Occurrence and significance of pathogenicity and fitness islands in environmental vibrios

Savannah Klein†, Shannon Pipes† and Charles R. Lovell*

Abstract
Pathogenicity islands (PAIs) are large genomic regions that contain virulence genes, which aid pathogens in establishing infections. While PAIs in clinical strains (strains isolated from a human infection) are well-studied, less is known about the occurrence of PAIs in strains isolated from the environment. In this study we describe three PAIs found in environmental Vibrio vulnificus and Vibrio parahaemolyticus strains, as well as a genomic fitness island found in a Vibrio diabolicus strain. All four islands had markedly different GC profiles than the rest of the genome, indicating that all of these islands were acquired via lateral gene transfer. Genes on the PAIs and fitness island were characterized. The PAI found in V. parahaemolyticus contained the tdh gene, a collagenase gene, and genes involved in the type 3 secretion system II (T3SS2). A V. vulnificus environmental strain contained two PAIs, a small 25 kbp PAI and a larger 143 kbp PAI. Both PAIs contained virulence genes. Toxin–antitoxin (TA) genes were found in all three species: on the V. diabolicus fitness island, and on the V. parahaemolyticus and V. vulnificus PAIs.

Keywords: Vibrio parahaemolyticus, Vibrio vulnificus, Pathogenicity islands

Introduction
Vibrio parahaemolyticus and Vibrio vulnificus can cause illnesses in humans, with an estimated 80,000 cases occurring annually in the United States (Scallan et al. 2011; CDC 2017). The hospitalization and mortality rates of V. parahaemolyticus gastroenteritis are 22% and 1%, respectively (Scallan et al. 2011). Although cases are usually mild and tend to resolve themselves after 1–3 days, V. parahaemolyticus is responsible for the majority of vibriosis cases (Scallan et al. 2011). V. vulnificus cases are less common; only about 100 occur each year in the United States. However, the hospitalization and mortality rates of this bacterium are much higher, at 92% and 35%, respectively (Scallan et al. 2011). V. vulnificus also causes sepsis and necrotizing fasciitis if it enters the body through an open wound. The majority of reported V. vulnificus cases are from wound infections (45%) and septicemia (43%); only 5% are gastroenteritis (Scallan et al. 2011). The mortality rate of V. vulnificus when it invades the bloodstream (sepsis) increases to 60%. Pathogenesis of both species is complex, and while some virulence factor genes have been implicated, the mechanisms underlying V. vulnificus and V. parahaemolyticus virulence are not well understood (Broberg et al. 2011; Lovell 2017; Klein and Lovell 2016).

Pathogenicity islands (PAIs), a subgroup of genomic islands that aid in and contribute to pathogenesis, have been found in clinical strains of both V. vulnificus and V. parahaemolyticus. PAIs are large chromosomal regions that are flanked by tRNA genes, and are usually associated with mobile genetic elements, such as phage, plasmid, integron, and transposon genes. A genomic island must contain at least one virulence gene, or gene that contributes to pathogenesis, to be considered a PAI. The size of PAIs ranges from 10 to 200 kbp (Schmidt and Hensel 2004; Hacker and Kaper 2000; Hacker and Carniel 2000) and the average Vibrio genome is 4.5 mbp (Pipes et al. in preparation), meaning that a single PAI could make up as much as 4% of a Vibrio genome. PAIs are
flanked by highly conserved tRNA genes that act as both integration and excision sites. The majority (~75%) of PAIs discovered have tRNA flanking sequences (Schmidt and Hensel 2004; Hacker and Kaper 2000). Additionally, tRNA loci are often found on extrachromosomal elements, such as plasmids and bacteriophages. This indicates that the most likely mechanism for extrachromosomal element insertion is homologous recombination between the extrachromosomal element tRNA and PAI flanking tRNA loci (Hacker and Kaper 2000).

