Post-Genomic Analysis of Members of the Family Vibrionaceae

E. FIDELMA BOYD, MEGAN R. CARPENTER, NITYANANDA CHOWDHURY, ANALUISA L. COHEN, BRANDY L. HAINES-MENGES, SAI S. KALBURGE, JOSEPH J. KINGSTON, J.B. LUBIN, SERGE Y. ONGAGNA-YHOMBI, and W. BRIAN WHITAKER

Department of Biological Sciences, University of Delaware, Newark, DE 19716

ABSTRACT
Similar to other genera and species of bacteria, whole genomic sequencing has revolutionized how we think about and address questions of basic Vibrio biology. In this review we examined 36 completely sequenced and annotated members of the Vibrionaceae family, encompassing 12 different species of the genera Vibrio, Aliivibrio, and Photobacterium. We reconstructed the phylogenetic relationships among representatives of this group of bacteria by using three housekeeping genes and 16S rRNA sequences. With an evolutionary framework in place, we describe the occurrence and distribution of primary and alternative sigma factors, global regulators present in all bacteria. Among Vibrio we show that the number and function of many of these sigma factors differs from species to species. We also describe the role of the Vibrio-specific regulator ToxRS in fitness and survival. Examination of the biochemical capabilities was and still is the foundation of classifying and identifying new Vibrio species. Using comparative genomics, we examine the distribution of carbon utilization patterns among Vibrio species as a possible marker for understanding bacteria-host interactions. Finally, we discuss the significant role that horizontal gene transfer, specifically, the distribution and structure of integrons, has played in Vibrio evolution.

INTRODUCTION
Within the family Vibrionaceae, 12 genera have thus far been proposed: Aliivibrio, Allomonas, Beneckea, Catenococcus, Echinimonas, Enterovibrio, Grimontia, Listonella, Lucibacterium, Photobacterium, Salinivibrio, and Vibrio (1). Nearly 140 species have been described and over 110 species names with standing in nomenclature are found within the genus Vibrio (2, 3, 4). Members of the genus Vibrio are abundant in various marine habitats worldwide and are associated with a wide range of living organisms, from corals, sea grass, and sponges to zooplankton, shellfish, and fish, as well as humans. Several species are pathogenic to aquatic life and humans, commonly in the form of vibriosis and gastrointestinal illnesses, respectively. One of the most significant human infections is the diarrheal disease cholera. In 2011 the World Health Organization (WHO) reported that global cholera incidents were on the rise, causing an estimated 3 million to 5 million cases and 100,000 to 200,000 deaths per year. Many Vibrio species are commensals or symbionts, while many are also halophiles and are capable of bioluminescence, and most are motile. The ecological diversity of Vibrionaceae is propelled in part by phenotypic innovations driven by both adaptive changes and horizontal gene transfer.
transfer (HGT). The accessibility and rapid acquisition of whole genome sequences now allows us to investigate the evolutionary events that contribute to *Vibrio* diversity, versatility, and abundance.

As of May 2013, the National Center for Biotechnology Information (NCBI) genome database contained 36 completely sequenced and annotated members of the *Vibrionaceae* family, encompassing 12 different species of the *Vibrio*, *Aliivibrio*, and *Photobacterium* genera. All species examined to date have two chromosomes, which is now an accepted characteristic of this family. There are currently nearly 700 genome sequences in progress of members of the *Vibrionaceae* family. General characteristics of the completely sequenced genomes are in progress of the *Vibrionaceae* family. The Jukes-Cantor substitution model was used to compute evolutionary distances (8). The bootstrap test was repeated 1,000 times and values over 50% are shown next to branches (9). The concatenated tree grouped the 60 examined strains into 11 distinct clades similar to groupings of *Vibrionaceae* previously described (Fig. 1). The concatenated tree was in general congruent with the concatenated HK tree; however, it displayed shorter branch lengths, indicating that this gene is more highly conserved. In the concatenated tree, A total of 60 strains encompassing 37 different species were used to construct a concatenated phylogenetic tree of three housekeeping (HK) genes and a tree based on 16S rRNA (Fig. 1 and Fig. 2). Genes from chromosomes I and II were used in the generation of the concatenated tree and include RNA polymerase subunit beta (*rpoB*), malate dehydrogenase (*mdh*), both found on chromosome I, and dihydroorotase (*pyrC*) found on chromosome II. All trees were generated using the Neighbor-Joining method in MEGA 5.0 (6, 7). The Jukes-Cantor method was used to compute evolutionary distances (8). The bootstrap test was repeated 1,000 times and values over 50% are shown next to branches (9). The concatenated tree grouped the 60 examined strains into 11 distinct clades similar to groupings of *Vibrionaceae* previously described (Fig. 1). The 16S rRNA tree was in general congruent with the concatenated HK tree; however, it displayed shorter branch lengths, indicating that this gene is more highly conserved. In the 16S rRNA tree, minor differences were noted within the Campellii and Parahaemolyticus clades. *Vibrio* sp. AND4 and *Vibrio* sp. EJY3 do not group within these clades as they do in the concatenated HK tree. They did, however, branch off a shared node, although distinctly (Fig. 2).

