to the population level. *Front. Microbiol.* **5**, 1–26 (2014).
4. Thompson, F. L., Iida, T. & Swings, J. Biodiveristy of Vibrios. *Microbiol. Mol. Biol. Rev.* **68**, 403–431 (2004).
5. Thompson, J. R. & Polz, M. F. Dynamics of Vibrio Populations and Their Role in Environmental Nutrient Cycling. in *The Biology of Vibrios* 190–203 (American Society of Microbiology, 2006). doi:10.1128/9781555815714.ch13.
6. Wang, R. *et al.* The pathogenesis, detection, and prevention of Vibrio parahaemolyticus. *Front. Microbiol.* **6**, 1–13 (2015).
7. Shimohata, T. & Takahashi, A. Diarrhea induced by infection of Vibrio parahaemolyticus. *J. Med. Invest.* **57**, 179–182 (2010).
8. Nilsson, W. B., Paranjpye, R. N., Hamel, O. S., Hard, C. & Strom, M. S. Vibrio parahaemolyticus risk assessment in the Pacific Northwest: It's not what's in the water. *FEMS Microbiol. Ecol.* **95**, 1–8 (2019).
9. Hiyoshi, H., Kodama, T., Iida, T. & Honda, T. Contribution of Vibrio parahaemolyticus virulence factors to cytotoxicity, enterotoxicity, and lethality in mice. *Infect. Immun.* **78**, 1772–1780 (2010).
10. Scallan, E. *et al.* Foodborne illness acquired in the United States. *Emerg. Infect. Dis.* **17**, 1338–1340 (2011).
11. Xu, F., Sevingy, J., Jones, S. H., Cooper, V. S. & Whistler, C. A. The major Atlantic pathogenic lineage of Vibrio parahaemolyticus evolved by lateral acquisition of DNA derived from a Pacific native population. (2017).
12. Chowdhury, F. R., Nur, Z., Hassan, N., Seidlein, L. & Dunachie, S. Pandemics, pathogenicity and changing molecular epidemiology of cholera in the era of global warming. *Ann. Clin. Microbiol. Antimicrob.* **16**, 1–6 (2017).
13. Maugeri, T. L., Caccamo, D. & Gugliandolo, C. Potentially pathogenic vibrios in brackish waters and mussels. *J. Appl. Microbiol.* **89**, 261–266 (2000).
14. Morris, J. G. & Acheson, D. Cholera and Other Types of Vibriosis: A Story of Human Pandemics and Oysters on the Half Shell. *Clin. Infect. Dis.* **37**, 272–280 (2003).
15. CDC. Vibrio Species Causing Vibriosis. https://www.cdc.gov/vibrio/ (2019).
16. Ali, M., Nelson, A. R., Lopez, A. L. & Sack, D. A. Updated global burden of cholera in endemic countries. *PLoS Negl. Trop. Dis.* **9**, 1–13 (2015).
17. Centers for Disease Control and Prevention (CDC). Cholera - Vibrio cholerae infection. https://www.cdc.gov/cholera/index.html.

18. Daniels, N. A. *et al.* Vibrio parahaemolyticus Infections in the United States, 1973–1998 . *J. Infect. Dis.* **181**, 1661–1666 (2002).
19. Jones, J. L. *et al.* Biochemical, serological, and virulence characterization of clinical and oyster Vibrio parahaemolyticus isolates. *J. Clin. Microbiol.* **50**, 2343–2352 (2012).
20. O'Boyle, N. & Boyd, A. Manipulation of intestinal epithelial cell function by the cell contact-dependent type III secretion systems of Vibrio parahaemolyticus. *Front. Cell. Infect. Microbiol.* **3**, 1–15 (2014).
21. Ralph, A. & Currie, B. J. Vibrio vulnificus and V. parahaemolyticus necrotising fasciitis in fishermen visiting an estuarine tropical northern Australian location. *J. Infect.* **54**, e111–e114 (2007).
22. Xu, F. *et al.* Genetic characterization of clinical and environmental Vibrio parahaemolyticus from the Northeast USA reveals emerging resident and non-indigenous pathogen lineages. *Front. Microbiol.* **6**, 1–15 (2015).
23. Zhang, L. & Orth, K. Virulence determinants for Vibrio parahaemolyticus infection. *Curr. Opin. Microbiol.* **16**, (2013).
24. Honda, T. & Iida, T. The pathogenicity of Vibrio parahaemolyticus and the role of the thermostable direct haemolysin and related haemolysins. *Rev. Med. Microbiol.* **4**, 106–113 (1993).
25. Tena, D. *et al.* Fulminant necrotizing fasciitis due to Vibrio parahaemolyticus. *J. Med. Microbiol.* **59**, 235–238 (2010).
26. Nair, G. B. *et al.* Global dissemination of Vibrio parahaemolyticus serotype O3:K6 and its serovariants. *Clin. Microbiol. Rev.* **20**, 39–48 (2007).
27. Daniels, N. A. & Shafaie, A. A Review of Pathogenic Vibrio Infections. **17**, (2000).
28. Honda, T., Iida, T., Akeda, Y. & Kodama, T. Sixty Years of Vibrio parahaemolyticus Research. *Microbe Mag.* **3**, 462–466 (2008).
29. Tran, L. *et al.* Determination of the infectious nature of the agent of acute hepatopancreatic necrosis syndrome affecting penaeid shrimp. *Dis. Aquat. Organ.* **105**, 45–55 (2013).
30. Lee, C.-T. *et al.* The opportunistic marine pathogen Vibrio parahaemolyticus becomes virulent by acquiring a plasmid that expresses a deadly toxin . *Proc. Natl. Acad. Sci.* **112**, 10798–10803 (2015).
31. Khimmakthong, U. & Sukkarun, P. The spread of Vibrio parahaemolyticus in tissues of the Pacific white shrimp Litopenaeus vannamei analyzed by PCR and histopathology. *Microb. Pathog.* **113**, 107–112 (2017).
32. Cui, Y. *et al.* Epidemic clones, oceanic gene pools, and eco-LD in the free living marine pathogen vibrio parahaemolyticus. *Mol. Biol. Evol.* **32**, 1396–1410 (2015).
33. Ellis, C. N. *et al.* Influence of Seasonality on the Genetic Diversity of Vibrio parahaemolyticus in New Hampshire Shellfish Waters as Determined by Multilocus Sequence Analysis. *Appl. Environ. Microbiol.* **78**, 3778–3782 (2012).
34. Hazen, T. H., Pan, L., Gu, J. D. & Sobecky, P. A. The contribution of mobile genetic elements to the evolution and ecology of Vibrios. *FEMS Microbiol. Ecol.* **74**, 485–499 (2010).
35. Paranjpye, R., Hamel, O. S., Stojanovski, A. & Liermann, M. Genetic Diversity of Clinical and Environmental Vibrio parahaemolyticus Strains from the Pacific Northwest. *Appl. Environ. Microbiol.* **78**, 8631–8638 (2012).
36. Parveen, S. *et al.* Seasonal distribution of total and pathogenic Vibrio parahaemolyticus in