There is considerable evidence that PAIs are acquired horizontally via one or more lateral transfer events. Within some PAIs there is evidence of one large transfer event, while other PAIs are more “mosaic-like.” The “mosaic-like” composition of certain PAIs is caused by multiple, independent lateral transfer events (Hacker and Kaper 2000; Schmidt and Hensel 2004). PAIs usually differ in codon usage biases and have a markedly lower or higher GC content than the rest of the genome (Schmidt and Hensel 2004; Hacker and Kaper 2000; Hacker et al. 1997; Hacker and Carniel 2000). This supports the idea that recognizable PAIs are incorporated into a genome via lateral gene transfer from a dissimilar or unrelated organism (donor) having differing GC content and codon usage than the recipient (Schmidt and Hensel 2004). However, PAI GC content may not differ from that of the core genome if the donor and recipient microorganisms are closely related (Hacker and Kaper 2000). Dissimilarities in base composition confirm that detectable lateral transfer of PAIs must have been of recent origin, as insufficient time for genetic drift has passed (Schmidt and Hensel 2004).

PAIs have been found in clinical strains of V. vulnificus and V. parahaemolyticus (e.g. Makino et al. 2003; Wang et al. 2006; Sugiyma et al. 2008; Quirke et al. 2006; Cohen et al. 2007). Nine PAIs have been identified in V. parahaemolyticus, with VPAI-1 and VPAI-7 (V. parahaemolyticus pathogenicity island one and V. parahaemolyticus pathogenicity island seven) being the most studied (Ceccarelli et al. 2013). VPAI-1 is a 22 kbp island that is found on chromosome 1 in some strains, and chromosome 2 in others (Wang et al. 2006; Chen et al. 2011). This observation provides evidence for the mobility of this genomic island. VPAI-7 is the largest Vibrio genomic island found to date. This island contains the virulence factors TDH (thermostable direct hemolysin) and type III secretion system 2 (T3SS2) (Makino et al. 2003; Sugiyma et al. 2008). Other names for VPAI-7 include VPalx or tdhVPA (Xu et al. 2017) and parts of VPAI-7 have been found in other Vibrio species, such as Vibrio mimicus (Gennari et al. 2011).

Genomic islands have been found in V. vulnificus clinical strains YJ016 and CMCP6, with 14 regions ranging in size from 14 to 117 kbp. A superintergen (SI) and nine V. vulnificus genomic islands (VV1–1 to VVI–IX) have been found in these clinical strains. PAIs have not been detected in environmentally derived V. vulnificus strains (Quirke et al. 2006). V. vulnificus VVI–1 has been found in the Vibrio cholerae biotype El Tor and O139 serogroup. The functional role of this island has not been determined but its presence in V. cholerae supports the idea that these regions can be transferred to other closely related species (O’Shea et al. 2004).

Work on Vibrio PAIs is heavily skewed toward clinical strains, with the pathogenic potential of naturally-occurring (environmental) strains rarely considered. In this study, we characterized four genomic islands found in environmental Vibrio strains: a PAI within a V. parahaemolyticus strain, two novel PAIs within a V. vulnificus strain, and a novel fitness island found in a Vibrio diabolicus strain. Environmental Vibrio strains, and the PAIs within them, could serve as reservoirs for virulence genes.

Materials and methods
Strain isolation
Environmental V. parahaemolyticus and V. diabolicus strains were isolated previously (Gutierrez West et al. 2013; Klein et al. 2014) from the pristine North Inlet salt marsh estuary near Georgetown, SC, USA (33°20′N, 79°12′W). Environmental V. vulnificus strains were also isolated near Georgetown, SC; however, they were isolated from lower salinity waters in Winyah Bay and the Waccamaw River. Water samples were plated on CHROMagar Vibrio (DRG International, NJ, USA) for isolation of V. vulnificus strains following the US Food and Drug Administration protocol (DePaola and Kaysner 2004). Vibrio strains were routinely cultivated on saline Luria Agar (SLA; per L; 10 g tryptone, 5 g yeast extract, 27 g NaCl, 15 g Bacto Agar). V. parahaemolyticus TS-8-11-4 and V. diabolicus JBS-8-11-1 were deposited into the DSMZ Public Culture Collection and were assigned their respective accession numbers: DSM 107522 and DSM 107521.

