O-antigen diversity and lateral transfer of the wbe region among Vibrio splendidus isolates

Citation: Wildschutte, H., Preheim, S. P., Hernandez, Y. and Polz, M. F. (2010), O-antigen diversity and lateral transfer of the wbe region among Vibrio splendidus isolates. Environmental Microbiology, 12: 2977–2987. doi: 10.1111/j.1462-2920.2010.02274.x

As Published: http://dx.doi.org/10.1111/j.1462-2920.2010.02274.x

Publisher: Blackwell Publishing

Persistent URL: http://hdl.handle.net/1721.1/60921

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Title: O-Antigen Diversity and Lateral Transfer of the *wbe* Region Among *Vibrio splendidus* Isolates

Running Title: O-antigen Diversity Among *Vibrio splendidus*

Hans Wildschutte[1*], Sarah Pacocha Preheim[1], Yasiel Hernandez[2], and Martin F. Polz[1]

[1]Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139.

[2]Department of Oceans and Human Health, RSMAS, University of Miami, 4600 Rickenbacker Causeway, Miami, FL 33149

* To whom correspondence should be addressed:

Civil and Environmental Engineering

Massachusetts Institute of Technology

77 Massachusetts Avenue, Building 48-208

Cambridge, MA 02139

Phone: (617) 253-8650

FAX: (617) 258-8850

22 Email: hansw@mit.edu

Summary

The O-antigen is a highly diverse structure expressed on the outer surface of Gram-negative bacteria. The products responsible for O-antigen synthesis are encoded in the *wbe* region, which exhibits extensive genetic diversity. While heterogeneous O-antigens are observed within *Vibrio* species, characterization of these structures has been devoted almost exclusively to pathogens. Here, we investigate O-antigen diversity among coastal marine *Vibrio splendidus*-like isolates. The *wbe* region was first identified and characterized using the sequenced genomes of strains LGP32, 12B01, and Med222. These regions were genetically diverse, reflective of their expressed O-antigen. Additional isolates from physically distinct habitats in Plum Island Estuary (MA, USA), including within animal hosts and on suspended particles, were further characterized based on multilocus sequence analysis (MLSA) and O-antigen profiles. Results showed serotype diversity within an ecological setting. Among 48 isolates which were identical in three MLSA genes, 41 showed *gpm* genetic diversity, a gene closely linked to the *wbe* locus, and at least 12 expressed different O-antigen profiles further suggesting *wbe* genetic diversity. Our results demonstrate O-antigen hyper-variability among these environmental strains and suggest that frequent lateral gene transfer generates *wbe* extensive diversity among *V. splendidus* and its close relatives.

Introduction

The O-antigen, a polysaccharide chain composed of repeated units of 2-6 sugars, protrudes from the surface of Gram-negative bacteria as the outermost portion of lipopolysaccharide (LPS). This outer membrane structure is in direct physical interaction

with the surrounding substrates and thus subject to environmental selective pressures. Consequently, O-antigens exhibit high diversity in basic composition and shape, largely due to the variation of monosaccharide building blocks, their linkage into repeat units, and the number of units (Reeves et al., 1996; Chatterjee and Chaudhuri, 2004). For example, hundreds of serotypes, or conspecific strains which encode and express distinct O-antigens, have been observed for *Escherichia coli* (Samuel and Reeves, 2003), *Salmonella enterica* (Popoff, 2001), and *Vibrio cholerae* (Chatterjee and Chaudhuri, 2004). This phenotypic diversity manifests in the *wbe* chromosomal region which ranges in size from ~40 to 70 kilobases (kb) reflecting differences in both shared and non-homologous gene content located within *wbe* regions. While shared *wbe* genes differ based on mutations, non-homologous genes result from lateral gene transfer (LGT) (Reeves et al., 1996; Stroeher et al., 1998).

Historically, O-antigen diversity among pathogenic bacteria was proposed to be influenced by selective pressure exerted by the host immune system in which strains expressing rare or novel structures evade immune detection and cause disease (Reeves, 1995). This hypothesis explains O-antigen diversity among pathogens that undergo phase variation which increases bacterial fitness by evasion within a host (Maskell et al., 1991; Meyer, 1991; Lukácová et al., 2008), but fails to explain serotype diversity among other pathogens that express stable O-antigens, such as *E. coli* O157, *S. enterica* serovar Typhi, and *V. cholerae* O1 and O139 which cause bacteremia, typhoid fever and cholera, respectively. Although conspecific strains may carry virulence genes, these serotypes are thought to be non-pathogenic (Guhathakurta et al., 1999; Bakhshi et al., 2008; Rahman et al., 2008; Ottaviani et al., 2009). Moreover, most isolates, including pathogenic ones,

spend the majority of their lifecycle in an environment not attributed to causing disease suggesting that other ecological selective pressures influence O-antigen diversity. For instance, O-antigen diversity among *S. enterica*, which spend most of its time as a gut commensal, may be maintained by intestinal amoeboid predation (Wildschutte et al., 2004; Wildschutte and Lawrence, 2007). Vibrios are marine microbes that have multiple lifestyles and survive either free-living, particle associated, or within animal hosts. Selective pressures may exist such as phage and protist predation, competition for attachment to particulate carbon sources in nutrient deprived waters, or from habitat differences encountered when traveling from hosts to the water column. Thus, knowledge of ecology may be necessary to understand bacterial genetic and structural diversity.