### Table 1: Genome features of sequenced *Vibrio* species

<table>
<thead>
<tr>
<th>Species</th>
<th>Size (Mb)</th>
<th>GC (%)</th>
<th>Plasmid</th>
<th># of genes</th>
<th>tRNA genes</th>
<th>rRNA genes</th>
<th>Int</th>
<th>Tnp</th>
<th>Isolation source</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>V. harveyi</em> BAA-1116</td>
<td>6.1</td>
<td>45.4</td>
<td>1</td>
<td>6252</td>
<td>121</td>
<td>24</td>
<td>23</td>
<td>52</td>
<td>Marine ocean</td>
<td>Unknown</td>
</tr>
<tr>
<td><em>Vibrio</em> sp. EJY3</td>
<td>5.5</td>
<td>44.9</td>
<td>None</td>
<td>4935</td>
<td>121</td>
<td>28</td>
<td>20</td>
<td>22</td>
<td>West Sea coast</td>
<td>Korea</td>
</tr>
<tr>
<td><em>Vibrio</em> sp. Ex25</td>
<td>5.1</td>
<td>44.9</td>
<td>None</td>
<td>4676</td>
<td>124</td>
<td>34</td>
<td>10</td>
<td>0</td>
<td>Hydrothermal vent</td>
<td>East Pacific Rise</td>
</tr>
<tr>
<td><em>V. parahaemolyticus</em> RIMD2210633</td>
<td>5.1</td>
<td>45.4</td>
<td>None</td>
<td>5032</td>
<td>126</td>
<td>34</td>
<td>6</td>
<td>8</td>
<td>Homo sapiens</td>
<td>Thailand</td>
</tr>
<tr>
<td><em>V. splendidus</em> LGP32</td>
<td>5.0</td>
<td>43.9</td>
<td>None</td>
<td>4572</td>
<td>114</td>
<td>26</td>
<td>7</td>
<td>27</td>
<td>Hollow oyster</td>
<td>France</td>
</tr>
<tr>
<td><em>V. vulnificus</em> CMCP6</td>
<td>5.1</td>
<td>46.7</td>
<td>None</td>
<td>4665</td>
<td>111</td>
<td>28</td>
<td>7</td>
<td>44</td>
<td>Homo sapiens</td>
<td>South Korea</td>
</tr>
<tr>
<td><em>V. vulnificus</em> MO6-24/O</td>
<td>5.0</td>
<td>46.9</td>
<td>None</td>
<td>4701</td>
<td>111</td>
<td>28</td>
<td>6</td>
<td>19</td>
<td>Homo sapiens</td>
<td>USA</td>
</tr>
<tr>
<td><em>V. vulnificus</em> YJ016</td>
<td>5.3</td>
<td>46.7</td>
<td>1</td>
<td>5202</td>
<td>112</td>
<td>28</td>
<td>13</td>
<td>48</td>
<td>Homo sapiens</td>
<td>Taiwan</td>
</tr>
<tr>
<td><em>V. anguillarum</em> 775</td>
<td>4.1</td>
<td>44.5</td>
<td>None</td>
<td>3836</td>
<td>84</td>
<td>21</td>
<td>4</td>
<td>80</td>
<td>Coho salmon</td>
<td>Pacific Ocean</td>
</tr>
<tr>
<td><em>V. furnissii</em> NCTC 11218</td>
<td>4.9</td>
<td>50.7</td>
<td>4</td>
<td>4075</td>
<td>105</td>
<td>37</td>
<td>11</td>
<td>438</td>
<td>Atlantic cod</td>
<td>Norway</td>
</tr>
<tr>
<td><em>V. cholerae</em> IEC224</td>
<td>4.1</td>
<td>47.5</td>
<td>None</td>
<td>3787</td>
<td>98</td>
<td>25</td>
<td>11</td>
<td>28</td>
<td>Homo sapiens</td>
<td>Brazil</td>
</tr>
<tr>
<td><em>V. cholerae</em> LMA3894-4</td>
<td>3.7</td>
<td>48.0</td>
<td>None</td>
<td>3270</td>
<td>92</td>
<td>25</td>
<td>4</td>
<td>6</td>
<td>Tucunduaba creek</td>
<td>Brazil</td>
</tr>
<tr>
<td><em>V. cholerae</em> M66-2</td>
<td>3.9</td>
<td>47.6</td>
<td>None</td>
<td>3812</td>
<td>97</td>
<td>22</td>
<td>5</td>
<td>20</td>
<td>Homo sapiens</td>
<td>Indonesia</td>
</tr>
<tr>
<td><em>V. cholerae</em> MJ-1236</td>
<td>4.2</td>
<td>47.32</td>
<td>None</td>
<td>3894</td>
<td>98</td>
<td>22</td>
<td>13</td>
<td>19</td>
<td>Homo sapiens</td>
<td>Bangladesh</td>
</tr>
<tr>
<td><em>V. cholerae</em> N16961</td>
<td>4.0</td>
<td>47.5</td>
<td>None</td>
<td>3998</td>
<td>98</td>
<td>25</td>
<td>6</td>
<td>21</td>
<td>Homo sapiens</td>
<td>Bangladesh</td>
</tr>
<tr>
<td><em>V. cholerae</em> O395</td>
<td>4.1</td>
<td>47.5</td>
<td>None</td>
<td>4055</td>
<td>96</td>
<td>25</td>
<td>12</td>
<td>28</td>
<td>Homo sapiens</td>
<td>India</td>
</tr>
<tr>
<td><em>V. cholerae</em> 2010EL-1786</td>
<td>4.1</td>
<td>47.5</td>
<td>None</td>
<td>3946</td>
<td>79</td>
<td>22</td>
<td>10</td>
<td>26</td>
<td>Homo sapiens</td>
<td>Haiti</td>
</tr>
<tr>
<td><em>V. salmonicida</em> LF11238</td>
<td>4.7</td>
<td>39.0</td>
<td>4</td>
<td>4075</td>
<td>105</td>
<td>37</td>
<td>11</td>
<td>438</td>
<td>Atlantic cod</td>
<td>Norway</td>
</tr>
<tr>
<td><em>V. fischeri</em> ES114</td>
<td>4.3</td>
<td>38.4</td>
<td>1</td>
<td>3984</td>
<td>119</td>
<td>37</td>
<td>5</td>
<td>1</td>
<td>Squid light organ</td>
<td>Hawaii</td>
</tr>
<tr>
<td><em>V. fischeri</em> MJ11</td>
<td>4.5</td>
<td>38.2</td>
<td>1</td>
<td>4175</td>
<td>124</td>
<td>12</td>
<td>4</td>
<td>1</td>
<td>Squid light organ</td>
<td>Japan</td>
</tr>
<tr>
<td><em>P. profundum</em> SS9</td>
<td>6.4</td>
<td>41.7</td>
<td>1</td>
<td>5746</td>
<td>169</td>
<td>48</td>
<td>9</td>
<td>212</td>
<td>2500 meters</td>
<td>Sulu Trough</td>
</tr>
</tbody>
</table>
Additionally, in the Orientalis clade, *V. caribbenticus* and *V. mignipulchritudo* are grouped together in this clade but not in the HK gene tree (Fig. 1 and Fig. 2).

The Campbellii clade encompasses four species, *V. campbellii*, *V. harveyi*, *V. rotiferianus*, and *Vibrio* sp. AND4 (Fig. 1). Within this clade are two strains named *V. cholerae* HENC-03 and *V. cholerae* HENC-01, which are sequenced strains from Haiti (Fig. 1). Our analysis of additional HK loci from these strains and whole genome sequence analysis between HENC-03, HENC-01, and *V. harveyi* indicated that these strains are bona fide members of the Campbellii clade and should be renamed. Recently, it was shown, based on comparative genome hybridization and multilocus locus sequence analyses, that both *V. harveyi* strains ATCC BAA-116 and HY01 are *V. campbellii* isolates (10). In our sequence analysis, *V. harveyi* ATCC BAA-116 clustered closely with *V. campbellii* DS40M4. Based on HK and 16S rRNA data, HENC-03 appears to be a *V. harveyi* isolate and HENC-01 appears to be a *V. campbellii* isolate. For both the HK and 16S rRNA trees, the Parahaemolyticus clade branches with the Campbellii clade (Fig. 1 and Fig. 2). This clade contains four species, *V. parahaemolyticus* (3 strains), *V. algimolyticus*, which branched with *Vibrio* sp. Ex25, and *Vibrio* sp. EYJ3 which clustered within this group but was distinct from the other representatives (Fig. 1).