Whole genome sequencing
Genomic DNA was isolated through the Wizard Genomic DNA Purification kit following the protocol for Gram negative organisms (Promega, Madison, WI, USA). After DNA was extracted, DNA quantity was measured via Quibit fluorimetry. Libraries were prepared and then sequenced using an Illumina MiSeq (V3 26300 base) at the Indiana University Center for Genomic Studies as a part of the Genome Consortium for Active Teaching NextGenSequencing Group (GCAT-SEEK) shared run (Buonaccorsi et al. 2011, 2014). Sequencing reads
were filtered (median phred score 0.20), trimmed (phred score 0.16), and assembled using the paired-end de novo assembly option in NextGENe V2.3.4.2 (SoftGenetics, State College, PA, USA). The assembled genomes were uploaded to the Rapid Annotation with Subsystem Technology (RAST) web service (Aziz et al. 2008; Overbeek et al. 2005, 2014) for analysis, guided contig reordering and assembly improvement. Genomes were aligned based on completed sequences using dotplot comparisons. Whole genome sequence data obtained from this work were submitted to the NCBI GenBank and assigned the accession numbers: PKQA00000000, PKPY00000000, and PKPZ00000000.

PAI detection and characterization
The fully sequenced genomes were uploaded to TUBIC (Tianjin University Bioinformatics Center) to determine their GC profiles (http://tubic.tju.edu.cn/). This tool displays GC content variation across a genome and can be useful for identifying genomic regions that differ from the rest of the genome in GC content (Gao and Zhang 2006). Genomic islands that were detected via TUBIC were isolated and the island nucleotide sequence was uploaded to RAST to identify and characterize the specific genes found on the genomic islands (http://rast.nmpdr.org/). NCBI GenBank was also used to characterize genomic island genes (http://www.ncbi.nlm.nih.gov/genbank/). Gene sequences of interest were edited and maximum-likelihood trees were constructed using the Kimura 2-parameter model with Mega version 7. DNAPlotter was used to visualize the circular chromosomes of the Vibrio strains (Carver et al. 2009).

Results
V. parahaemolyticus island
Vibrio parahaemolyticus strain TS-8-11-4 was isolated from salt marsh sediments (Gutierrez West et al. 2013; Klein et al. 2014) at the pristine North Inlet estuary in South Carolina, USA. This strain had a genome of 4.98 mbp; chromosome 1 was 3.19 mbp in length and chromosome 2 was 1.78 mbp in length. The majority of the genome contained a GC content of 45.57%, which is typical for V. parahaemolyticus (Farmer and Janda 2005). However, this strain contained a 223 kbp island that had a markedly lower GC content (41.5%) not typical of V. parahaemolyticus. The majority (69%) of genes on the TS-8-11-4 PAI could not be assigned specific identities and were thus designated hypothetical. The genomic island of V. parahaemolyticus TS 8-11-4 was on the second chromosome of and it harbored virulence genes (Fig. 1a). The virulence factor genes that were found on this island included the thermostable direct hemolysin gene, genes involved in the type three secretion system II (T3SS2), a collagenase gene, as well as capsule production genes.

V. vulnificus islands
Vibrio vulnificus strain WR-2-BW was isolated near Georgetown, SC from Waccamaw River waters. Its genome (4.96 mbp) contained two chromosomes, the first chromosome larger (2.96 mbp) than the second (1.99 mbp). The average GC content of V. vulnificus ranges from 46 to 48% (Farmer and Janda 2005), and the average GC content of strain WR-2-BW was 46.83%. Two regions within the genome had GC contents that were markedly lower from the rest of the genome. The first region had a GC content of 38.2% and the second

![Fig. 1](image.png)

Fig. 1  a–c Circular presentation of the second chromosome of a Vibrio parahaemolyticus environmental strain TS-8-11-4, b Vibrio vulnificus environmental strain WR-2-BW, and c Vibrio diabolicus environmental strain JBS-8-11-1. Track 1, forward coding sequences; track 2, reverse coding sequences; track 3, tRNA genes; track 4, red, pathogenicity islands, blue, genomic fitness islands; track 5, virulence and virulence-associated genes; track 6, genes involved in toxin–antitoxin systems; track 7, mobile genetic elements. Virulence and virulence-associated genes are numbered and are defined via the center text boxes.
region had a GC content of 42.5%; both of which are lower than the typical GC content of *V. vulnificus* strains. These regions were found on the second chromosome (Fig. 1b). The first region was a 25 kbp island and the second region was a 143 kbp island. The 25 kbp island was 30 genes in length, and had two genes that had virulence-related functions, which include a putative LPS biosynthesis protein gene and an O-antigen flippase *wzx* gene. The 143 kbp island contained the cytolysin gene *vvhB*, a chitinase gene, *tdlD*/*tdlE* proteolytic complex genes, and Type IV secretory pathway components. The 143 kbp genomic island was comprised of 160 genes in total, 63% of which were characterized as hypothetical or had unknown function.