While O-antigen characterization has been well documented among individual pathogenic *Vibrio* strains including *V. cholera* O1 and O139 (Stroeher et al., 1998; Chatterjee and Chaudhuri, 2004), serotype diversity at the population level remains less studied. The *Vibrio splendidus* clade represents the dominant vibrioplankton group in the temperate coastal ocean (Thompson et al., 2004a; Thompson et al., 2004b; Thompson et al., 2005) and has been found free-living and associated with numerous marine substrates including suspended organic particles, zooplankton, mussels and crabs [Preheim *et al.*, submitted; (Thompson et al., 2005; Hunt et al., 2008)]. Since isolates survive in various habitats, O-antigen diversity may persist among strains because certain structures provide fitness benefits against different selective pressures. To initially characterize O-antigen diversity and establish that different serotypes occur among *V. splendidus*-like isolates, we used the published genomes of LGP32, 12B01, and Med222 to identify and define the *wbe* region and show that its genetic diversity reflects O-antigen differences. These

environmental strains were isolated from different geographic locations; LGP32 was isolated from an oyster pond in France, 12B01 from Plum Island Estuary (PIE) of coastal Massachusetts, and Med222 from the Mediterranean Sea (Le Roux et al., 2009). We extended this study to *V. splendidus*-like environmental isolates within the PIE to determine if O-antigen diversity persists among strains within a geographical area but from diverse marine habitats including different body regions of crabs and mussels, and zooplankton. Combined methods of multilocus sequencing analysis (MLSA) and O-antigen profiling were used to show that O-antigen hyper-variability exists among *V. splendidus*-like isolates. Sequence analysis of the *gpm* gene, a housekeeping gene closely linked to the *wbe* locus, was used to investigate LGT about the *wbe* region. Extensive *gpm* genetic divergence as well as phylogenetic incongruencies between MLSA and *gpm* tree topologies, suggest a more frequent transfer of the *wbe* region compared to MLSA housekeeping genes among our environmental isolates and with LGP32, 12B01, and Med222. Together, these methods provide an excellent means for discriminating between closely related isolates and may prove useful in linking bacterial diversity to ecological parameters.

Results

Genetic Diversity of the *V. splendidus wbe* Locus.

The *wbe* loci of the *V. splendidus*-like strains LGP32, 12B01, and Med222 were identified and determined to be bounded by the *gmhD* and *gpm* genes (Figure 1a). The *gmhD* gene product (also referred to as *rfaD*) encodes an epimerase involved in heptose synthesis and is required for core LPS in many Gram-negative bacteria (Coleman, 1985;

115 Stroeher et al., 1998). Among annotated vibrios, the *gmhD* ORF has been shown to have
116 strong linkage to the *wbe* region (Stroeher et al., 1998). Initially using *gmhD* as a guide,
117 we identified the *wbe* regions in LGP32, 12B01, and Med222. For each strain, this locus
118 was found on the larger of two chromosomes, which contains core loci involved in
119 cellular processing, signaling, and metabolism (Le Roux et al., 2009). The *wbe* regions
120 differ in size between strains by almost 20 kb: the 12B01 *wbe* is the largest at 54.4 kb,
121 Med222 is 43 kb, and LGP32 is 37 kb. Although the ORFs within these regions have
122 predicted functions in the synthesis, linkage, and modification of sugars, the wide range
123 in size is largely due to non-homologous *wbe* gene content between strains (Figure 1a).
124 While pairwise comparisons of ORFs flanking the *wbe* region were highly conserved,
125 many ORFs within our predicted *wbe* region were non-homologous with respect to each
126 region suggesting gain and/or loss through lateral gene transfer and further supporting our
127 identification of each *wbe* coding region.

128 Homologous ORFs were identified within the *wbe* region between strains with
129 LGP32 as a reference (Table 1). Separate analyses were conducted using 12B01 or
130 Med222 as the reference (Tables S1 and S2). Three gene groups show similarity among
131 the strains (indicated by gray shading in Figure 1a; also refer to Table 1). The first group
132 (I) is represented in LGP32 as ORFs labeled 1-7. Group I ORFs, which include the *gmhD*
133 gene required in LPS synthesis (see above) were found in all three strains, suggesting
134 conserved functions among these strains. Other predicted Group I gene products include
135 a regulator and a transferase. Given their conserved location relative to *gmdH,* these may
136 be involved in assembling heptose into core, which was found in LPS from all three
137 strains (Table 2). Group II (LGP32 ORFs 12-14), is shared between 12B01 and LGP32.