Chesapeake Bay oysters and waters. *Int. J. Food Microbiol.* **128**, 354–361 (2008).
37. Johnson, C. N. *et al.* Ecology of vibrio parahaemolyticus and vibrio vulnificus in the coastal and estuarine waters of Louisiana, Maryland, Mississippi, and Washington (United States). *Appl. Environ. Microbiol.* **78**, 7249–7257 (2012).
38. Zimmerman, A. M. *et al.* Variability of total and pathogenic vibrio parahaemolyticus densities in Northern Gulf of Mexico water and oysters. *Appl. Environ. Microbiol.* **73**, 7589–7596 (2007).
39. Turner, J. W. *et al.* Population Structure of Clinical and Environmental Vibrio parahaemolyticus from the Pacific Northwest Coast of the United States. *PLoS One* **8**, 1–10 (2013).
40. Banerjee, S. K. *et al.* Phenotypic and genotypic characterization of Canadian clinical isolates of Vibrio parahaemolyticus collected from 2000 to 2009. *J. Clin. Microbiol.* **52**, 1081–1088 (2014).
41. Johnson, C. N. *et al.* Relationships between environmental factors and pathogenic vibrios in the northern gulf of Mexico. *Appl. Environ. Microbiol.* **76**, 7076–7084 (2010).
42. Centers for Disease Control and Prevention (CDC). Preliminary FoodNet Data on the Incidence of Infection with Pathogens Transmitted Commonly Through Food --- 10 States, 2006. *MMWR Morb Mortal Wkly Rep* **56**, 336–339 (2007).
43. Centers for Disease Control and Prevention (CDC). Cholera and Other Vibrio Illness Surveillance (COVIS). https://www.cdc.gov/vibrio/surveillance.html (2016).
44. Mahoney, J. C., Gerding, M. J., Jones, S. H. & Whistler, C. A. Comparison of the pathogenic potentials of environmental and clinical Vibrio parahaemolyticus strains indicates a role for temperature regulation in virulence. *Appl. Environ. Microbiol.* **76**, 7459–7465 (2010).
45. DePaola, A., Kaysner, C. A., Bowers, J. & Cook, D. W. Environmental Investigations of Vibrio parahaemolyticus in. *Appl. Environ. Microbiol.* **66**, 4649–4654 (2000).
46. 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.* **281**, 603 (1999).
47. Martinez-Urtaza, J., Bowers, J. C., Trinanes, J. & DePaola, A. Climate anomalies and the increasing risk of Vibrio parahaemolyticus and Vibrio vulnificus illnesses. *Food Res. Int.* **43**, 1780–1790 (2010).
48. Martinez-Urtaza, J. *et al.* Spread of Pacific Northwest Vibrio parahaemolyticus Strain. *N Engl J Med* **47**, 549–562 (2013).
49. Newton, A. *et al.* 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 Morb Mortal Wkly Rep* **63**, 335–336 (2014).
50. Baker-Austin, C. *et al.* Emerging Vibrio risk at high latitudes in response to ocean warming. *Nat. Clim. Chang.* **3**, 73–77 (2013).
51. Kaysner, C. A., DePaola, A. & Jones, J. BAM: Vibrio. *Bacteriol. Anal. Man.* (2019).
52. DePaola, A. *et al.* Molecular, serological, and virulence characteristics of Vibrio parahaemolyticus isolated from environmental, food, and clinical sources in North America and Asia. *Appl. Environ. Microbiol.* **69**, 3999–4005 (2003).
53. Johnson, C. N. *et al.* Genetic relatedness among tdh + and trh + vibrio parahaemolyticus cultured from gulf of mexico oysters (crassostrea virginica) and surrounding water and sediment. *Microb. Ecol.* **57**, 437–443 (2009).

54. Gutierrez West, C. K., Klein, S. L. & Lovell, harles R. High frequency of virulence factor genes tdh, trh, and tlh in Vibrio parahaemolyticus strains isolated from a pristine estuary. *Appl. Environ. Microbiol.* **79**, 2247–2252 (2013).
55. Schmidt, H. & Hensel, M. Kiessling_Scasso_1996GeoRF (1).pdf. **17**, 14–56 (2004).
56. Juhas, M. *et al.* Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* **33**, 376–393 (2009).
57. Didelot, X. & Maiden, M. C. J. Impact of recombination on bacterial evolution. *Trends Microbiol.* **18**, 315–322 (2010).
58. Esteves, K. *et al.* Highly diverse recombining populations of vibrio cholerae and vibrio parahaemolyticus in french mediterranean coastal lagoons. *Front. Microbiol.* **6**, 1–17 (2015).
59. Gavilan, R. G., Zamudio, M. L. & Martinez-Urtaza, J. Molecular Epidemiology and Genetic Variation of Pathogenic Vibrio parahaemolyticus in Peru. *PLoS Negl. Trop. Dis.* **7**, (2013).
60. González-Escalona, N. *et al.* Determination of molecular phylogenetics of Vibrio parahaemolyticus strains by multilocus sequence typing. *J. Bacteriol.* **190**, 2831–2840 (2008).
61. Hanage, W. P., Fraser, C., Tang, J., Connor, T. R. & Corander, J. Hyper-Recombination, Diversity, and Antibiotic Resistance in Pneumococcus. *Science (80-.).* **324**, 1454–1457 (2009).
62. Sheppard, S. K., McCarthy, N. D., Falush, D. & Maiden, M. C. J. Convergence of Campylobacter species: Implications for bacterial evolution. *Science (80-.).* **320**, 237–239 (2008).
63. Goldenfeld, N. & Woese, C. Biology ' s next revolution ∗. 1–3.
64. Blokesch, M. & Schoolnik, G. K. Serogroup conversion of Vibrio cholerae in aquatic reservoirs. *PLoS Pathog.* **3**, 0733–0742 (2007).
65. Chen, Y. *et al.* Comparative genomic analysis of Vibrio parahaemolyticus: Serotype conversion and virulence. *BMC Genomics* **12**, 294 (2011).
66. Robinson, D. A. & Enright, M. C. Evolution of S. aureus by Large Chromosomal Repeats. *Society* **186**, 1060–1064 (2004).
67. Antonova, E. S. & Hammer, B. K. Genetics of Natural Competence in Vibrio cholerae and other Vibrios. *Microbiol. Spectr.* **3**, 1–18 (2015).
68. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
69. Blokesch, M. Chitin colonization, chitin degradation and chitin-induced natural competence of Vibrio cholerae are subject to catabolite repression. *Environ. Microbiol.* **545**, 1898–1912 (2012).
70. Broberg, Christopher A. , Calder, Thomas J., Orth, K. Vibrio parahaemolyticus cell biology and pathogenicity determinants. **49**, 1841–1850 (2009).
71. Kishishita, M. *et al.* Sequence Variation in the Thermostable Direct Hemolysin-. **58**, 2449–2457 (1992).
72. Park, K. S. *et al.* Cytotoxicity and Enterotoxicity of the Thermostable Direct Hemolysin-Deletion Mutants of Vibrio parahaemolyticus. *Microbiol. Immunol.* **48**, 313–318 (2004).
73. Shirai, H. *et al.* Molecular epidemiologic evidence for association of thermostable direct hemolysin (TDH) and TDH-related hemolysin of Vibrio parahaemolyticus with gastroenteritis. *Infect. Immun.* **58**, 3568–3573 (1990).