The Splendidus and Orientalis clades branched together on the HK tree. Three different species were found within the Splendidus clade, including *V. splendidus*, represented by 3 strains, with strain ATCC33789 divergent from the other two strains LPG32 and 12B01 examined. *Vibrio* sp. MED222 clustered closely with *V. splendidus* LGP32 suggesting that it is likely a *V. splendidus* isolate. *V. cyclitrophicus* also branched within this clade (Fig. 1). The Orientalis clade contained six highly divergent species that were not closely related to one another: *V. corallilyticus*, *V. parahaemolyticus* strain 16, *V. sinaloensis*, *V. brasiliensis*, *V. orientalis*, and *V. tubiashii*. These species were also present in the same clade in the 16S rRNA tree. The Orientalis clade represents at least six different groupings. *V. parahaemolyticus* strain 16 is misidentified and should be renamed *Vibrio* sp. 16, because it is unrelated to *V. parahaemolyticus* on both the HK and 16S rRNA trees.

The Scophthalmi clade contains three species based on both HK and 16S rRNA trees, *V. ichthyoienteri* which is distantly related to *V. scophthalmi* and *Vibrio* sp. N418, which are highly related to one another and are likely the same species (Fig. 1). The Anguillarum/Ordalii group contains *V. anguillarum* and *V. ordalii* that are both closely related to one another and also may represent a single species. In the 1985 paper by MacDonell and Colwell, based on the 5S rRNA sequence, *V. anguillarum* was placed in a new genus *Listonella* along with *V. damselae* and *V. pelagia*. Since then *L. damselae* has now been placed in the genus *Photobacterium* based on the 16S rRNA gene sequence. We retain the name *V. anguillarum* in agreement with Thompson and colleagues since this species is nested within *Vibrio* clades and is highly related to *V. ordalii* (2, 3, 4, 11). It was also demonstrated that *L. pelagia* is highly related to *V. splendidus* and clearly belongs in the genus *Vibrio* (2, 3, 4, 11). The Vulnificus clade consists of three *V. vulnificus* strains that are all highly related to one another suggesting that there is limited genetic diversity within this species. This clade does not cluster closely with any other clade examined in the HK gene tree. The Cholerae clade consists of four species: *V. cholerae* and *V. mimicus*, which are distinct species, and *Vibrio* sp. RC341 and *Vibrio* sp. RCS86, which cluster more closely with *V. mimicus* than *V. cholerae*. The Fischeri and Photobacterium clades formed divergent groups from the other *Vibrio* species clades in both HK and 16S rRNA trees (Fig. 1 and Fig. 2).

Overall, among the housekeeping genes examined from these strains, *rpoB* has the highest amount of conserved sites and *pyrC* (chromosome II) has the least (Table 2), congruent with the high variability among genes found on chromosome II. As expected, the 16S rRNA gene has the highest overall percentage of conserved sites (83.8%). The most variability was found in the Photobacterium clade for all 4 genes examined demonstrating that Photobacterium species are highly divergent from one another and may represent more than one genus (Table 2). The next most highly variable clade is Orientalis, which showed a high number of polymorphic sites among all four genes examined (Table 2). The most highly conserved clades include Anguillarum, Cholerae, Furnissii, and Vulnificus. When all clades are examined, the majority of the variability is represented by parsimony-informative sites. When each clade is examined separately, the majority of the variability is due to singletons, particularly in the *pyrC* gene in the following clades: Photobacterium, Orientalis, Fischerii, Scophthalmi, and Splendidus (Table 2).

In the following sections, we will use the concatenated HK tree to describe and discuss the distribution of different factors and traits within *Vibrio* species that has allowed them to occupy diverse niches. First, we describe the distribution of sigma factors and their role in global gene regulation in response to different environmental...
and host signals. Then, we discuss the *Vibrio*-specific
two-component regulator ToxRS and its role in gene
regulation among different species. The presence of
*Vibrio* in diverse niches and their interactions with eu-
karyotic hosts is driven by the need to exploit new food
sources. Thus, we will describe a subset of nutritional
traits that specific species have adopted that enable and
enhance host-*Vibrio* interactions. Last, we will describe
the role of horizontal gene transfer in the evolution of
*Vibrio* with specific attention to superintegron dis-
tribution and structure.

**DISTRIBUTION AND FUNCTIONALITY
OF SIGMA FACTORS THROUGHOUT
THE VIBRIONACEAE FAMILY**

Sigma factors are essential dissociable subunits of the
bacterial RNA polymerase (RNAP). In bacteria, RNAP is
not capable of initiating transcription or recognizing
specific promoters without a sigma factor forming a
holoenzyme. The sigma factor subunit has domains,
which enables it to recognize and direct the RNAP to
specific promoters. Once the holoenzyme is bound to
the promoter, transcription initiation can occur.

There are two major families of sigma factors that will
be discussed in this section: the sigma-70 family and the
sigma-54 family. The sigma-70 family encompasses the
largest group of sigma factors, and includes a primary sigma-70, responsible for transcription of housekeeping
genes, and several alternative sigma factors, which reg-
ulate transcription in response to various stimuli (12).
The sigma-54 family encompasses one sigma factor
designated RpoN in Gram-negative bacteria. RpoN is
functionally similar but structurally divergent from the
sigma-70 family.

Alternative sigma factors enable the cell to globally alter transcription in response to various stimuli, in-
cluding stress conditions that may occur within the bacterium’s natural environment. The repertoire of
alternative sigma factors available varies widely between
different species of *Vibrio*, including closely related
species, and may reflect the lifestyle of a particular spe-
cies (Table 3). For example, some *Vibrio* species possess
as few as 8 sigma factors while others possess as many as
12 sigmas. All *Vibrio* species contain a single copy of the
housekeeping sigma-70 factor, known as RpoD, and a
single copy of RpoH, which regulates the cytosolic heat
shock response. A phylogeny of *Vibrio* species based on
these proteins has an identical topology to that for the
HK tree, demonstrating that these are highly conserved
proteins and are ancestral to the group (data not shown).
All *Vibrio* species analyzed also contained a copy of the
*Escherichia coli* RpoS homologue, which in this
species is known as the stress response sigma factor. We
identified a number of *Vibrio* species that also encoded
an additional divergent copy of RpoS, two of which,
RpoX and RpoQ, are described in the literature (Fig. 3)
(13, 14). FliA, which regulates polar flagellum synthesis
or FliAP in *Vibrio*, is present in all *Vibrio* species. A
number of *Vibrio* species also possess a second flagella
sigma factor known as FliAL or LafS, which regulates
the synthesis of the lateral flagella in those organisms
(Table 3 and Fig. 3). The polar flagellar system is re-
quired for swimming motility in liquid media whereas
the lateral system is required for swarming motility in
viscous or solid media. The two flagellar systems have
been studied extensively in *V. parahaemolyticus* and
were shown to be distinct systems. Knockout deletions
of the polar flagellum system do not affect the lateral
flagella system and vice versa (15, 16, 17, 18).