**V. diabolicus island**

*Vibrio diabolicus* strain JBS-8-11-1 was isolated previously from North Inlet salt marsh sediments (Gutierrez West et al. 2013; Klein et al. 2014). Its genome (5.04 mbp) was comprised of two chromosomes, the first (3.23 mbp) being larger than the second (1.81 mbp). Its GC content was typical of other *V. diabolicus* genomes (44.91%) (Goudenege et al. 2014), except for a 182 kbp island, located on chromosome 2, which had a GC content of 40.8%. Eighty-two percentage of the island consisted of hypothetical genes. This island harbored no known virulence genes; it is hereafter referred to as a fitness island (Fig. 1c). Three genes, a phage DNA synthesis gene, a phage DNA replication gene, and a gene encoding a phage capsid protein, were located very close to each other on the fitness island. Thirteen genes involved in toxin–antitoxin (TA) systems were located on the fitness island.

**Discussion**

The genomic island of *V. parahaemolyticus* TS 8-11-4 was deemed a PAI due to the presence of virulence genes on this island, despite its environmental origin (Schmidt and Hensel 2004; Hasan et al. 2010; Dobrindt et al. 2004). The thermostable direct hemolysin gene (*tdh*) was found on this island, as well as genes involved in the type three secretion system II (T3SS2). Both the *tdh* gene and T3SS2 complex are the two major virulence factors implicated in *V. parahaemolyticus* pathogenesis (Makino et al. 2003; Park et al. 2004; Yanagihara et al. 2010). A collagenase gene was found on the island; collagenase is thought to be involved in *V. parahaemolyticus* virulence (Gode-Portratz et al. 2011). The genomic island of *V. parahaemolyticus* strain TS-8-11-4 is a PAI, and more specifically, because it contains *tdh* and T3SS2 genes, we designate this island as a VPAI-7 (VPaIα or VPAI-7 (VPaIα or tdh).

Four genes involved in capsule production, as well as one integrase gene, and a Na+/H+ antiporter (*nhaA*) were also found on this PAI. Capsules aid pathogens in evasion of host immune defenses, establishing infections, and survival in harsh environments, such as the stomach. *V. parahaemolyticus* virulence is correlated with capsule production (Broberg et al. 2011; Letchumanan et al. 2014). One capsule gene had high homology with Gram positive capsule production genes. This is interesting because vibrios are Gram negative organisms, so this gene may have been acquired laterally. An integrase gene was found near the center of the island. Integrase genes are associated with PAIs and function to integrate foreign DNA into the genome (Hacker and Kaper 2000). Usually VPAI-7 does not contain an integrase gene, but a few transposon genes instead (Ceccarelli et al. 2013). Finally, we determined that a *nhaA* gene is located on this genomic island. *nhaA* genes encode Na+/H+ antiporters, which transport ions to balance pH. Na+/H+ antiporters aid *V. cholerae* in environmental persistence (Vimont and Berche 2000) and are essential for *Yersinia pestis* virulence (Minato et al. 2013).

Similar to *V. parahaemolyticus*, the two islands found for the *V. vulnificus* WR-2-BW strain are characterized as PAIs due to the presence of virulence genes and virulence-related genes. Two of these genes had virulence-related functions, a putative LPS biosynthesis protein gene and an O-antigen flippase *wzx* gene. These genes are virulence-associated factors, as they do not directly cause host cell damage, but they do contribute to pathogenesis, aiding in the establishment of infections. Lipopolysaccharide (LPS) is a main component of the outer membrane of Gram negative bacteria, and is a known pyrogen (fever-producing agent) (McPherson et al. 1991; Jones and Oliver 2009). Phylogenies show that the LPS biosynthesis protein gene from *V. vulnificus* WR-2-BW was closely related to an LPS biosynthesis protein gene from a *Vibrio corallililyticus* species. The O-antigen flippase *wzx* gene is part of the major class of O-antigen gene clusters, and it encodes a hydrophobic protein with 12 potential transmembrane segments (Liu et al. 1996).