74. Panicker, G., Call, D. R., Krug, M. J. & Bej, A. K. Detection of pathogenic Vibrio spp. in shellfish by using multiplex PCR and DNA microarrays. *Appl. Environ. Microbiol.* **70**, 7436–7444 (2004).
75. Nishibuchi, M. & Kaper, J. B. Thermostable direct hemolysin gene of Vibrio parahaemolyticus: a virulence gene acquired by a marine bacterium. *Infect. Immun.* **63**, 2093–9 (1995).
76. Honda, T., Ni, Y., Miwatani, T., Adachi, T. & Kim, J. The thermostable direct hemolysin of Vibrio parahaemolyticus is a pore-forming toxin. *Can. J. Microbiol.* **38**, 1175–1180 (1992).
77. Raghunath, P. Roles of thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH) in Vibrio parahaemolyticus. *Front. Microbiol.* **5**, 2010–2013 (2014).
78. Nishibuchi, M. & Kaper, J. B. Duplication and variation of the thermostable direct haemolysin (tdh) gene in Vibrio parahaemolyticus. *Mol. Microbiol.* **4**, 87–99 (1990).
79. Alipour, M., Issazadeh, K. & Soleimani, J. Isolation and identification of Vibrio parahaemolyticus from seawater and sediment samples in the southern coast of the Caspian Sea. *Comp. Clin. Path.* **23**, 129–133 (2014).
80. Miyamoto, Y. *et al.* In vitro hemolytic characteristic of Vibrio parahaemolyticus: its close correlation with human pathogenicity. *J. Bacteriol.* **100**, 1147–1149 (1969).
81. Nishibuchi, M., Fasano, A., Russell, R. G. & Kaper, J. B. Enterotoxigenicity of Vibrio parahaemolyticus with and without genes encoding thermostable direct hemolysin. *Infect. Immun.* **60**, 3539–3545 (1992).
82. Dondapati, S. K., Wüstenhagen, D. A., Strauch, E. & Kubick, S. Cell-free production of pore forming toxins: Functional analysis of thermostable direct hemolysin from Vibrio parahaemolyticus. *Eng. Life Sci.* **18**, 140–148 (2018).
83. Shinoda, S. Sixty Years from the Discovery of Vibrio parahaemolyticus and Some Recollections. 数学.
84. Izutsu, K. *et al.* Comparative genomic analysis using microarray demonstrates a strong correlation between the presence of the 80-kilobase pathogenicity island and pathogenicity in kanagawa phenomenon-positive Vibrio parahemolyticus strains. *Infect. Immun.* **76**, 1016–1023 (2008).
85. Iida, T. *et al.* Evidence for genetic linkage between the ure and trh genes in Vibrio parahaemolyticus. *J. Med. Microbiol.* **46**, 639–645 (1997).
86. Iida, T. *et al.* Close proximity of the tdh, trh and we genes on the chromosome of Vibrio para haemolyticus. 2517–2523 (1998).
87. Park, K.-S. *et al.* Genetic characterization of DNA region containing the trh and ure genes of Vibrio parahaemolyticus. *Infect. Immun.* **68**, 5742–5748 (2000).
88. Nelapati, S., Nelapati, K. & Chinnam, B. K. Vibrio parahaemolyticus- An emerging foodborne pathogen-A Review. *Vet. World* **5**, 48–63 (2012).
89. Chun, D., Chung, J. K., Tak, R. & Seol, S. Y. Nature of the Kanagawa phenomenon of Vibrio parahaemolyticus. *Infect. Immun.* **12**, 81–87 (1975).
90. Twedt, R. M., Novelli, R. E., Spaulding, P. L. & Hall, H. E. Comparative Hemolytic Activity of Vibrio parahaemolyticus and Related Vibrios. *Infect. Immun.* **1**, 394–399 (1970).
91. Bhoopong, P. *et al.* Variability of properties of Vibrio parahaemolyticus strains isolated from individual patients. *J. Clin. Microbiol.* **45**, 1544–1550 (2007).
92. Banerjee, S. K., Rutley, R. & Bussey, J. Diversity and dynamics of the Canadian coastal

Vibrio community: An emerging trend detected in the temperate regions. *J. Bacteriol.* **200**, 1–4 (2018).
93. Ming, X., Yamamoto, K. & Honda, T. Construction and characterization of an isogenic mutant of Vibrio parahaemolyticus having a deletion in the thermostable direct hemolysin- related hemolysin gene (trh). *J. Bacteriol.* **176**, 4757–4760 (1994).
94. Nishibuchi, M., Janda, J. M. & Ezaki, T. The thermostable direct hemolysin gene (tdh) of Vibrio hollisae is dissimilar in prevalence to and phylogenetically distant from the tdh genes of other vibrios: Implications in the horizontal transfer of the tdh gene. *Microbiol. Immunol.* **40**, 59–65 (1996).
95. Wang, D. *et al.* Genome sequencing reveals unique mutations in characteristic metabolic pathways and the transfer of virulence genes between v. mimicus and v. cholerae. *PLoS One* **6**, (2011).
96. Klein, S. L., Gutierrez West, C. K., Mejia, D. M. & Lovell, C. R. Genes similar to the vibrio parahaemolyticus virulence-related Genes tdh, tlh, and vscC2 Occur in Other vibrionaceae species isolated from a pristine estuary. *Appl. Environ. Microbiol.* **80**, 595–602 (2014).
97. Ronholm, J. Genomic Features in Clinical Vibrio parahaemolyticus Isolates Lacking Recognized Virulence Factors and Environmental Isolates are Dissimilar. **82**, (2015).
98. Okada, N. *et al.* Identification and characterization of a novel type III secretion system in trh-positive vibrio parahaemolyticus strain TH3996 reveal genetic lineage and diversity of pathogenic machinery beyond the species level. *Infect. Immun.* **77**, 904–913 (2009).
99. Konieczna, I. *et al.* Bacterial Urease and its Role in Long-Lasting Human Diseases. *Curr. Protein Pept. Sci.* **13**, 789–806 (2012).
100. Steyert, S. R. & Kaper, J. B. Contribution of urease to colonization by shiga toxin-producing Escherichia coli. *Infect. Immun.* **80**, 2589–2600 (2012).
101. Debowski, A. W. *et al.* Helicobacter pylori gene silencing in vivo demonstrates urease is essential for chronic infection. 1–13 (2018).
102. Sidebotham, R. L. & Baron, J. H. Hypothesis: Helicobacter pylori, urease, mucus, and gastric ulcer. *Lancet* 193–195 (1990).
103. Cai, Y. L. & Ni, Y. X. Purification, characterization, and pathogenicity of urease produced by Vibrio parahaemolyticus. *J. Clin. Lab. Anal.* **10**, 70–73 (1996).
104. Gal-Mor, O. & Finlay, B. B. Pathogenicity islands: A molecular toolbox for bacterial virulence. *Cell. Microbiol.* **8**, 1707–1719 (2006).
105. Makino, K. *et al.* Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V cholerae. *Mech. Dis.* **361**, 743–749 (2003).
106. Hurley, C. C., Quirke, A. M., Reen, F. J. & Boyd, E. F. Four genomic islands that mark post-1995 pandemic Vibrio parahaemolyticus isolates. *BMC Genomics* **7**, 1–19 (2006).
107. Xu, F. *et al.* Parallel evolution of two clades of a major Atlantic endemic Vibrio parahaemolyticus pathogen lineage by independent acquisition of related pathogenicity islands. *Appl. Environ. Microbiol.* **83**, AEM.01168-17 (2017).
108. Boyd, E. F. *et al.* Molecular analysis of the emergence of pandemic Vibrio parahaemolyticus. *BMC Microbiol.* **8**, 1–14 (2008).
109. Ronholm, J., Petronella, N., Chew Leung, C., Pightling, A. W. & Banerjeea, S. K. Genomic features of environmental and clinical Vibrio parahaemolyticus isolates lacking recognized virulence factors are dissimilar. *Appl. Environ. Microbiol.* **82**, 1102–1113 (2016).