The extracytoplasmic function (ECF) family of sigma
factors is highly divergent from the primary and primary
alternative sigma factors just described. This can be seen
visually in *Fig. 3*, with the ECF sigma factors forming
distinct divergent clades. The number and type of ECFs
varies even among closely related species with some
species possessing 3 ECF factors and others possessing
as many as 5. ECF factors typically respond to cell

---

**FIGURE 1** Evolutionary relationships of *Vibrio* species based the concatenated house-
keeping genes *rpoB, mdh*, and *pyrC*. The evolutionary history was inferred by using the
Neighbor-Joining method. The optimal tree with the sum of branch length = 2.24860938
is shown. The percentage of replicate trees in which the associated taxa clustered
together in the bootstrap test (1,000 replicates) is shown next to the branches. The tree
is drawn to scale, with branch lengths in the same units as those of the evolutionary
distances used to infer the phylogenetic tree. The evolutionary distances were computed
by using the Jukes-Cantor method and are in the units of the number of base substitutions
per site. The analysis involved 60 nucleotide sequences. Codon positions included were
1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair.
There were a total of 6,012 positions in the final data set. Evolutionary analyses were
conducted in MEGAS (120). doi:10.1128/microbiolspec.VE-0009-2014.f1
PHYLOGENY AND FUNCTION OF THE ALTERNATIVE SIGMA FACTOR RpoS

The role of RpoS has been characterized in a number of Vibrio species where it has been found to function in response to starvation stress conditions. In V. cholerae, it was shown that RpoS played a role in hydrogen peroxide, hyperosmolarity, and carbon starvation stress response, as well as in the production of HA/protease that processes the cholera toxin (21). It was also suggested that RpoS was not required for mouse intestinal survival (21). Another study examining in vivo colonization of V. cholerae showed that RpoS was required for colonization in a suckling CD1 mouse model, as well as played a role in enhanced resistance to hydrogen peroxide but not acid stress (22). It is important to note that these two in vivo studies utilized different strains as well as different growth conditions, which may account for the differences in the role of RpoS in colonization. More recently it was shown that RpoS is important for the mucosa escape response, because it was demonstrated that expression of motility and chemotaxis genes was upregulated by RpoS during this response in vivo (23). In the rpoS deletion mutant, levels of choler toxin were 10- to 100-fold higher. Taken together, these data suggested that RpoS represses virulence and stimulates motility to facilitate transmission (23).

In V. vulnificus, it has been shown that RpoS protects cells from acid stress, oxidative stress, and nutrient starvation and also that RpoS positively regulates extracellular enzymes such as albuminase, caseinase, and elastase, which may be required for survival under certain environmental conditions as well as for host adaptations (24, 25). In V. anguillarum, it has been shown that RpoS positively regulated expression of metalloproteases, and an rpoS deletion mutant showed reduced virulence in zebra fish (26).

In V. parahaemolyticus, it was demonstrated that RpoS had a limited involvement in acid stress and did not play a role in osmotic stress (27). However, RpoS plays a role in the viable nonculturable state for both V. vulnificus and V. parahaemolyticus because it was found to be constitutively expressed even below 15°C (28, 29). The effect on colonization of an rpoS deletion mutation in V. parahaemolyticus was examined in both a mouse model of colonization and in oyster colonization, and in both animal models RpoS was not required (30, 31, 32). Analysis of an rpoS deletion mutant in V. harveyi, found that this mutant was more sensitive to stationary phase stress and high concentrations of ethanol in comparison with the wild type, but was unaffected by high osmolarity or hydrogen peroxide stresses (10). In V. alginolyticus an rpoS mutant was more sensitive than wild type to ethanol, hyperosmolarity, and hydrogen peroxide (33).

Taken together, these studies demonstrated that RpoS can play a role in starvation, osmolarity, ethanol, hydrogen peroxide, and acid stress responses depending on the species. These data suggest that RpoS has evolved different roles among different Vibrio species, and the

**FIGURE 2** Evolutionary relationships of taxa based on 16S rDNA sequence. The evolutionary history was inferred by using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.47043464 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches (30). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed by using the Jukes-Cantor method and are in the units of the number of base substitutions per site (52). The analysis involved 58 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 1,564 positions in the final data set. Evolutionary analyses were conducted in MEGA5 (120). doi:10.1128/microbiolspec.VE-0009-2014.f2
specific role of RpoS in each potentially has evolved to allow changes in the lifestyles of the species. Consistent with this are the several studies that have shown variation in the RpoS regulon among these species (21, 22, 23, 34).

As shown in the phylogenetic analysis in Fig. 3 and Table 3, a number of the Vibrio species possess additional RpoS-like sigma factors. This divergence from the canonical RpoS suggests that it may also have diverged in function. In our analysis of the available Vibrio sequenced genomes, we identified five RpoS-like copies that were divergent from each other and RpoS (Fig. 3). RpoX was originally identified in V. alginolyticus and was shown to be involved in biofilm formation and stress response (13). We found that the distribution of RpoX is confined to a handful of species: all V. alginolyticus and V. splendidus (Fig. 3). We identified a homologue of RpoX in Vibrio sp. MED222, Vibrio spp. HENC-01, HENC-02, and HENC-03, which are not shown in Fig. 3. RpoQ was characterized in V. fischeri, and has 45% amino acid identity to the RpoS protein in this species (14). Overexpression of RpoQ in V. fischeri led to increased chitinase activity but decreased motility and luminescence (14). It was shown that RpoQ was controlled by LuxO via LitR, a LuxR homologue (14). We found that RpoQ is confined to V. fischeri, V. salmonicida, and V. shilonii isolates. We identified a third divergent RpoS-like sequence that was present in V. harveyi 1DA3 (locus tag Vh03010) and has not been characterized to date. We identified closely related homologues of Vh03010 in V. alginolyticus, V. splendidus, Vibrio sp. MED222, Vibrio spp. HENC-01, HENC-02, and HENC-03, V. cyclitropicus, and V. shilonii isolates. As stated earlier, all Vibrio species contain a copy of the canonical RpoS. We found three species that had three RpoS-like proteins, V. shilonii, P. damsela, and Photobacterium sp. SKA3, and one species that had 4 RpoS-like proteins, V. angustum. These RpoS-like sigma factors have diverged significantly from RpoS, suggesting that they may have evolved functions distinct from that of RpoS.