Gene products in this group have proposed functions in polysaccharide export. Interestingly, these ORFs were not identified in Med222, suggesting this strain uses a different system for O-antigen export. Finally, Group III (LGP32 ORFs 26-29), has homologues in both 12B01 and Med222; however, in these strains the ORFs are not adjacent to one another. Within Med222 Group III ORFs are represented as ORFs 8, 9, 13, and 14. In 12B01, these ORFs are observed twice, at ORFs 9-12 and 45-48, suggesting a duplication event or two independent transfers. The predicted functions of these genes are involved in the glucose and rhamnose synthesis pathways, which we verified to be incorporated into the O-antigen of each strain (Table 2). Besides these similarities, most ORFs among LGP32, 12B01 and Med222 are non-homologous genes with respect to each *wbe* region, and likely encode different proteins that help assemble diverse O-antigens. Taken together, our results indicate the overall *wbe* composition is diverse among these closely related strains.

The *wbe* loci of Gram-negative bacteria are typically marked by JUMP (Just Upstream of Many Polysaccharide regions) sites, which include a short conserved signal sequence for DNA uptake and are thought to be involved in LGT during transformation of competent cells (Hobbs and Reeves, 1994; Snyder et al., 2007). These short conserved sequences reside just prior to *wbe* regions of other vibrios (González-Fraga et al., 2008). Genome searching revealed JUMP sites to be exclusively located within our defined *wbe* region of LGP32, 12B01 and Med222, just prior to a series of ORFs transcribed in one direction (Figure 1b). The LGP32 JUMP site is located downstream of putative O-antigen transporter genes. In 12B01 and Med222, this sequence is immediately upstream of ORFs 8 and 9, respectively. Interestingly, 12B01 has another very similar JUMP sequence just

upstream of ORFs 45-48 which is homologous to the ORFs 9-12 (Figure 1a). The conserved JUMP site sequence and its location just prior to *wbe* gene clusters transcribed in the same direction suggest that these sites are involved in the transfer of multiple *wbe* encoded genes during a single LGT event.

O-Antigen Structural Variability Reflects *wbe* Genetic Diversity.

In other vibrios, the *wbe* gene region has been shown to encode proteins responsible for O-antigen synthesis (Stroeher et al., 1998; Chatterjee and Chaudhuri, 2004). Different structures are phenotypically manifested through the incorporation of dissimilar monosaccharides and their linkage into polysaccharide units. Thus, variation in *wbe* gene content (i.e., ORFs encoding monosaccharide synthesis, transferases, and transporters) is likely to influence the O-antigen expressed by a strain. Given the observed *wbe* genetic diversity between LGP32, 12B01, and Med222 (Figure1a and Tables 1, S1 and S2), we next analyzed the LPS core and O-antigen expressed by each strain through silver staining. This method allows visualization of differences in O-antigen repeat units through differential banding patterns, such that different profiles represent dissimilar O-antigens. Different O-antigen profiles were observed among LGP32, 12B01 and Med222 indicating each strain is of a distinct serotype (Figure 1c).

To address whether the differences in O-antigen profiles could be attributed to the inclusion of monosaccharides unique to each strain, the glycosyl residues belonging to the LGP32, 12B01 and Med222 O-antigens were determined through combined gas chromatography and mass spectrometry (Merkle and Poppe, 1994). For all strains, we were able to identify monosaccharides common to the LPS core (heptose and glucose), and those typically included in the O-antigen (galactose, rhamnose and ribose) (Table 2)

(Stroeher et al., 1998; Samuel and Reeves, 2003; Chatterjee and Chaudhuri, 2004). Overall, these shared residues represent most of the conserved regions among LGP32, 12B01, and Med222 (Figure 1a and Table 1). We also detected residues not shared by all strains. For example, glucuronic acid, which has been shown to be included in the O-antigen of other Gram-negatives (Samuel and Reeves, 2003; Chatterjee and Chaudhuri, 2004), was detected in 12B01, and an unidentified amino sugar was unique to Med222 (Table 2). These residues are likely to contribute at least partially to the observed differences in O-antigen structures (Figure 1c). Together, these results support that *wbe* genotypic diversity contributes to phenotypic diversity between serotypes.

Serotype Diversity Among Closely Related *V. splendidus*-like Isolates.

O-antigen diversity was observed among *V. splendidus*-like strains LGP32, 12B01 and Med222 (Figure 1) which were originally isolated from diverse geographical regions (Table S3) (Le Roux et al., 2009). Our recent study of population-level diversity among vibrios in the PIE affords the opportunity to determine O-antigen diversity among closely related, co-existing strains (Preheim *et al.*, submitted). We chose 114 representatives within the *V. splendidus* clade from several marine habitats (Table S3), including zooplankton, crabs, and mussels (Pacocha, et al. 2010) to investigate serotype diversity.

As an estimate of overall relatedness of these 114 strains, concatenated nucleotide sequences of the *adk*, *hsp60*, and *mdh* housekeeping genes were used for MLSA and a maximum likelihood tree was generated (Figure 2a). Isolates had either different sequence types (ST) (n=37) meaning they were closely related based on nucleotide changes within the genes used for MLSA or they shared a ST with another strain (n=77) suggesting genetically identical or clonal isolates. Overall, we observed relatively little

genetic divergence among all these strains, as inferred from branch lengths and position relative to LGP32, 12B01 and Med222, thus limiting the time-scale for genome and O-antigen variation to accrue. Serotype diversity was characterized by visualizing O-antigen profiles for a total of 53 PIE isolates consisting of 37 with different STs as well as 16 that shared STs (Figure 2a); another 61 isolates that shared either ST 3, 12, or 243 were characterized in separate analyses (see below). Silver staining revealed at least 9 different O-antigen structures from these isolates (Figures S1), and in some cases isolates of the same ST were found to express different structures (for example, 9ZC32 and 9ZC73; 9CH134 and 9CHC140), thus confirming multiple serotypes of closely related *V. splendidus*-like strains are present within PIE.