110. Espejo, R. T., García, K. & Plaza, N. Insight into the origin and evolution of the Vibrio parahaemolyticus pandemic strain. *Front. Microbiol.* **8**, 1–6 (2017).
111. Iida, T. & Yamamoto, K. Cloning and expression of two genes encoding highly homologous hemolysins from a Kanagawa phenomenonpositive Vibrio parahaemolyticus T4750 strain. *Gene* **93**, 9–15 (1990).
112. Galán, J. E. Common themes in the design and function of bacterial effectors. **74**, 1–19 (2009).
113. Green, E. R. & Mecsas, J. Bacterial Secretion Systems – An overview. *Am. Soc. Microbiol.* **4**, 1–32 (2016).
114. Gunasinghe, S. D., Webb, C. T., Elgass, K. D., Hay, I. D. & Lithgow, T. Super-resolution imaging of protein secretion systems and the cell surface of gram-negative bacteria. *Front. Cell. Infect. Microbiol.* **7**, (2017).
115. Thomas, S., Holland, I. B. & Schmitt, L. The Type 1 secretion pathway - The hemolysin system and beyond. *Biochimica et Biophysica Acta - Molecular Cell Research* vol. 1843 1629–1641 (2014).
116. Bankapalli, L. K., Mishra, R. C., Singh, B. & Raychaudhuri, S. Identification of Critical Amino Acids Conferring Lethality in VopK, a Type III Effector Protein of Vibrio cholerae: Lessons from Yeast Model System. *PLoS One* **10**, e0141038 (2015).
117. Leo, J. C., Grin, I. & Linke, D. Type V secretion: Mechanism(S) of autotransport through the bacterial outer membrane. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 1088–1101 (2012).
118. Russell, A. B., Peterson, S. B. & Mougous, J. D. Type VI secretion system effectors: Poisons with a purpose. *Nat. Rev. Microbiol.* **12**, 137–148 (2014).
119. Troisfontaines, P. & Cornelis, G. R. Type III Secretion: More Systems Than You Think. *Physiology* **20**, 326–339 (2005).
120. Cornelis, G. R. The type III secretion injectisome. *Nat. Rev. Microbiol.* **4**, (2006).
121. Marlovits, T. C. & Stebbins, C. E. Type III secretion systems shape up as they ship out. *Curr. Opin. Microbiol.* **13**, 47–52 (2010).
122. Park, K.-S. K. K. S. *et al.* Functional characterization of two type III secretion systems of Vibrio parahaemolyticus. *Infect. Immun.* **72**, 6659 (2004).
123. Burdette, D. L., Seemann, J. & Orth, K. Vibrio VopQ induces PI3-kinase-independent autophagy and antagonizes phagocytosis. *Mol. Microbiol.* **73**, 639–649 (2009).
124. Kodama, T. *et al.* Identification and characterization of VopT, a novel ADP-ribosyltransferase effector protein secreted via the Vibrio parahaemolyticus type III secretion system 2. *Cell. Microbiol.* **9**, 2598–2609 (2007).
125. O 'boyle, N., Boyd, A. & Zhang, L. Manipulation of intestinal epithelial cell function by the cell contact-dependent type III secretion systems of Vibrio parahaemolyticus. (2014) doi:10.3389/fcimb.2013.00114.
126. Ribet, D. & Cossart, P. How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes Infect.* **17**, 173–183 (2015).
127. Krachler, A. M. & Orth, K. Functional characterization of the interaction between bacterial adhesin Multivalent Adhesion Molecule 7 (MAM7) protein and its host cell ligands. *J. Biol. Chem.* **286**, 38939–38947 (2011).
128. O'Boyle, N., Houeix, B., Kilcoyne, M., Joshi, L. & Boyd, A. The MSHA pilus of Vibrio parahaemolyticus has lectin functionality and enables TTSS-mediated pathogenicity. *Int. J. Med. Microbiol.* **303**, 563–573 (2013).
129. Krachler, A. M., Ham, H. & Orth, K. Outer membrane adhesion factor multivalent

adhesion molecule 7 initiates host cell binding during infection by Gram-negative pathogens. *Proc. Natl. Acad. Sci.* **108**, 11614–11619 (2011).
130. Hsieh, Y.-C. *et al.* Study of capsular polysaccharide from Vibrio parahaemolyticus. *Infect. Immun.* **71**, 3329–3336 (2003).
131. Kirn, T. J., Jude, B. A. & Taylor, R. K. A colonization factor links Vibrio cholerae environmental survival and human infection. *Nature* **438**, 863–866 (2005).
132. Nagayama, K., Oguchi, T., Arita, M. & Honda, T. Purification and characterization of a cell-associated hemagglutinin of Vibrio parahaemolyticus. *Infect. Immun.* **63**, 1987–1992 (1995).
133. Bhowmick, R. *et al.* Intestinal adherence of Vibrio cholerae involves a coordinated interaction between colonization factor GbpA and mucin. *Infect. Immun.* **76**, 4968–4977 (2008).
134. Buttner, D. Protein Export According to Schedule: Architecture, Assembly, and Regulation of Type III Secretion Systems from Plant- and Animal-Pathogenic Bacteria. *Microbiol. Mol. Biol. Rev.* **76**, 262–310 (2012).
135. Qadri, F. *et al.* Adaptive and Inflammatory Immune Responses in Patients Infected with Strains of Vibrio parahaemolyticus . *J. Infect. Dis.* **187**, 1085–1096 (2003).
136. Pastorelli, L., Salvo, C. De, Mercado, J. R., Vecchi, M. & Pizarro, T. T. Central role of the gut epithelial barrier in the pathogenesis of chronic intestinal inflammation: Lessons learned from animal models and human genetics. *Front. Immunol.* **4**, 1–22 (2013).
137. Nelson, E. T., Clements, J. D. & Finkelstein, R. A. Vibrio cholerae adherence and colonization in experimental cholera: electron microscopic studies. *Infect. Immun.* **14**, 527–547 (1976).
138. Ritchie, J. M. *et al.* Inflammation and disintegration of intestinal villi in an experimental model for vibrio parahaemolyticus-induced diarrhea. *PLoS Pathog.* **8**, (2012).
139. Zhou, X. *et al.* Remodeling of the intestinal brush border underlies adhesion and virulence of an enteric pathogen. *MBio* **5**, 1–9 (2014).
140. Bugalhão, J. N., Mota, L. J. & Franco, I. S. Bacterial nucleators: actin' on actin. *Pathog. Dis.* **73**, ftv078 (2015).
141. Hiyoshi, H. *et al.* VopV, an F-actin-binding type III secretion effector, is required for vibrio parahaemolyticus-induced enterotoxicity. *Cell Host Microbe* **10**, 401–409 (2011).
142. Liverman, A. D. B. *et al.* Arp2/3-independent assembly of actin by Vibrio type III effector VopL. *Proc. Natl. Acad. Sci.* **104**, 17117–17122 (2007).
143. Zhang, L. *et al.* Type III Effector VopC Mediates Invasion for Vibrio Species. *Cell Rep.* **1**, 453–460 (2012).
144. Okada, R. *et al.* The Vibrio parahaemolyticus effector VopC mediates Cdc42 dependent invasion of cultured cells but is not required for pathogenicity in an animal model of infection. **4**, 139–148 (2014).
145. Hiyoshi, H. *et al.* Interaction between the Type III Effector VopO and GEF-H1 Activates the RhoA-ROCK Pathway. *PLOS Pathog.* **11**, e1004694 (2015).
146. Kustermans, G., Piette, J. & Legrand-Poels, S. Actin-targeting natural compounds as tools to study the role of actin cytoskeleton in signal transduction. *Biochem. Pharmacol.* **76**, 1310–1322 (2008).
147. Zhou, X. *et al.* A Vibrio parahaemolyticus T3SS Effector Mediates Pathogenesis by Independently Enabling Intestinal Colonization and Inhibiting TAK1 Activation. *Cell Rep.* **3**, 1690–1702 (2013).