A cytolysin secretion gene, *vvhB*, was found also found on the 143 kbp *V. vulnificus* island. Cytolysins lyse erythrocytes by forming small pores in the cytoplasmic membrane or binding to cholesterol to interrupt potassium and sodium ion channels (Choi et al. 2002). In *V. vulnificus*, the expression and mechanism of cytolysins *vvhA* and *vvhB* are not fully understood, however, they are both believed to play a role in pathogenicity (Choi et al. 2002). They are homologous to a known *V. cholerae* El Tor hemolysin (Choi et al. 2002; Yamamoto et al. 1990). Phylogenies show that the *vvhB* gene in the *V. vulnificus*
WR-2-BW strain was 99% identical to other *V. vulnificus* *vvhB* genes from other strains.

Other genes of interest on the 143 kbp PAI include a chitinase gene, *tldD/tldE* proteolytic complex genes, and type IV secretory pathway components. In *Escherichia coli*, it was shown that the TldD and TldE proteins could be involved in regulating gyrase function as well as aiding in proteolytic activity (Allali et al. 2002). The chitinase gene had a 99% blast identity score to the chitinase gene found in the *V. vulnificus* YJ016 strain; however, the chitinase gene in YJ016 is located on the first chromosome and WR-2-BW’s chitinase gene is located in the second chromosome. Chitinous exoskeletal materials of invertebrates can be a source of carbon and nitrogen for bacteria; vibrios in particular have a well-known association with marine copepods (Kaneko and Colwell 1975; Lovell 2017). *V. cholerae* has a well-studied association with copepods, which commonly serve as a vector of cholera infections in Bangladesh water systems (Tamplin et al. 1990). Chitinase has been identified as part of the mechanism for adsorption and attachment to copepods, which relates to its ability to colonize its host and degrade the host exoskeleton, increasing the overall ecological fitness of the vibrios (Huq et al. 1983; Nalin et al. 1979; Bhowmick et al. 2006).

*Vibrio diabolicus* had a large genomic island that did not contain any virulence factors or virulence associated genes, which we defined as a fitness island, as it contained genes that would aid the organism in persistence in the environment. Toxin–antitoxin (TA) systems are found either on plasmids, genomic islands, or within the chromosome and are made up of closely linked toxin and antitoxin genes. The encoded labile antitoxin protects the host from the stable toxin, while competitor cells that do not have the TA system (and respective antitoxin) are eliminated (Hayes 2003; Van Melderen and Saavedra 2005).

---

**Fig. 2** Maximum-likelihood phylogeny (Kimura 2-parameter model) of *doc* toxin genes and *phd* antitoxin genes. Bold indicates sequences obtained from this study. The bootstrap values represent 1000 replications, and values of less than 50 are not shown. The reference sequences were acquired from NCBI GenBank.
Sometimes TA systems are referred to as “addiction modules” because the host cell is dependent on the antitoxin (Van Melderen and Saavedra 2009). The toxin and respective antitoxin loci are usually found neighboring each other, often overlapping (Hayes 2003). Seven type II TA toxins were found on JBS-8-11-1’s fitness island, along with their neighboring respective antitoxins. Type I TAs include RNA antitoxins, while type II TAs have protein antitoxins (Hayes 2003). The relE, yafQ, and yoeB toxin genes encode mRNA interferase endoribonucleases; all three of these toxin genes were detected on this fitness island. The doc toxin gene (death on curing) inhibits translation by blocking translation elongation at the 30S ribosomal subunit (Lui et al. 2008); three copies of the doc toxin gene and three copies of its antitoxin partner gene, phd (prevent host death) were found on JBS-8-11-1’s fitness island. doc toxin genes and phd antitoxin genes are widespread in vibrios and were also found on V. parahaemolyticus strain TS-8-11-4’s PAI as well as V. vulnificus strain WR-2-BW’s PAI (Fig. 2).