Given the number of O-antigen structures observed within the closely related isolates, and that some strains with the same ST showed different O-antigen profiles (Figures S1 and 2a), we next examined serotype diversity among strains having the same ST to further constrain the time scale of O-antigen variation. Strains with ST 3 (n=25) and ST 12 (n=23) were isolated from multiple habitats including crabs, mussels, and zooplankton while ST 243 (n=13) originated from one individual crab. Silver staining was used to examine serotype diversity among isolates belonging to each ST (Figure 3). For ST 243 we observed no differences in O-antigen banding patterns (Figure 3a) possibly due to clonal expansion within a single host specimen. Surprisingly, a variety of O-antigen profiles were observed for ST 3 (Figure 3b and c) and 12 (Figure 3d and e), and overall we estimate at least 12 different serotypes within these groups alone. This is interpreted as a conservative estimate since profiles that appear similar may not absolutely represent the same O-antigens. These results demonstrate that diverse *V.*

splendidus-like serotypes occur within PIE, and further suggest O-antigen hyper-variability among strains, as isolates of identical ST can express distinct structures.

Genetic Diversity of the *gpm* Gene and *wbe* LGT.

Our sequence analysis of the LGP32, 12B01 and Med222 genomes led to the identification of putative JUMP sites within the *wbe* regions (Figure 1a and b), which have been implicated in LGT between other bacteria (Hobbs and Reeves, 1994). As a means to investigate if LGT is a possible mechanism of *wbe* diversity and to discriminate between serotypes having the same ST, we performed a phylogenetic analysis of *gpm*, a housekeeping gene in close linkage to *wbe* and required in glycolysis (Figure 1a). Overall, we observed a ~6-fold increase in divergence of *gpm* coding sequence (4.6%), compared to MLSA divergence (0.81%) among our environmental strains (Table S4). Furthermore, 81 unique STs were observed among the sample set using *gpm* sequence analysis while only 37 unique STs were identified based on *adk*, *hsp60* and *mdh* alone.

Using an approach similar to our initial MSLA with the *adk*, *hsp60* and *mdh* *genes*, we generated a *gpm*-based maximum likelihood tree as a means to infer strain relatedness with respect to *wbe* (Figure 2b). Compared to the MLSA generated tree, the *gpm* gene tree exhibits longer branch lengths as a result of increased *gpm* nucleotide change. More importantly, the topology of the *gpm* tree (Figure 2b) differs from the MLSA tree (Figure 2a). For instance, strains with identical sequences based on MLSA genes (ST 3 and 12) are scattered throughout the *gpm* based tree, showing they are not genetically identical even if they have similar O-antigen profiles; and LGP32, 12B01 and Med222 appear more closely related to environmental isolates in the *gpm*-based tree as opposed to MLSA. In addition, strain 9CSC152 is more closely related to the SWAT3

outgroup based on the *gpm* gene (Figure 2b) than to the PIE environmental isolates (Figure 2a). These results suggest that *wbe* transfer occurs frequently across the *V. splendidus* clade. Of the 61 strains belonging to STs 3, 12, and 243, a total of 41 unique *gpm* sequences were observed: 24 of 25 isolates for ST 3, and 17 of 23 for ST 12 (Figure 2b). Identical strains based on *gpm* were mostly of ST 243, again suggesting clonality. In combination with the identification of *wbe* JUMP sites within the available *V. splendidus*-like genomes, O-antigen hyper-variability among PIE isolates with the same STs, and *gpm* gene diversity along with tree incongruencies between MLSA and *gpm* sequences, these data indicate frequent LGT of *wbe* loci within the PIE marine column resulting in multiple *V. splendidus*-like serotypes.

Discussion

Marine bacteria constantly encounter diverse habitats while carried through the water column. Ecological selective pressures ranging from predation to surface adherence likely exist on spatial scales and may influence O-antigen diversity among serotypes. The ability to change an O-antigen through LGT of the *wbe* region may offer advantages in fitness across diverse environments. Using the sequenced genomes of LGP32, 12B01, and Med222, we showed characteristics of LGT such as non-homologous genetic differences between strains and the presence of JUMP sites, which are believed to facilitate *wbe* gene transfer (Figure 1). With the acquisition of *wbe* regions, entire functional pathways involving the synthesis of different O-antigen structures can be gained with the potential result of serotype conversion. We have previously shown that *V. splendidus* is found in different marine environments such as free-living within the water

column, attached to suspended particles, and on marine hosts [Preheim et al., submitted, (Thompson et al., 2005; Hunt et al., 2008)]. We suggest that the acquisition and expression of different *wbe* regions among *V. splendidus* and its close relatives could influence bacterial fitness through environmental interactions by the O-antigen resulting in the maintenance of O-antigen diversity.