### DISTRIBUTION AND FUNCTION OF RpoH, FliA, AND FIIAL

Although all Vibrio species contain a single E. coli RpoH homologue, this sigma factor has remained relatively uncharacterized across the Vibrio species. The V. cholerae RpoH can functionally complement an E. coli rpoH mutant, suggesting that it has evolved a similar role in both species (35). A global gene expression and phenotypic analysis of a V. cholerae rpoH deletion mutant by Slamti and colleagues suggested that, in this species, rpoH promotes growth at temperatures ranging from 15 to 42°C (36). The rpoH deletion mutant could only be made in the presence of a plasmid-borne inducible copy of rpoH suggesting that this is an essential gene. Not surprisingly, the rpoH mutant was severely attenuated in a suckling mouse model of colonization (36).
The flagellar sigma factors FliAP and FliAL have been best characterized in *V. parahaemolyticus* (15, 16, 17). A study in 1993 by McCarter and coworkers identified the genes involved in the swarming flagellar motor of *V. parahaemolyticus* using transposon mutagenesis (16). They identified sigma-28 (FliAL) that is required for the synthesis of lateral flagella (16). In *V. parahaemolyticus*, FliAP has been shown to be important in the development of the single polar flagellum required for swimming motility, and three of four flagellins are dependent on FliAP. The other flagellin expression is dependent on RpoN (15). Both fliAP and fliAL are regulated by RpoN, which is discussed below, and an *rpoN* mutant is nonflagellated and defective in swimming and swimming motility in *V. parahaemolyticus* (18). Both the FliAP and FliAL sigma factors are present in all members of the Campelli and Parahaemolyticus clades, indicating that the retention of two flagella systems is conserved in these clades. The presence of both flagella regulators has an unusual distribution. The two regulators are present in all members of the Campelli, Parahaemolyticus, and Furnissii clades and in all *V. mimicus* isolates, as well as in *V. corallilyticus*, *V. brasilensis*, *V. caribbenticus*, and several *Photobacterium* species. These data suggest that the two flagellar systems are ancestral to the Campelli, Parahaemolyticus, and Cholerae clades and lost from *V. cholerae* isolates.

**DISTRIBUTION AND FUNCTION OF THE SIGMA-54 FAMILY SIGMA FACTOR RpoN**

RpoN, the only member of the sigma-54 family, has been characterized in *V. alginolyticus*, *V. anguillarum*, *V. cholerae*, *V. harveyi*, *V. fischeri*, and *V. parahaemolyticus* (18, 37, 38, 39, 40, 41, 42, 43). In *V. cholerae* RpoN has been shown to be required for motility, nitrogen metabolism, biofilm formation, quorum sensing, type 6 secretion system (T6SS) synthesis, and virulence (41, 42, 44, 45, 46, 47, 48). RpoN is required for colonization in an infant mouse model of cholera, and the defect was not entirely related to the lack of motility or glutamine synthetase expression (44). Klose and colleagues recently determined that the *V. cholerae* RpoN regulon includes more than 500 genes in serogroup O1 pathogenic isolates (48). They showed that RpoN played a role in the regulation of flagella synthesis, ammonium assimilation, virulence factor synthesis, and dicarboxylic acid metabolism (48). A study published in 2012 showed that RpoN differentially regulates the type IV secretion system and flagella operons in *V. cholerae* serogroup O37 strain V52 (42). They identified 68 RpoN binding sites and 82 operons that were positively controlled by RpoN. The 82 operons identified included genes for motility, all T6SS *hcp* 1 and *hcp* 2 operons, nitrogen metabolism, quorum-sensing regulator HapR (LuxR homologue), formate dehydrogenase, phage shock, and over 40 genes of unknown function indicating the significant contributions of RpoN to cellular processes in *V. cholerae* (42). In *V. fischeri*, RpoN was shown to play a role in motility, biofilm formation, luminescence, and colonization of the squid host (40, 49). The *rpoN* mutant examined in these studies was nonmotile, produced less biofilm, grew poorly on minimal media, and could not colonize its symbiotic squid host. The *rpoN* mutant also had 3- to 4-fold higher levels of bioluminescence, suggesting that RpoN normally represses bioluminescence (40, 49). In *V. anguillarum*, it was shown that RpoN is essential for flagellum production as well as virulence in fish. An *rpoN* mutant exhibited loss of motility, but no loss of virulence when fish were intraperitoneally injected; however, virulence was significantly reduced when fish were orally infected.

**TABLE 3** Number of primary and alternative sigma factors present among *Vibrio* species

<table>
<thead>
<tr>
<th>Species</th>
<th>Total No. Sigma Factors</th>
<th>Type of Sigma Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Primary</td>
<td>Alternative</td>
</tr>
<tr>
<td><em>V. alginolyticus</em></td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td><em>V. parahaemolyticus</em></td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td><em>V. harveyi</em></td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td><em>V. vulnificus</em></td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td><em>V. anguillarum</em></td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td><em>V. cholerae</em></td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td><em>V. furnissii</em></td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td><em>V. splendidus</em></td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td><em>V. salmonicida</em></td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td><em>V. fischeri</em></td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>
FIGURE 3 Phylogenetic tree constructed from the alignment of sigma-70 family sigma factors by using the amino acid sequences of the highly conserved domains 2 and 4. The program MEGA was used to construct a neighbor-joining tree using the Poisson model, complete deletion, and a bootstrap value of 1,000 (120). Abbreviations are: Ec, E. coli; Vp, V. parahaemolyticus; Vc, V. cholerae; Vv, V. vulnificus; Vh, V. harveyi; Vf, V. fischeri; Va, V. alginolyticus; Vs, V. splendidus; Vaa, V. anguillarum; Vfu, V. furnissii.

Phylogenetic analysis shows that there are two main subfamilies of sigma factors, primary and extracytoplasmic function (ECF) type. Phylogenetic analysis also shows that predicted RpoD, RpoF, RpoS, RpoH, and RpoE sigma factors encoded by each Vibrio species analyzed do cluster with and are homologues to those found in E. coli. The analysis also shows additional primary-like alternative sigma factors found in five of the species that are closely related to RpoS. The analysis also demonstrates that there are multiple putative ECF factors found within the Vibrio species analyzed, forming distinct clades.

In *V. anguillarum*, RpoN is also required for protease secretion, exopolysaccharide production, and biofilm formation (43). RpoN is required for the formation of the polar flagellum synthesis and motility in *V. alginolyticus* (37).

RpoN was shown to play a role in quorum sensing in *V. harveyi* via the sigma-54 activator LuxO. When LuxO activates RpoN to allow formation of a functional holoenzyme, this leads to expression of LuxR, the regulator of luminescence and other phenotypes (39). This was also shown to be the case in *V. cholerae, V. alginolyticus*, and *V. parahaemolyticus*. In *V. alginolyticus* RpoN regulate hcp1, a hallmark of T6SS as well as other T6SS genes and motility via VasH, a sigma-54 activator protein (42, 50). As mentioned previously, in *V. cholerae*, T6SS gene clusters are also under the control of RpoN via VasH, and in *V. parahaemolyticus* RpoN is also required for hcp1 expression, which is repressed by OpaR (LuxR), the quorum-sensing output regulator in *V. parahaemolyticus* that also represses the T3SS1 cluster (41, 42, 50, 51, 52).