Lateral gene transfer in environmental strains
PAIs are present in environmental Vibrio strains and are most likely acquired via lateral gene transfer. All four of the islands described here have significant lower GC content than the rest of the genome, providing evidence that these islands originated from a foreign source and were transferred into these genomes relatively recently. Additional evidence includes mobile genetic elements, such as phage and plasmid genes, integrases, and transposons. Virulence loci on VPAI-7 have been detected in environmental species that do not cause human infections: Vibrio mimicus, Vibrio harveyi, and Vibrio natriegens (Gennari et al. 2011; Klein et al. 2014). Clearly, lateral transfer of individual virulence loci and/or entire PAIs is occurring between and among environmental vibrios. It is well documented that V. cholerae enters a natural competency state in the presence of chitin or under low-nutrient conditions (Hazan et al. 2010; Metzger and Blokesch 2016); however, less is known about uptake of exogenous DNA by other Vibrio species. Further studies examining the rates of lateral transfer among vibrios in the environment are needed. Vibrios survive, persist, and can undergo rapid population expansions (bloom) in coastal ecosystems. Consequently, the pathogenicity loci (and potential of said loci to be transferred laterally) of naturally occurring environmental strains are clearly important.

Abbreviations
PAI: pathogenicity island; TA: toxin–antitoxin; LPS: lipopolysaccharide; T3SS2: type three secretion system II; tidh: thermostable direct hemolysin gene.

Authors’ contributions
All authors (SK, SP, and CL) conceived and designed the experiments. SK was responsible for sample collection and strain isolation. SP performed the whole genome sequence annotation and assembly. Both authors (SK, SP) used bioinformatic techniques and tools to identify and characterize PAIs in environmental Vibrio genomes. CL directed the project and obtained funding; all authors (SK, SP, and CL) wrote the paper. All authors read and approved the final manuscript.

Acknowledgements
We would like to thank the Belle W. Baruch Institute for access to the pristine North Inlet Estuary near Georgetown, SC, USA. We would also like to thank Dr. Daniel Tufford for access to environmental V. vulnificus strains and Dr. Bob Friedman for helpful bioinformatic discussions.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
Data will be made available through publication and GenBank.

Compliance with ethical standards
This article does not contain any studies with human participants or animals performed by the author.

Consent for publication
All authors listed on this manuscript have read and agreed to the publication of this research.

Funding
Genomic sequencing was supported by the office of the Vice President for Research, University of South Carolina, 2015 South Carolina Floods Research Initiative.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 16 May 2018 Accepted: 19 October 2018
Published online: 30 October 2018

References
Allali N, Afif H, Couturier M, Van Melderen L (2002) The highly conserved TldD and TldE proteins of Escherichia coli are involved in microcin B17 processing and in CcdA degradation. J Bacteriol 184(12):3224–3231
Aziz RK, Bartels D, Best AA, DeJongh M, Disz T (2008) The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75
Bhowmick R, Ghosal A, Chatterjee NS (2006) Effect of environmental factors on expression and activity of chitinase genes of vibrios with special reference to Vibrio choleras. J Appl Microbiol 103:97–108
Broberg CA, Calder TJ, Orth K (2011) Vibrio parahaemolyticus cell biology and pathogenicity determinants. Microbes Infect 13:992–1001
Buonaccorsi VP, Boyle MD, Grove D, Paul C (2011) GCATSEEK: genome consortium for active teaching of undergraduates through increased faculty access to next-generation sequencing data. CBE Life Sci Educ 10:342–345
Buonaccorsi VP, Peterson M, Lamendella G, Newman JD, Trun N (2014) Vision change through the genome consortium for active teaching using next-generation sequencing (GCAT-SEEK). CBE Life Sci Educ 13:1–2
Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J (2009) DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25(1):119–120
Ceccarelli D, Hasan NA, Huq A, Colwell RR (2013) Distribution and dynamics of epidemic and pandemic Vibrio parahaemolyticus virulence factors. Front Cell Infect Microbiol. https://doi.org/10.3389/fcimb.2013.00097
Centers for Disease Control and Prevention (2017) Vibrio species causing vibriosis. Atlanta, GA. https://www.cdc.gov/vibrioidx/index.html