148. Yu, B., Cheng, H. C., Brautigam, C. A., Tomchick, D. R. & Rosen, M. K. Mechanism of actin filament nucleation by the bacterial effector VopL. *Nat. Struct. Mol. Biol.* **18**, 1068–1074 (2011).
149. Ridley, A. J. & Hall, A. The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* **70**, 389–399 (1992).
150. de Souza Santos, M., Salomon, D. & Orth, K. T3SS effector VopL inhibits the host ROS response, promoting the intracellular survival of Vibrio parahaemolyticus. *PLoS Pathog.* **13**, 1–25 (2017).
151. Yang, H. *et al.* A novel mouse model of enteric vibrio parahaemolyticus infection reveals that the type iii secretion system 2 effector vopc plays a key role in tissue invasion and gastroenteritis. *MBio* **10**, 1–19 (2019).
152. Kim, M. *et al.* Bacterial interactions with the host epithelium. *Cell Host Microbe* **8**, 20–35 (2010).
153. Zenonos, K. & Kyprianou. RAS signaling pathways, mutations and their role in colorectal cancer. *World J. Gastrointest. Oncol.* **5**, 97 (2013).
154. Nandan, M. O. & Yang, V. W. An Update on the Biology of RAS/RAF Mutations in Colorectal Cancer. *Curr. Colorectal Cancer Rep.* **7**, 113–120 (2011).
155. GITTER, A. H. *et al.* Epithelial Barrier Defects in HT-29/B6 Colonic Cell Monolayers Induced by Tumor Necrosis Factor-α. *Ann. N. Y. Acad. Sci.* **915**, 193–203 (2006).
156. Pédron, T. & Sansonetti, P. Commensals, Bacterial Pathogens and Intestinal Inflammation: An Intriguing Ménage à Trois. *Cell Host Microbe* **3**, 344–347 (2008).
157. Boneca, I. G. *et al.* A critical role for peptidoglycan N-deacetylation in Listeria evasion from the host innate immune system. *Proc. Natl. Acad. Sci.* **104**, 997–1002 (2007).
158. Raffatellu, M. *et al.* Lipocalin-2 resistance of Salmonella enterica serotype Typhimurium confers an advantage during life in the inflamed intestine. *Cell Host Microbe* **5**, 476–486 (2010).
159. Stecher, B. & Hardt, W. D. Mechanisms controlling pathogen colonization of the gut. *Curr. Opin. Microbiol.* **14**, 82–91 (2011).
160. Reddick, L. E. & Alto, N. M. Bacteria fighting back: How pathogens target and subvert the host innate immune system. *Mol. Cell* **54**, 321–328 (2014).
161. Dev, A., Iyer, S., Razani, B. & Cheng, G. NF-κB and Innate Immunity. in *Chemistry of Heterocyclic Compounds* vol. 0 115–143 (2010).
162. Sakurai, H. Targeting of TAK1 in inflammatory disorders and cancer. *Trends Pharmacol. Sci.* **33**, 522–530 (2012).
163. Kajino, T. *et al.* Protein phosphatase 6 down-regulates TAK1 kinase activation in the IL-1 signaling pathway. *J. Biol. Chem.* **281**, 39891–39896 (2006).
164. Trosky, J. E. *et al.* Inhibition of MAPK signaling pathways by VopA from Vibrio parahaemolyticus. *J. Biol. Chem.* **279**, 51953–51957 (2004).
165. Kodama, T. *et al.* Identification of two translocon proteins of Vibrio parahaemolyticus type III secretion system 2. *Infect. Immun.* **76**, 4282–4289 (2008).
166. Calder, T. *et al.* Vibrio type III effector VPA1380 is related to the cysteine protease domain of large bacterial toxins. *PLoS One* **9**, 1–8 (2014).
167. Fujino, T. *et al.* On the Bacteriological Examination of Shirasu-Food Poisoning. *Med. J. Osaka Univ.* **4**, 299–304 (1953).
168. Infectious Disease Surveillance Center. Vibrio parahaemolyticus, Japan, 1996–1998.

Infect. Agents Surveill. Rep. **20**, 233 (1999).
169. Bhuiyan, N. A. *et al.* Prevalence of the pandemic genotype of Vibrio parahaemolyticus in Dhaka, Bangladesh, and significance of its distribution across different serotypes. *J. Clin. Microbiol.* **40**, 284–286 (2002).
170. Deepanjali, A., Kumar, H. S., Karunasagar, I. & Karunasagar, I. Seasonal variation in abundance of total and pathogenic Vibrio parahaemolyticus bacteria in oysters along the southwest coast of India. *Appl. Environ. Microbiol.* **71**, 3575–3580 (2005).
171. Li, Y. *et al.* Vibrio study in China. **20**, 2012–2015 (2014).
172. Liu, X., Chen, Y., Wang, X. & Ji, R. Foodborne disease outbreaks in China from 1992 to 2001 national foodborne disease surveillance system. *Wei Sheng Yan Jiu* **33**, 725–727 (2004).
173. Matsumoto, C. *et al.* Pandemic spread of an O3:K6 clone of Vibrio parahaemolyticus and emergence of related strains evidenced by arbitrarily primed PCR and toxRS sequence analyses. *J. Clin. Microbiol.* **38**, 578–585 (2000).
174. Wong, H.-C. *et al.* Characterization of Vibrio parahaemolyticus Isolates Obtained from Foodborne Illness Outbreaks during 1992 through 1995 in Taiwan. *J. Food Prot.* **63**, 900–906 (2000).
175. Shiyang, C., Shuqiung, L. & Lifang, Z. Occurance of Vibrio parahaemolyticus in seawater and some seafoods in the coastal area of Qingdao. **21**, 43–50 (1991).
176. Okuda, J. *et al.* Emergence of a unique O3:K6 clone of Vibrio parahaemolyticus in Calcutta, India, and isolation of strains from the same clonal group from Southeast Asian travelers arriving in Japan. *J. Clin. Microbiol.* **35**, 3150–3155 (1997).
177. Abbott, S. L. *et al.* 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 (1989).
178. Chowdhury, N. R. *et al.* Molecular evidence of clonal Vibrio parahaemolyticus pandemic strains. *Emerg. Infect. Dis.* **6**, 631–636 (2000).
179. Chowdhury, N. R., Stine, O. C., Morris, J. G. & Nair, G. B. Assessment of Evolution of Pandemic Vibrio parahaemolyticus by Multilocus Sequence Typing. *J. Clin. Microbiol.* **42**, 1280–1282 (2004).
180. Okuda, J., Ishibashi, M., Abbott, S. L., Janda, J. M. & Nishibuchi, M. Analysis of the thermostable direct hemolysin (tdh) gene and the tdh- related hemolysin (trh) genes in urease-positive strains of Vibrio parahaemolyticus isolated on the west coast of the United States. *J. Clin. Microbiol.* **35**, 1965–1971 (1997).
181. Ansaruzzaman, M. *et al.* Pandemic serovars (O3:K6 and O4:K68) of Vibrio parahaemolyticus associated with diarrhea in Mozambique: Spread of the pandemic into the African continent. *J. Clin. Microbiol.* **43**, 2559–2562 (2005).
182. Quilici, M. L., Robert-Pillot, A., Picart, J. & Fournier, J. M. Pandemic Vibrio parahaemolyticus O3:K6 spread, France. *Emerg. Infect. Dis.* **11**, 1148–1149 (2005).
183. González-Escalona, N. *et al.* Vibrio parahaemolyticus Diarrhea, Chile, 1998 and 2004. *Emerg. Infect. Dis.* **11**, 2004–2006 (2005).
184. Velazquez-Roman, J., León-Sicairos, N., Flores-Villaseñor, H., Villafaña-Rauda, S. & Canizalez-Roman, A. Association of pandemic Vibrio parahaemolyticus O3:K6 present in the coastal environment of Northwest Mexico with cases of recurrent diarrhea between 2004 and 2010. *Appl. Environ. Microbiol.* **78**, 1794–1803 (2012).
185. Vezzulli, L. *et al.* Climate influence on Vibrio and associated human diseases during the