Our recent studies of the function of RpoN in *V. parahaemolyticus* RIMD2210633, a clinical serogroup O3:K6 isolate, have uncovered a novel role for this sigma (18). We constructed an in-frame deletion of rpoN (VP2670) and examined the effects *in vivo* using our newly developed streptomycin-treated adult mouse model of colonization (18, 31). We showed that deletion of rpoN rendered *V. parahaemolyticus* aflagellar and therefore nonmotile, and also resulted in reduced biofilm formation and a defect in glutamine synthetase production. We performed *in vivo* competition assays between the rpoN mutant and a wild type marked with the beta-galactosidase gene lacZ (strain WBWlacZ) to screen for white and blue colonies. We demonstrated that the rpoN mutant colonized significantly more proficiently than WBWlacZ.

Interestingly, when we constructed mutants defective in the polar flagellum biosynthesis by knocking out the fltIAP sigma factor and the lateral flagellar synthesis system by deletion of fltIAL sigma, these mutants also outcompeted WBWlacZ, but not to the same level as the rpoN mutant. These data suggested that lack of motility is not the sole cause of the fitness effect. Previous work in *E. coli* also showed that motility mutants were better *in vivo* colonizers, and this work indicated that the motility mutants were metabolically more fit (53). Therefore, we determined whether there were differences in growth of the rpoN mutant and the wild-type strain (18). In an *in vitro* growth competition assay in mouse intestinal mucus, the rpoN mutant also outcompeted the wild type and exhibited faster doubling times when grown in mucus and on individual components of mucus. We determined the doubling times of the mutant and wild type in mucus sugars and showed that the mutant had faster doubling times. We then examined the expression pattern of genes in the pathways for the catabolism of mucus sugars and showed that these genes had significantly higher expression levels in ΔrpoN than in wild type (18).

In summary, it appears that the RpoN regulon is vast and complex within the *Vibrio* genus. Common themes are found such as a requirement for RpoN for flagellation, motility, biofilm formation, and regulation of T3SS, T6SS, and the quorum-sensing regulator LuxR. Differences arise in the role of RpoN in the virulence of a particular species, which probably reflects the different requirements for colonization and virulence among the species and the animal models used rather than differences in RpoN regulation per se. Overall, studies indicate that, similar to other genera, sigma factors are an important component of global transcriptional regulation in *Vibrio* with much variability seen in the number of alternative sigma factors. This variation may be reflective of the different lifestyles of each species and may suggest different requirements for gene transcription regulation in response to changing cellular and extracellular conditions.

**PHYLOGENY AND FUNCTION OF ECF-TYPE SIGMA FACTORS**

The most well-studied ECF-type sigma factor is RpoE, and it is required for cell envelope stress response. RpoE has been characterized in studies in *V. vulnificus*, *V. harveyi*, *V. angustum*, and *V. parahaemolyticus* and multiple studies in *V. cholerae* (54, 55, 56, 59). We identified a close homologue of the *E. coli* RpoE in all sequenced *Vibrio* indicating this ECF is conserved (Fig. 3). In *V. cholerae*, RpoE was shown to play an important role in intestinal survival and virulence in an infant mouse model (55). In this first study, an rpoE mutant was created in O395, a biotype classical *V. cholerae* strain, and was highly attenuated for virulence and sensitive to ethanol stress. The rpoE mutant, similar to wild type, was not defective against heat, bile salts, hydrogen peroxide, polymyxin B, osmolarity, and pH (5.5 to 10) stresses (55). In contrast, a study by Waldor’s group, examining an rpoE mutant in a biotype El Tor background strain, showed that this mutant was sensitive to both a bioactive peptide P2 and polymyxin B antimicrobial peptides (56). They determined that the
outer membrane protein OmpU was a key determinant of basal rpoE expression and proposed that misfolded OmpU protein in the periplasm may be the signal that allows release of RpoE tethered to the cell membrane (56). With the use of a next-generation high-throughput sequencing, they found that most rpoE mutants could be constructed only in the presence of suppressor mutations, suggesting that rpoE is an essential gene in this V. cholerae El Tor strain (58). The authors found that, of all of the independently constructed rpoE mutants, 75% had a suppressor mutation in the promoter region of ompU (58).

In V. vulnificus, an rpoE mutant was more sensitive to membrane-perturbing agents such as ethanol and SDS, as well as hydrogen peroxide, and there was a slight increase in the sensitivity to heat and cold shock, but this mutant was not attenuated for virulence in mice (57). It was suggested that rpoE is essential in V. harveyi because an rpoE mutant could not be constructed (59). Instead, mutants were constructed in the rpoE regulatory region rseABC. It was demonstrated that the overexpression of rpoE resulted in a reduction of hemolytic activity and attenuation for colonization in shrimp (59).

In V. parahaemolyticus, we constructed an in-frame deletion of the E. coli rpoE homologue from V. parahaemolyticus RIMD2210633 (32). We found that rpoE is essential for survival during cell envelope stress conditions such as polymyxin B, hydrogen peroxide, and ethanol stresses but was not required for acid, bile salt, and SDS stress responses. Our in vivo colonization analysis in a streptomycin-treated adult mouse model of colonization demonstrated that the rpoE mutant was severely defective compared with the wild-type strain (32). These data suggest that RpoE is required for in vivo colonization probably through its role in cell envelope stress tolerance. Unlike in V. cholerae, it appears that, in V. parahaemolyticus, OmpU may not be a signal for RpoE release from the membrane since our ompU deletion mutant strain was resistant to polymyxin B and ethanol stress unlike the rpoE mutant, which is sensitive to both these stresses (32).

We identified several divergent ECF-type sigmas among Vibrio species, and these along with flIAL were variably present among different clades. These divergent ECFs may not be required for extractoplasmic cell envelope stress but may have evolved for a range of functions. V. parahaemolyticus contained four additional divergent ECF-type sigmas: VP0055, VP2210, VP2358, and VPA1690. VP2210 was present in all sequenced Vibrio species with the exception of V. salmonicida and its function remains unknown. In a number of Vibrio species this gene is embedded among fatty acid oxidation and transport genes, which may suggest a possible role for this regulator. VP2358 is also widely distributed, present in most species with the exception of V. fischeri, V. salmonicida, and Photobacterium species (Fig. 3). The function of VP2358 is also unknown and the genes that surround this sigma include a putative antisigma factor that has a cupin domain and may be involved in sensing reactive oxygen species. Other genes surrounding this sigma are involved in DNA damage repair and amino acid biosynthesis. Homologues of VP0055 are present in all members of the Campellii and Parahaemolyticus clades as well as in V. vulnificus, V. fischeri, V. salmonicida, Vibrio sp. 16, V. sinolensis, V. furnissii, and Photobacterium AK15. The function of VP0055 is unknown; however, in all Vibrio species that possess this sigma, the genes surrounding it are genes involved in gluconate metabolism. VPA1690 is only present in V. parahaemolyticus, V. algilnolyticus, and Vibrio sp. Ex25 strains. The genes immediately adjacent to this sigma are conserved hypothetical proteins and transcriptional regulators, as well as genes involved in alkyl hydroperoxide reduction, which may suggest a possible function.