To investigate LGT among environmental *V. splendidus*-like strains, we chose closely related and even identical strains based on MLSA to constrain O-antigen variability. Related strains were on average 0.81% divergent based on the concatenated *adk, hsp60,* and *mdh* sequences consisting of 1254 base pairs (Table S4), while strains having the same sequences (such as ST 3 and 12 strains) were devoid of mutations. Even with this mutational constraint, extensive genetic diversity was observed in the *gpm* gene amongst ST 3 and ST 12 isolates –an average and maximum nucleotide divergence of *gpm* was 5.25% and 13.5%, respectively. Because *gpm* is closely linked to the *wbe* region (Figure 1), selective sweeps are precluded and *gpm* diversity is likely maintained through hitchhiking with the *wbe* locus. Furthermore, incongruencies between MLSA and *gpm* phylogenies (Figure 2) and the presence of disparate O-antigens within a ST (Figure 3d and e) suggest that LGT occur at the *wbe* chromosomal location.

High rates of transfer within the *wbe* region, as suggested by extensive genetic diversity of the *gpm* gene, provide a means for serotype selection. We did not observe a predominant serotype among or within hosts (except for one crab where clonal expansion of strains with ST 243 is evident) which supports our recent study that *V. splendidus* are generalists among invertebrate hosts (Preheim *et al.*, submitted). However, we did observe closely related serotypes (expressing the same O-antigen) among different hosts.

For instance, strains 9CS34 (ST 12), 9CG23 (ST 3), and 9CSC94 (ST 3) from crab specimens 2, 2, and 5, respectively, were of the same serotype; and 9CG33 (ST 12), 9MHC17 (ST 12), and 9MHC23 (ST 12) from different mussels and a crab were of another serotype. If O-antigen selection occurs in the water column, prior to association within a host, then closely related serotypes could be found dispersed among different marine invertebrates. Continued studies to identify possible selective pressures influencing O-antigen and *wbe* diversity in the marine environment are being investigated.

Serotypes expressing the same O-antigen usually have different *gpm* sequences resulting from mutations within *gpm* or because of its close linkage to *wbe* making it susceptible to lateral transfer while preventing *gpm* selective sweeps. However, it is possible for strains to have the same *gpm* sequences yet dissimilar O-antigens. We amplified 483 base pairs of the *gpm* gene starting 168 base pairs downstream from the start site; if recombination occurs before the amplified region or within the *wbe* locus, then *gpm* gene sequences may be identical. For instance, a group of six strains from crab specimen 7 were identical based on *gpm* gene analysis (Figure 2b). It was expected that these were clonal isolates because all were of ST 12 based on MLSA; however, the O-antigen profiles among these strains differ. For example, 9CHC127, 9CHC133, and 9CS146 show one profile, while 9CSC139, 9CSC158, 9CS151 show another (Figure 3d and e). This is also seen with (1) 9CS134 and CS126 and (2) 9CS24 and 9MG29 which have the same *gpm* sequence but dissimilar O-antigen profiles. These results suggest that LGT occurred within the *wbe* region without involving the *gpm* gene. Furthermore, we would predict that genetic diversity of genes surrounding the *wbe* region would decrease with distance from the *wbe* locus if this region is under strong selection.

Our results suggest that the O-antigen hyper-variability observed among environmental *V. splendidus*-like serotypes reflects LGT-driven diversity of the *wbe* region. Frequent *wbe* transfer is evident among these strains and as well as the more distantly related LGP32, 12B01, Med222, all within the *V. splendidus* clade. The selective pressures that maintain O-antigen diversity remain unknown but may be related to phage infection, protist predation, or ecological interactions during life history in the water column. MLSA approaches that include loci with hyper-variable outer membrane structures have improved capacity to discriminate among otherwise identical STs and can provide greater insight into ecologically relevant differentiation among closely related strains.

Experimental Procedures

Strain Isolation and Growth Media.

Water samples and invertebrates were collected from Plum Island Sound Estuary, Ipswich, MA in the spring and fall of 2008 as described in (Preheim *et al.*, submitted). Briefly, seawater samples were collected at high tide in 4 L bottles from the shore. Zooplankton was isolated by filtering 100 L of seawater through a 64 μm mesh net. Samples were rinsed three times with sterile seawater, washed into a 50 ml conical tube and kept at ambient temperature in the dark until processing ~2 hours later. Living and dead zooplankton were differentiated by eye under a dissecting microscope based on movement and 10-140 individuals of each category were picked from each 100 L concentrate. Collections also included four male green crabs (*Carcinus maenas*); eight male and four female shore crabs (*Hemigrapsus sanguineus*; and sixteen blue mussels