past half-century in the coastal North Atlantic . *Proc. Natl. Acad. Sci.* **113**, E5062–E5071 (2016).
186. Yeung, M. P. S. & Boor, K. J. Epidemiology, Pathogenesis, and Prevention of Foodborne Vibrio parahaemolyticus Infections. *Foodborne Pathog. Dis.* **1**, 74–88 (2004).
187. Martinez-Urtaza, J. *et al.* Characterization of pathogenic Vibrio parahaemolyticus isolates from clinical sources in Spain and comparison with Asian and North American pandemic isolates. *J. Clin. Microbiol.* **42**, 4672–4678 (2004).
188. Cabrera-García, M. E., Vázquez-Salinas, C. & Quiñones-Ramírez, E. I. Serologic and molecular characterization of Vibrio parahaemolyticus strains isolated from seawater and fish products of the gulf of Mexico. *Appl. Environ. Microbiol.* **70**, 6401–6406 (2004).
189. Molenda, J. R. *et al.* Vibrio pamhaemolyticus gastroenteritis in Maryland: Laboratory aspects. *Appl. Microbiol.* **24**, 444–8 (1972).
190. PubMLST. www.pubmlst.org/vparahaemolyticus.
191. Ansede-Bermejo, J., Gavilan, R. G., TriÑanes, J., Espejo, R. T. & Martinez-Urtaza, J. Origins and colonization history of pandemic Vibrio parahaemolyticus in South America. *Mol. Ecol.* **19**, 3924–3937 (2010).
192. Ellingsen, A. B., Olsen, J. S., Granum, P. E., Rørvik, L. M. & González-Escalona, N. Genetic characterization of trh positive Vibrio spp. isolated from Norway. *Front. Cell. Infect. Microbiol.* **3**, 1–10 (2013).
193. García, K. *et al.* Rise and fall of pandemic Vibrio parahaemolyticus serotype O3: K6 in southern Chile. *Environ. Microbiol.* **15**, 527–534 (2013).
194. González-Escalona, N., Gavilan, R. G., Brown, E. W. & Martinez-Urtaza, J. Transoceanic spreading of pathogenic strains of Vibrio parahaemolyticus with distinctive genetic signatures in the recA gene. *PLoS One* **10**, 1–13 (2015).
195. Haendiges, J. *et al.* 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 (2016).
196. Martinez-Urtaza, J. *et al.* Emergence of asiatic vibrio diseases in south america in phase with El Niño. *Epidemiology* **19**, 829–837 (2008).
197. Xu, F. *et al.* Sequence Type 631 Vibrio parahaemolyticus, an Emerging Foodborne Pathogen in North America. *J. Clin. Microbiol.* 645–648 (2017).
198. Haendiges, J. *et al.* Characterization of Vibrio parahaemolyticus clinical strains from Maryland (2012-2013) and comparisons to a locally and globally diverse V. parahaemolyticus strains by whole-genome sequence analysis. *Front. Microbiol.* **6**, 1–11 (2015).
199. Nolan, C. M. *et al.* Vibrio parahaemolyticus gastroenteritis: An outbreak associated with raw oysters in the pacific northwest. *Diagn. Microbiol. Infect. Dis.* **2**, 119–128 (1984).
200. CDC. Outbreak of Vibrio parahaemolyticus infections associated with eating raw oysters - Pacific Northwest, 1997. *J. Am. Med. Assoc.* **280**, 126–127 (1998).
201. Centers for Disease Control and Prevention (CDC). Vibrio parahaemolyticus infections associated with consumption of raw shellfish--three states, 2006. *MMWR. Morb. Mortal. Wkly. Rep.* **55**, 854–6 (2006).
202. McDermott, J. E. *et al.* Minireview: Computational prediction of type III and IV secreted effectors in gram-negative bacteria. *Infect. Immun.* **79**, 23–32 (2011).
203. Dean, P. Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiol. Rev.* **35**, 1100–1125 (2011).

204. Lilic, M., Vujanac, M. & Stebbins, C. E. A common structural motif in the binding of virulence factors to bacterial secretion chaperones. *Mol. Cell* **21**, 653–664 (2006).
205. Notti, R. Q., Stebbins, C. E., Erec Stebbins, C. & Stebbins, C. E. The Structure and Function of Type III Secretion Systems. *Microbiol. Spectr.* **4**, 1–30 (2016).
206. Ghosh, P. Process of Protein Transport by the Type III Secretion System. *Society* **68**, 771–795 (2004).
207. Lower, M. & Schneider, G. Prediction of Type III Secretion Signals in Genomes of Gram-Negative Bacteria. **4**, 1–9 (2009).
208. Goldberg, T., Rost, B. & Bromberg, Y. Computational prediction shines light on type III secretion origins. *Sci. Rep.* **6**, 1–10 (2016).
209. Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **46**, D493–D496 (2018).
210. Marchler-Bauer, A. *et al.* CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45**, D200–D203 (2017).
211. Cortes, C. & Vapnik, V. Support-Vector Networks. *Mach. Learn.* **297**, 273–297 (1995).
212. Petty, N. K. *et al.* The Citrobacter rodentium genome sequence reveals convergent evolution with human pathogenic Escherichia coli. *J. Bacteriol.* **192**, 525–538 (2010).
213. Connolly, J. P. R. *et al.* Host-associated niche metabolism controls enteric infection through fine-tuning the regulation of type 3 secretion. *Nat. Commun.* **9**, (2018).
214. Zurawski, D. V., Mumy, K. L., Faherty, C. S., McCormick, B. A. & Maurelli, A. T. Shigella flexneri type III secretion system effectors OspB and OspF target the nucleus to downregulate the host inflammatory response via interactions with retinoblastoma protein. *Mol. Microbiol.* **71**, 350–368 (2009).
215. Fukazawa, A. *et al.* GEF-H1 mediated control of NOD1 dependent NF-κB activation by Shigella effectors. *PLoS Pathog.* **4**, 19–24 (2008).
216. Coletta, A. *et al.* Low-complexity regions within protein sequences have position-dependent roles. *BMC Syst. Biol.* **4**, (2010).
217. Toll-Riera, M., Radó-Trilla, N., Martys, F. & Albá, M. M. Role of low-complexity sequences in the formation of novel protein coding sequences. *Mol. Biol. Evol.* **29**, 883–886 (2012).
218. Yadav, B. S. *et al.* Bioinformatics-based study on prokaryotic, archaeal and eukaryotic nucleic acid-binding proteins for identification of low-complexity and intrinsically disordered regions. *Front. Life Sci.* **9**, 2–16 (2016).
219. DePristo, M. A., Zilversmit, M. M. & Hartl, D. L. On the abundance, amino acid composition, and evolutionary dynamics of low-complexity regions in proteins. *Gene* **378**, 19–30 (2006).
220. Karlin, S., Brocchieri, L., Bergman, A., Mrázek, J. & Gentles, A. J. Amino acid runs in eukaryotic proteomes and disease associations. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 333–338 (2002).
221. Moxon, E. R., Rainey, P. B., Nowak, M. A. & Lenski, R. E. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4**, 24–33 (1994).
222. Verstrepen, K. J., Jansen, A., Lewitter, F. & Fink, G. R. Intragenic tandem repeats generate functional variability. *Nat. Genet.* **37**, 986–990 (2005).
223. Alberts, B. *et al. Membrane Proteins - Molecular Biology of the Cell*. (Garland Science, 2002).
224. Nakjang, S., Ndeh, D. A., Wipat, A., Bolam, D. N. & Hirt, R. P. A novel extracellular