**VIBRIO-SPECIFIC GLOBAL REGULATOR ToxRS**

Apart from sigma factors, bacteria also utilize two-component regulators to help control gene transcription in response to changing environmental conditions. Among Vibrio species, the number of two-component systems varies from species to species. The ToxRS system is a two-component regulatory system present in all Vibrio and has been characterized extensively in V. cholerae and to a much lesser extent in V. parahaemolyticus (27, 30, 31, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73). We reconstructed the evolutionary history of ToxR for the same set of isolates that we examined for the HK and 16S rRNA gene trees (Fig. 4). We obtained a ToxR phylogeny that had a topology similar to HK and 16S rRNA trees for all the major clades described, indicating that this regulatory system is ancestral to the family Vibrionaceae (Fig. 4).

In V. cholerae the ToxRS regulon can be separated into two branches, the ToxT-dependent and -independent pathways (73, 74, 75, 76). ToxR has been shown to act as a coactivator of the toxT gene, which encodes a transcription factor that positively regulates the expression of the genes for cholera toxin and the toxin
coregulated pilus (TCP), which are essential virulence and colonization factors, respectively (60, 61, 66, 73). ToxR regulation of the outer membrane porin OmpU has been shown to be important for survival in the face of organic acid stress, bile stress, and antimicrobial peptides in this species (69, 70, 72). As discussed above, there is evidence that indicates that OmpU in V. cholerae may act as a signal in the activation of RpoE, by an as yet unknown mechanism (56, 58). The role of OmpU in in vivo survival is not entirely known, because V. cholerae ompU deletion mutants only exhibit a slight decrease in colonization in the infant mouse model of colonization (56, 58). In V. splendidus, OmpU was shown to be required for antimicrobial peptide resistance and to be an essential component for colonization of the oyster host (78, 79).

Unlike V. cholerae, the V. parahaemolyticus ToxRS regulon does not contain the ToxT pathway and its regulon. Klose and coworkers demonstrated that a toxR mutant strain had an altered outer membrane profile, specifically OmpU (70). Our group demonstrated that ToxRS positively regulates the V. parahaemolyticus OmpU homologue and is sensitive to acid and bile salt stresses (27, 31). We showed that under acid conditions toxRS and ompU were highly expressed. In a toxRS mutant strain, ompU expression was severely reduced and cells were sensitive to both acid and bile salt stresses. The toxRS mutant compared with the wild-type strain by the use of a pretreated streptomycin adult mouse model exhibited a severe defect in intestinal colonization (27, 31). In addition, we demonstrated that the V. parahaemolyticus ToxRS system was a negative regulator of the T3SS-1 contained in chromosome I. This regulation by ToxRS was indirect and was mediated by the LeuO transcription factor homologue encoded by vp0350. ToxRS was shown to be a positive regulator of leuO expression, and LeuO is a negative regulator of T3SS-1 gene expression (31, 80). This ToxRS pathway is thus far unique to V. parahaemolyticus (31).

CARBON UTILIZATION: SUGARS AND AMINO SUGARS

The adaptability to new nutritional sources is crucial for a species to switch between macro- and microenvironments. Survival in the nutrient-limited conditions of the water column and marine sediments requires that an organism efficiently utilizes the available nutrients. Adaptation to host environments also requires the ability to switch from nutrient-poor to nutrient-rich and from free-living to host-associated conditions and also requires that a species can compete with the resident microbiota for nutrients. In the next section, we will discuss some of the major nutrient capabilities of Vibrio species and how generalization and specialization within and between species may have aided niche expansion.

Chitin is a very abundant substance found in the marine environment, with an estimated 10 x 10^{11} tons produced annually. Chitin is composed primarily of a polymer of β-1,4 N-acetylgalucosamine (GlcNAc), and Vibrio are key to the degradation of chitin, which contributes significantly to nitrogen and carbon cycling in the marine ecosystem (14, 48, 52, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90). Chitin provides a ready-made source of GlcNAc, which is an important metabolite and source of energy for all Vibrio species. The ability to use chitin as a carbon and nitrogen source is universally found within Vibrio species. It has been shown in V. furnissii that there is a strong chemotactic response to GlcNAc oligomers, as well as the ability to adhere to chitin particles through GlcNAc-lectin association (81, 82, 83). Chitin comes in many varying forms in the wild, and Vibrio species have been shown to produce several chitinases to cope with this diversity. For example, V. harveyi has six chitinase genes, and, among Vibrio species in general, the number can range from 2 to 8. Extracellular chitinases break down the large polymers into smaller subunits of GlcNAc. The oligosaccharides are transported by a specific chitin porin (91). Chitin degradation appears to be tightly controlled in the presence of other energy sources, as seen in V. furnissii, where considerable chitinase release was only found in the absence of any other carbon source (86, 88, 91, 92, 93, 94, 95). In Vibrio species, chitin degradation is controlled by a chitin sensor (ChiS) kinase (87, 96).

Many Vibrio species also interact with a large number of eukaryotic hosts where chitin is absent. Many of these commensal and pathogenic host interactions involve bacterial colonization of the mucus-covered epithelium surfaces. For example, the mammalian gastrointestinal epithelium contains a mucus gel layer that is the site of attachment of some species. The main constituent of mucus is mucin, a glycoprotein composed of oligosaccharides that include ribose, mannose, hexuronates (glucuronate and glucuronate), galactose, fucose, arabinose, GlcNAc, N-acetylgalactosamine (GalNAc), and N-acetylneuraminic acid (Neu5Ac or sialic acid). Therefore, mucin constitutes a direct carbohydrate source that can offer ecological advantages to Vibrio gastrointestinal species especially if they can utilize a unique carbon source or one more efficiently than the resident commensal species. We examined the genome of sequenced
Vibrio species and found that the genes for the metabolism of mucin sugars are widely present. For example, we found that all Vibrio species contain the genes required for the metabolism (and transport) of ribose, mannose, gluconate, galactose, GlcNAc and GalNAc.