(*Mytilus edulis*). All animals were washed with sterile seawater and placed in a whirl pack and cooler until processing. For crabs, gill (one brachia), stomach (entire tissue) and hindgut (~4 cm beginning with anus) were collected following stunning prior to dissection (no anesthesia). For mussels, approximately 1.5 cm^2 of gill and hindgut (including the anus) tissue was collected. For both crabs and mussels, gastrointestinal (GI) contents were collected by flushing tissue with 4 ml sterile seawater with a syringe. Tissues were washed 3x with sterile seawater to ensure only attached bacteria were collected. Crab and mussel tissue and GI tract contents samples were homogenized in a tissue grinder, serially diluted (10- to 10,000-fold) in sterile seawater, and plated for isolation on *Vibrio*-selective marine TCBS media (BD Difco TCBS + 1% NaCl). A total of 160 isolates were picked from each sample type per season (20 per specimen) using the most dilute samples with sufficient growth. Isolated colonies were re-streaked 3x alternating 1% TSB media (BD Bacto + 2% NaCl) and marine TCBS media to ensure purity of isolates.

PCR Amplification for MLSA and *gpm* Analysis.

Partial amplification of the heat shock protein (*hsp60*), adenylate kinase (*adk*), and malate dehydrogenase (*mdh*) genes were performed with all isolates for MLSA. Primers were as follows: *adk*, 5'GTATTCCACAAATYTCTACTGG3' and 5'GCTTCTTTACCGTAGTA3; *hsp60*, 5'GAATTCGAIIIIGCIGGIGAYGGIACIACIAC3' and 5'CGCGGGATCCYKIYKITCICCRAAICCIGGIGCYTT; *mdh*, 5'GATCTGAGYCATATCCCWAC3' and 5'GCTTCWACMACYTCRGTACCCG3'. PCR amplification was carried out as previously described with annealing temperature at

41°C for *adk* and *hsp60* and 60°C for *mdh* and sequences were submitted to GenBank (Preheim *et al.*, submitted). For *gpm* gene analysis, partial gene amplification was performed using primers 5'GATGGYCAAATGGGTAACTC3' and 5'CAGCACGGTAGTTCATGAAG3'. PCR amplification was carried out for 30 cycles with an annealing temperature of 60°C. Amplicons were sequenced bidirectionally at the Bay Paul Center at the Marine Biological Laboratory, Woods Hole, MA with the same primers for each respective gene. *gpm* sequences were submitted to GenBank under accession numbers GU990234-GU990351.

Phylogenetic Tree Construction and Gene Divergence.

Concatenated *adk*, *hsp60*, and *mdh* for MLSA and single *gpm* gene sequences were used to generate sequence alignments and gene divergence matrices using default parameters in ClustalX. The Vibrionales bacterium SWAT-3 was used as the outgroup. Maximum likelihood trees were constructed from the alignment using PhyML set with HKY85 substitution parameters (Guindon and Gascuel, 2003). Bootstrapping was performed in 100 replicates and values >70% are shown.

Whole Cell Lysates and Silver Stain for Estimation of O-antigen Diversity.

Strains were grown overnight in 5 mL of TSB at room temperature (RT). When cultures reached an OD_{600} of 1.0, 1 mL of cells were aliquoted and spun at 13,000 rpm for 3 min. Cells were resuspended in 100 µl lysis buffer (1M Tris HCl; pH 6.8, 2% SDS; 4% β-mercaptoethanol; 10% glycerol), incubated at 100°C for 10 minutes, and then cooled to below 60°C. Lysates were treated with 1.3 µl of 20 mg/ml of proteinase K and incubated for 1 hr at 55 ^{o}C. Bromophenol blue was added to each lysate for visualization, and 14 µl each loaded to a precast 10-20% tricine Novex gel. Following electrophoresis, gels were

391 silver stained as previously described to visualize the O-antigen (Hitchcock and Brown,
392 1983). Briefly, each gel was fixed in 40% ethanol and 5% acetic acid for 1 hr, oxidized in
393 the fixative with 0.7% periodic acid, and then incubated in silver stain (0.6% silver nitrate;
394 0.14 M NaOH; 1 ml 37% ammonium hydroxide) for 10 min. Gels were developed by
395 incubation for 2 min in developing buffer (50 μM citric acid and 0.7% formaldehyde in
396 200 ml volume) at 40°C. All gels were repeatedly washed after each step as described.

397 **O-Antigen Glycosyl Composition Analysis.**

398 Procedures were carried out as previously described (Merkle and Poppe, 1994). Single
399 colonies of LGP32, 12B01, and Med222 were picked and grown overnight at RT in 75
400 mls of Difco TSB media. Cultures were pelleted by centrifugation for 10 min at 10,000
401 rpm and resuspended in 1 ml of water. Five 5 mls of 95% ethanol was added and cells
402 were incubated at room temperature for an hour. Each cell suspension was pelleted and
403 the supernatant removed. The Complex Carbohydrate Research Center at the University
404 of Georgia determined core and O-antigen glycosyl residues after acid hydrolysis of
405 purified LPS.