metallopeptidase domain shared by animal Host-Associated mutualistic and pathogenic microbes. *PLoS One* **7**, (2012).
225. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
226. Withey, J. H. & DiRita, V. J. Vibrio cholerae ToxT independently activates the divergently transcribed aldA and tagA genes. *J. Bacteriol.* **187**, 7890–7900 (2005).
227. Luo, Q. *et al.* Enterotoxigenic escherichia coli secretes a highly conserved mucin-degrading metalloprotease to effectively engage intestinal epithelial cells. *Infect. Immun.* **82**, 509–521 (2014).
228. Solovyev, V. & Salamov, A. AUTOMATIC ANNOTATION OF MICROBIAL GENOMES AND METAGENOMIC SEQUENCES. in *In Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies* 61–78 (Nova Science, 2011).
229. Frost, L. S., Paranchych, W. & Willetts, N. S. The complete sequence of the F traALE region that includes the gene for F pilin. *J. Bacteriol.* **160**, 395–401 (1984).
230. Hubbard, T. P. *et al.* Genetic analysis of Vibrio parahaemolyticus intestinal colonization. doi:10.1073/pnas.1601718113.
231. Zhou, X. *et al.* The Hydrophilic Translocator for Vibrio parahaemolyticus, T3SS2, Is Also Translocated. doi:10.1128/IAI.00402-12.
232. Page, A. J. *et al.* Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
233. Curak, J., Rohde, J. & Stagljar, I. Yeast as a tool to study bacterial effectors. *Curr. Opin. Microbiol.* **12**, 18–23 (2009).
234. Siggers, K. A. & Lesser, C. F. The Yeast Saccharomyces cerevisiae: A Versatile Model System for the Identification and Characterization of Bacterial Virulence Proteins. *Cell Host Microbe* **4**, 8–15 (2008).
235. Popa, C., Coll, N. S., Valls, M. & Sessa, G. Yeast as a Heterologous Model System to Uncover Type III Effector Function. *PLOS Pathog.* **12**, e1005360 (2016).
236. Livny, J. *et al.* Comparative RNA-Seq based dissection of the regulatory networks and environmental stimuli underlying Vibrio parahaemolyticus gene expression during infection. *Nucleic Acids Res.* **42**, 12212–12223 (2014).
237. Gotoh, K. *et al.* Bile acid-induced virulence gene expression of Vibrio parahaemolyticus reveals a novel therapeutic potential for bile acid sequestrants. *PLoS One* **5**, (2010).
238. Li, P. *et al.* Bile salt receptor complex activates a pathogenic type III secretion system. *Elife* **5**, 1–26 (2016).
239. Petronella, N. & Ronholm, J. The mechanisms that regulate Vibrio parahaemolyticus virulence gene expression differ between pathotypes. *Microb. Genomics* **4**, (2018).
240. Ma, W. & Guttman, D. S. Evolution of prokaryotic and eukaryotic virulence effectors. *Curr. Opin. Plant Biol.* **11**, 412–419 (2008).
241. Jimenez, A., Chen, D. & Alto, N. M. How Bacteria Subvert Animal Cell Structure and Function. *Annu. Rev. Cell Dev. Biol.* **32**, 373–397 (2016).
242. Pelaseyed, T. *et al.* The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunol. Rev.* **260**, 8–20 (2014).
243. McGuckin, M. A., Lindén, S. K., Sutton, P. & Florin, T. H. Mucin dynamics and enteric pathogens. *Nat. Rev. Microbiol.* **9**, 265–278 (2011).

244. Gagnon, M., Zihler Berner, A., Chervet, N., Chassard, C. & Lacroix, C. Comparison of the Caco-2, HT-29 and the mucus-secreting HT29-MTX intestinal cell models to investigate Salmonella adhesion and invasion. *J. Microbiol. Methods* **94**, 274–279 (2013).
245. Lievin-Le Moal, V. & Servin, A. L. Pathogenesis of Human Enterovirulent Bacteria: Lessons from Cultured, Fully Differentiated Human Colon Cancer Cell Lines. *Microbiol. Mol. Biol. Rev.* **77**, 380–439 (2013).
246. Valeri, M. *et al.* Pathogenic E. coli exploits SslE mucinase activity to translocate through the mucosal barrier and get access to host cells. *PLoS One* **10**, 1–14 (2015).
247. Conference, I. S. S. Notice of Illness Outbreaks, Shellfish Closures, Reopenings, & Recalls. http://www.issc.org/notices (2019).
248. Cox, A. M. & Gomez-Chiarri, M. Vibrio parahaemolyticus in Rhode Island coastal ponds and the estuarine environment of narragansett bay. *Appl. Environ. Microbiol.* **78**, 2996–2999 (2012).
249. Lovell, C. R. Ecological fitness and virulence features of Vibrio parahaemolyticus in estuarine environments. *Appl. Microbiol. Biotechnol.* **101**, 1781–1794 (2017).
250. Charro, N. & Mota, L. J. Approaches targeting the type III secretion system to treat or prevent bacterial infections. *Expert Opin. Drug Discov.* **10**, 373–387 (2015).
251. LB (Luria-Bertani) liquid medium. *Cold Spring Harb. Protoc.* **2006**, pdb.rec8141 (2006).
252. Visick, K. L., Hodge-Hanson, K. M., Tischler, A. H., Bennett, A. K. & Mastrodomenico, V. Tools for rapid genetic engineering of Vibrio fischeri. *Appl. Environ. Microbiol.* **84**, 26–28 (2018).
253. Pearce, D. A. & Sherman, F. Toxicity of Copper, Cobalt, and Nickel Salts Is Dependent on Histidine Metabolism in the Yeast Saccharomyces cerevisiae. **181**, 4774–4779 (1999).
254. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 1–9 (2009).
255. Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* **77**, 61–68 (1989).
256. Ausubel, M. *et al.* Current protocols in molecular biology. *Mol. Reprod. Dev.* **1**, 146–146 (1989).
257. Stabb, E. V. & Ruby, E. G. RP4-based plasmids for conjugation between Escherichia coli and members of the vibrionaceae. *Methods Enzymol.* **358**, 413–426 (2002).
258. Val, M. E., Skovgaard, O., Ducos-Galand, M., Bland, M. J. & Mazel, D. Genome engineering in Vibrio cholerae: A feasible approach to address biological issues. *PLoS Genet.* **8**, (2012).
259. Duerkop, B. A., Ulrich, R. L. & Greenberg, E. P. Octanoyl-homoserine lactone is the cognate signal for Burkholderia mallei BmaR1-BmaI1 quorum sensing. *J. Bacteriol.* **189**, 5034–5040 (2007).
260. Adin, D. M., Visick, K. L. & Stabb, E. V. Identification of a cellobiose utilization gene cluster with cryptic β-galactosidase activity in Vibrio fischeri. *Appl. Environ. Microbiol.* **74**, 4059–4069 (2008).
261. Newton, A. E. *et al.* Increase in Vibrio parahaemolyticus infections associated with consumption of Atlantic Coast shellfish — 2013. *Morb. Mortal. Wkly. Rep.* **63**, 335–336 (2014).
262. Frichot, E. & François, O. LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929 (2015).

263. Sun, S., Noorian, P. & McDougald, D. Dual role of mechanisms involved in resistance to predation by protozoa and virulence to humans. *Front. Microbiol.* **9**, 1–12 (2018).
264. Azam, F. *et al.* The Ecological Role of Water-Column Microbes in the Sea. *Mar. Ecol. - Prog. Ser.* **10**, 257–263 (1983).
265. Hahn, M. W. & Holfe, M. G. Grazing of protozoa and its effect on populations of aquatic bacteria. *FEMS Microbiol. Ecol.* **35**, 113–121 (2000).
266. Matz, C. & Jürgens, K. Effects of hydrophobic and electrostatic cell surface properties of bacteria on feeding rates of heterotrophic nanoflagellates. *Appl. Environ. Microbiol.* **67**, 814–820 (2001).
267. Erken, M., Weitere, M., Kjelleberg, S. & McDougald, D. In situ grazing resistance of Vibrio cholerae in the marine environment. *FEMS Microbiol. Ecol.* (2011) doi:10.1111/j.1574-6941.2011.01067.x.
268. Matz, C. & Kjelleberg, S. Off the hook - How bacteria survive protozoan grazing. *Trends Microbiol.* **13**, 302–307 (2005).
269. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
270. Samudrala, R., Heffron, F. & McDermott, J. E. Accurate prediction of secreted substrates and identification of a conserved putative secretion signal for type iii secretion systems. *PLoS Pathog.* **5**, (2009).
271. Arnold, R. *et al.* Sequence-Based Prediction of Type III Secreted Proteins. *PLoS Pathog.* **5**, (2009).
272. Zheng, L.-L. *et al.* A Comparison of Computational Methods for Identifying Virulence Factors. *PLoS One* **7**, e42517 (2012).
273. Matz, C., Nouri, B., McCarter, L. & Martinez-Urtaza, J. Acquired type III secretion system determines environmental fitness of epidemic vibrio parahaemolyticus in the interaction with bacterivorous protists. *PLoS One* **6**, (2011).
274. Morton, D. B., Dunphy, G. B. & Chadwick, J. S. REACTIONS OF HEMOCYTES OF IMMUNE AND NON-IMMUNE GALLERIA MELLONELLA LARVAE TO PROTEUS MIRABILIS. *Dev. Comp. Immunol.* **43**, 1986 (1987).
275. Ramarao, N., Nielsen-Leroux, C. & Lereclus, D. The insect Galleria mellonella as a powerful infection model to investigate bacterial pathogenesis. *J. Vis. Exp.* 1–7 (2012) doi:10.3791/4392.
276. Loh, J. M., Adenwalla, N., Wiles, S. & Proft, T. Galleria mellonella larvae as an infection model for group A streptococcus. *Virulence* **4**, 419–428 (2013).
277. Tsai, C. J. Y., Loh, J. M. S. & Proft, T. Galleria mellonella infection models for the study of bacterial diseases and for antimicrobial drug testing. *Virulence* **7**, 214–229 (2016).
278. Hacker, J. & Carniel, E. Ecological fitness , genomic islands and bacterial pathogenicity. **2**, 376–381 (2001).
279. Wagley, S. *et al. Galleria mellonella* as an infection model to investigate virulence of *Vibrio parahaemolyticus*. *Virulence* **9**, 1–11 (2017).
280. Albers, U., Reus, K., Shuman, H. A. & Hilbi, H. The amoebae plate test implicates a paralogue of lpxB in the interaction of Legionella pneumophila with Acanthamoeba castellanii. *Microbiology* **151**, 167–182 (2005).
281. Hill, L., Veli, N. & Coote, P. J. Evaluation of Galleria mellonella larvae for measuring the efficacy and pharmacokinetics of antibiotic therapies against Pseudomonas aeruginosa

infection. *Int. J. Antimicrob. Agents* **43**, 254–261 (2014).
282. Alghoribi, M. F., Gibreel, T. M., Dodgson, A. R., Beatson, S. A. & Upton, M. Galleria mellonella infection model demonstrates high lethality of ST69 and ST127 uropathogenic E. coli. *PLoS One* **9**, (2014).
283. Harding, C. R. *et al.* Legionella pneumophila pathogenesis in the Galleria mellonella infection model. *Infect. Immun.* **80**, 2780–2790 (2012).
284. Insua, J. L. *et al.* Modeling Klebsiella pneumoniae Pathogenesis by Infection of the Wax Moth Galleria mellonella. *Infect. Immun.* **81**, 3552–3565 (2013).
285. Vodovar, N., Acosta, C., Lemaitre, B. & Boccard, F. Drosophila: A polyvalent model to decipher host-pathogen interactions. *Trends Microbiol.* **12**, 235–242 (2004).
286. Dalhammar, G. & Steiner, H. Characterization of inhibitor A, a protease from Bacillus thuringiensis which degrades attacins and cecropins, two classes of antibacterial proteins in insects. **252**, 247–252 (1984).
287. Coburn, B., Sekirov, I. & Finlay, B. B. Type III secretion systems and disease. *Clin. Microbiol. Rev.* **20**, 535–549 (2007).
288. Erken, M., Lutz, C. & McDougald, D. The Rise of Pathogens: Predation as a Factor Driving the Evolution of Human Pathogens in the Environment. *Microb. Ecol.* **65**, 860–868 (2013).
289. Molmeret, M., Horn, M., Wagner, M., Santic, M. & Abu Kwaik, Y. Amoebae as Training Grounds for Intracellular Bacterial Pathogens. *Appl. Environ. Microbiol.* **71**, 20–28 (2005).
290. Laskowski-Arce, M. A. & Orth, K. Acanthamoeba castellanii promotes the survival of Vibrio parahaemolyticus. *Appl. Environ. Microbiol.* **74**, 7183–7188 (2008).
291. Baffone, W. *et al.* Detection of free-living and plankton-bound vibrios in coastal waters of the Adriatic Sea (Italy) and study of their pathogenicity-associated properties. *Environ. Microbiol.* **8**, 1299–1305 (2006).
292. Kaneko, T. & Colwell, R. R. Ecology of Vibrio parahaemolyticus in Chesapeake Bay. *J. Bacteriol.* **113**, 24–32 (1973).
293. Matz, C. *et al.* Marine biofilm bacteria evade eukaryotic predation by targeted chemical defense. *PLoS One* **3**, 1–7 (2008).
294. Page, R. & Holmes, E. *Molecular Evolution: A Phylogenetic Approach*. *Blackwell Science* (1998). doi:10.1046/j.1420-9101.1999.0072b.x.
295. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2015).
296. Reinoso Webb, C., Koboziev, I., Furr, K. L. & Grisham, M. B. Protective and pro-inflammatory roles of intestinal bacteria. *Pathophysiology* **23**, 67–80 (2016).
297. Katsonis, P. *et al.* Single nucleotide variations: Biological impact and theoretical interpretation. *Protein Sci.* **23**, 1650–1666 (2014).
298. Lebeuf-Taylor, E., McCloskey, N., Bailey, S. F., Hinz, A. & Kassen, R. The distribution of fitness effects among synonymous mutations in a gene under directional selection. *Elife* **8**, 1–16 (2019).
299. Ensminger, A. W., Yassin, Y., Miron, A. & Isberg, R. R. Experimental evolution of Legionella pneumophila in mouse macrophages leads to strains with altered determinants of environmental survival. *PLoS Pathog.* **8**, (2012).
300. Adiba, S., Nizak, C., van Baalen, M., Denamur, E. & Depaulis, F. From grazing resistance to pathogenesis: The coincidental evolution of virulence factors. *PLoS One* **5**, 1–10

(2010).
301. Long, M. *et al.* Infant Rabbit Model for Diarrheal Diseases. **8**, 444–454 (2016).
302. Darling, A. C. E. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
303. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
304. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
305. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
306. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: A genome comparison visualizer. *Bioinformatics* **27**, 1009–1010 (2011).