406 **Analysis of the *wbe* region.**

407 The *gmhD* gene from *V. cholera* strain N16961 (locus tag VC0240) was used to
408 determine the presence and location of *gmhD* gene in the sequenced genomes of *V.*
409 *splendidus*-like strains LGP32, 12B01, and Med222. We identified the *gmhD* gene in
410 each genome screened (LGP32, YP_002415885; 12B01, ZP_00989916; and Med222,
411 ZP_01065583) and used its location as a reference point to manually analyze adjacent
412 open reading frames (ORFs) for predicted functions involved in synthesis, linkage, and

modification of sugars. ORFs bounded by *gmhD* and *gpm* represented the *wbe* region of each strain.

Acknowledgements

The authors acknowledge Dr. Frederique Le Roux and Dr. Jarone Pinhassi for kindly sharing strains LGP32 and Med222, respectively. We would also like to sincerely thank Julia Wildschutte for her careful and considerate manuscript critiques. The project described was supported by the grant number F32GM084640 from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. Isolation of the O-antigen and its sugar determination was supported in part by the Department of Energy-funded (DE-FG09-93ER-20097) Center for Plant and Microbial Complex Carbohydrates. Additional funding was from the National Science Foundation supported Woods Hole Center for Oceans and Human Health (COHH) and grants from the Gordon and Betty Moore Foundation and the Department of Energy Genome to Life (GTL) program to M.F.P.

References

Bakhshi, B., Pourshafie, M.R., Navabakbar, F., Tavakoli, A., Shahcheraghi, F., Salehi, M. et al. (2008) Comparison of distribution of virulence determinants in clinical and environmental isolates of *Vibrio cholera*. *Iranian Biomedical Journal* **12**: 159-165.

436 Chatterjee, S.N., and Chaudhuri, K. (2004) Lipopolysaccharides of *Vibrio cholerae*: II.
437 Genetics of biosynthesis. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of*
438 *Disease* **1690**: 93-109.
439
440 Coleman, W.G. (1985) The *rfaD* Gene Codes for ADP-L-glycero-D-mannoheptose-6-
441 epimerase. *The Journal of Biological Chemistry* **258**: 1985-1990.
442
443 González-Fraga, S., Pichel, M., Binsztein, N., Johnson, J.J., Morris, J.G.J., and Stine, O.C.
444 (2008) Lateral gene transfer of O1 serogroup encoding genes of *Vibrio cholerae*. *FEMS*
445 *Microbiology Letters* **286**: 32-38.
446
447 Guhathakurta, B., Sasmal, D., Pal, S., Chakraborty, S., Nair, G.B., and Datta, A. (1999)
448 Comparative analysis of cytotoxin, hemolysin, hemagglutinin and exocellular enzymes
449 among clinical and environmental isolates of *Vibrio cholerae* O139 and non-O1, non-
450 O139. *FEMS Microbiology Letters* **179**: 401-407.
451
452 Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate
453 large phylogenies by maximum likelihood. *Systems Biology* **52**: 696-704.
454
455 Hitchcock, P.J., and Brown, T.M. (1983) Morphological heterogeneity among *Salmonella*
456 lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. *Journal of*
457 *Bacteriology* **154**: 269-277.
458

Hobbs, M., and Reeves, P.R. (1994) The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. *Molecular Microbiology* **12**: 855-856.

Hunt, D.E., David, L.A., Gevers, D., Preheim, S.P., Alm, E.J., and Polz, M.F. (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 1081-1085.

Le Roux, F., Zouine, M., Chakroun, N., Binesse, J., Saulnier, D., Bouchier, C. et al. (2009) Genome sequence of *Vibrio splendidus*: an abundant planctonic marine species with a large genotypic diversity. *Environmental Microbiology* **11**: 1959–1970.

Lukácová, M., Barák, I., and Kazár, J. (2008) Role of structural variations of polysaccharide antigens in the pathogenicity of Gram-negative bacteria. *Clinical Microbiology and Infection* **14**: 200-206.

Maskell, D.J., Szabo, M.J., Butler, P.D., Williams, A.E., and Moxon, E.R. (1991) Phase variation of lipopolysaccharide in *Haemophilus influenzae*. *Research in Microbiology* **142**: 719-724.

Merkle, R.K., and Poppe, I. (1994) Carbohydrate composition analysis of glycoconjugates by gas-liquid chromatography/mass spectrometry. *Methods in Enzymology* **230**: 1-15.

Meyer, T.F. (1991) Evasion mechanisms of pathogenic *Neisseriae*. *Behring Institute Mitteilungen* **88**: 194-199.

Ottaviani, D., Leoni, F., Rocchegiani, E., Santarelli, S., Masini, L., Di Trani, V. et al. (2009) Prevalence and virulence properties of non-O1 non-O139 *Vibrio cholerae* strains from seafood and clinical samples collected in Italy. *International Journal of Food Microbiology* **132**: 47-53.

Popoff, M.Y. (2001) *Antigenic Formulas of the Salmonella Serovars, 8th edition.* Paris: Institut Pasteur.

Preheim, S., Boucher, Y., Wildschutte, H., David, L., Veneziano, D., Alm, E., and Polz, M.F. (2010) Metapopulation structure of Vibrionaceae among coastal marine invertebrates. *Submitted to Environmental Microbiology*.

Rahman, M.H., Biswas, K., Hossain, M.A., Sack, R.B., Mekalanos, J.J., and Faruque, S.M. (2008) Distribution of genes for virulence and ecological fitness among diverse *Vibrio cholerae* population in a cholera endemic area: tracking the evolution of pathogenic strains. *DNA and Cell Biology* **27**: 347-355.

Reeves, P. (1995) Role of O-antigen variation in the immune response. *Trends in Microbiology* **3**: 381-386.

505 Reeves, P.R., Hobbs, M., Valvano, M.A., Skurnik, M., Whitfield, C., Coplin, D. et al.
506 (1996) Bacterial polysaccharide synthesis and gene nomenclature. *Trends in*
507 *Microbiology* **4**: 495-503.
508
509 Samuel, G., and Reeves, P. (2003) Biosynthesis of O-antigens: genes and pathways
510 involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydrate*
511 *Research* **338**: 2503-2519.
512
513 Snyder, L.A.S., McGowan, S., Rogers, M., Duro, E., O'Farrell, E., and Saunders, N.J.
514 (2007) The repertoire of minimal mobile elements in the *Neisseria* species and evidence
515 that these are involved in horizontal gene transfer in other bacteria. *Molecular Biology*
516 *and Evolution* **24**: 2802-2815.
517
518 Stroeher, U.H., Jedani, K.E., and Manning, P.A. (1998) Genetic organization of the
519 regions associated with surface polysaccharide synthesis in *Vibrio cholerae* O1, O139
520 and *Vibrio anguillarum* O1 and O2: a review. *Gene* **223**: 269-282.
521
522 Thompson, F.L., Iida, T., and Swings, J. (2004a) Biodiversity of vibrios. *Microbiology*
523 *and Molecular Biology Reviews* **68**: 403-431.
524
525 Thompson, J.R., Randa, M.A., Marcelino, L.A., Tomita-Mitchell, A., Lim, E., and Polz,
526 M.F. (2004b) Diversity and dynamics of a north atlantic coastal *Vibrio* community.
527 *Applied and Environmental Microbiology* **70**: 4103-4110.

Thompson, J.R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D.E., Benoit, J. et al. (2005) Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**: 1311-1313.

Wildschutte, H., and Lawrence, J.G. (2007) Differential *Salmonella* survival against communities of intestinal amoebae. *Microbiology* **153**: 1781-1789.

Wildschutte, H., Wolfe, D.M., Tamewitz, A., and Lawrence, J.G. (2004) Protozoan predation, diversifying selection, and the evolution of antigenic diversity in *Salmonella*. *Proceedings in the National Academy of Science USA* **101**: 10644-10649.

Figure Legends

Figure 1. The *wbe* genotypic and O-antigen phenotypic diversity of *V. splendidus* strains 12B01, Med222, and LGP32. (A) The regions exhibit extensive genetic diversity between the *gmhD* and *gpm* flanking genes. Each *wbe* region encodes similar and different genes whose putative functions are O-antigen construction. Rectangular boxes represent ORFs. ORFs depicted above and below the respective genome baseline indicated forward and reverse transcription, respectively. Grey lines between genomes indicate homology between those genes. Black bars above LGP32 ORFs identified as I, II, and III indicated regions of shared homology with other *wbe* loci. Open and closed circles represent JUMP sites. (B) JUMP sites shown for *V. cholera* 01, 12B01, Med222, and LGP32 which contains the conserved DNA uptake signal sequenced (USS). Circles represent respective JUMP sequence locations in the *wbe* region. Bold and shaded sequences represent the conserved USS in *V. cholera* and *V. splendidus*, respectively. (C) Silver stain showing different O-antigen profiles; lanes 1, molecular marker; 2, *Salmonella enterica* LT2; 3, *Escherichia coli* K12; 4, 12B01; 5, Med222; and 6, LGP32.

Figure 2. Maximum likelihood trees of *V. splendidus*-like strains isolated from different marine habitats. (A) Phylogenetic relatedness based on MLSA of concatenated *adk*, *mdh* and *hsp60* partial gene sequences consisting of 1254 base pairs. The strains with ST 3, 12 or 243 are boxed grey and their ST# is present after their strain name. Sequenced genomes are bolded and marked with a *. (B) Phylogenetic relatedness based on *gpm* partial gene sequence consisting of 483 base pairs. Strains are labeled

564 according to the season and animal sample of isolation: Fall and spring is designated by
565 9 or 4, respectively, and the specific animal sample is identified by CG, crab gills; CH,
566 crab intestines; CHC, crab intestinal lining; CS, crab stomach; CSC, crab stomach lining;
567 MHC, mussel intestinal lining; ZC, zooplankton. Colors represent the individual host or
568 zooplankton sample they were isolated from.
569
570 **Figure 3. Silver stains showing the O-antigen profiles of *V. splendidus*-like**
571 **environmental isolates.** (A) Strains isolated from an individual crab host having MLSA
572 ST 243 express the same O-antigen. (B and C) ST 3 and (D and E) ST 12 strains isolated
573 from either crabs, mussels, or zooplankton show similar and different O-antigen profiles.
574 Strains were isolated from different individual hosts as indicated by numbers and strains
575 nomenclature, as described in Figure 2.