In the case of the catabolism of glucuronate, arabinose, fucose, and sialic acid, a wide variation in the presence of these genes was found. While species from 8 of the major clades identified in Fig. 1 carry genes for glucuronate catabolism, these genes were missing from all representatives in the Cholerae, Fischeri, and Orientalis clades. These data suggest that the ability to utilize glucuronate as a nutrient source is ancestral to the family Vibrionaceae and was likely lost from V. cholerae, V. fischeri, and V. orientalis. The genes for arabinose catabolism are only present in two species, V. parahaemolyticus and V. furnissii, which suggests that this is may be a recently acquired trait in the family Vibrionaceae. In V. parahaemolyticus the arabinose catabolism gene cluster is present in a 25-kb region encompassing VPA1652 to VPA1679 that also contains a ferric uptake system (111). These genes are present in chromosome II and were missing from all representatives in the Cholerae, Fischeri, and Orientalis clades. These data suggest that the ability to utilize glucuronate as a nutrient source is ancestral to the family Vibrionaceae and was likely lost from V. cholerae, V. fischeri, and V. orientalis. The genes for arabinose catabolism are only present in two species, V. parahaemolyticus and V. furnissii, which suggests that this is may be a recently acquired trait in the family Vibrionaceae. In V. parahaemolyticus the arabinose catabolism gene cluster is present in a 25-kb region encompassing VPA1652 to VPA1679 that also contains a ferric uptake system (111). This region was identified by genome comparisons with V. cholerae and V. vulnificus which lack this region (97). The ability to utilize arabinose is not present in all strains of this species; for example, the environmental strain UCM-V493 does not contain the arabinose gene cluster (98).

The genes for fucose catabolism were present in only V. vulnificus, V. mimicus, V. nigripulchritudo, V. shilonii, V. harveyi HENC-01, and Vibrio sp. EJY3. Our examination of the fucA gene from this pathway shows that its closest homologue is present in enteric species. In V. vulnificus YJ016, the fucose catabolism gene cluster is present within a 117-kb island. The genomic island is inserted at a tRNA serine locus, contains an integrase, and has an overall GC content of 43% (99). The island region encompasses VV2149 to VV2262 in YJ016 and is largely absent in MO6-24/O and CMCP6, but the fucose catabolism cluster is present in MO6-24/O. Taken together, these data strongly suggest that this is a newly acquired trait in these species and not ancestral to the group.

N-Acetylmuramoylalanine, commonly known as sialic acid, is an amino sugar utilized in humans and is ubiquitous among metazoans of the deuterostome lineage (100, 101, 102). Sialic acid is an integral component of mucin. Genes involved in the catabolism of sialic acid are present in less than 25% of sequenced Vibrio genomes (103, 104). These genes are present in five clades of Vibrionaceae: Photobacterium, Fischeri, Cholerae, Vulnificus, and Orientalis clades. Within the Cholerae clade the ability to uptake and catabolize sialic acid is broadly distributed, present in many isolates of both V. cholerae and V. mimicus (103, 104, 105, 106). In these species, the genes are present within a pathogenicity island (PAI) named Vibrio pathogenicity island-2 (VPI-2), which is found only in pathogenic isolates (105, 106). The VPI-2 region has all the hallmarks of a PAI; it has a GC content lower than the rest of the genome and encodes an integrase that is adjacent to a tRNA serine locus (105, 106). In V. cholerae non-O1/non-O139 isolates and V. mimicus isolates that cause inflammatory diarrhea, a T3SS is present on the VPI-2 island (107, 108, 109, 110). Within V. vulnificus, the ability to uptake and catabolize sialic acid is present in clinical-type isolates but absent from environmental isolates (111). These genes are present in chromosome II and are not associated with a horizontally acquired region (111). In the Orientalis clade, three species, V. orientalis, Vibrio sp. 16, and V. sinaloensis, contain the genes for sialic acid catabolism. These genes are also present in V. ichthyoenteri, V. nigripulchritudo, V. shilonii, and Vibrio sp. MED222.

The ability to catabolize sialic acid as a sole carbon and energy source has been shown to be a significant in vivo fitness factor (111, 112, 113, 114). We demonstrated that deletion of the nanA gene in V. cholerae, which encodes the first enzyme in the sialic acid catabolic pathway, resulted in decreased colonization in comparison with wild type in an infant mouse model of cholera (112). It was also demonstrated in V. vulnificus that sialic acid catabolism is important for mouse intestinal colonization (115). These data indicate that the ability to utilize sialic acid, an abundant carbon and nitrogen source in the human gut, is important in in vivo survival. A tripartite ATP-independent periplasmic (TRAP) transporter, SiaPQM (vc1777-vc1779), is adjacent to the sialic acid catabolism genes contained within the VPI-2 in V. cholerae pathogenic isolates (103, 104, 106, 109, 112). Thomas and colleagues have clearly demonstrated that SiaPQM functions biochemically as a sialic acid transporter in vitro (114, 116). We showed a strong
correlation between the presence of VPI-2 and the ability of *Vibrio* to grow on sialic acid as a sole carbon source (113). We demonstrated that a strain of *V. cholerae* N16961 defective in *vc1777*, encoding the large membrane component of the TRAP transporter, does not grow on sialic acid (113). Interestingly, a report by Sharma and colleagues proposed that an entirely different TRAP transporter, DctPQM encoded by *vc1927-vc1929*, was the sole sialic acid transporter in *V. cholerae* (117). However, our bioinformatic, genomic analysis, and biochemical and genetic analyses determined that DctPQM is a C₄-dicarboxylate-specific TRAP transporter and disruption of *vc1929* results in a defect in growth on C₄-dicarboxylates but not sialic acid (113, 114). These data demonstrate that SiaPQM, not DctPQM, is the sole sialic acid transporter in *V. cholerae* (113, 114).

**HORIZONTAL GENE TRANSFER**

Horizontal gene transfer has played an important role in the emergence of different *Vibrio* species by allowing strains and species to expand their niche. In this next section, we will provide an overview of the major HGT events that have occurred in the *Vibrio* species whose complete genome sequence is available. Specifically, we will give a detailed account of the genetic diversity and phylogeny of superintegrons within these sequenced strains.

*V. cholerae* is the most widely studied *Vibrionaceae* species because of its current and historical impact on human health. Of the seven completely sequenced strains of *V. cholerae*, six were isolated from humans, N16961 (Bangladesh), O395 (India), IEC224 (Brazil), M66-2, MJ-1236 (Bangladesh), and 2010EL-1786 (Haiti), and one, *V. cholerae* LMA3894-4 (Brazil), was acquired from creek water (Table 1). Within the genome of cholerigenic isolates are three regions acquired by horizontal gene transfer that are essential for virulence (118). Cholera toxin, the cause of the secretory diarrhea, is encoded within a filamentous phage named CTXphi. The intestinal colonization factor, the type IV Toxin coregulated pilus (TCP), is encoded within a pathogenicity island named VPI-1 or the TCP island. Sialidase and sialic acid uptake and catabolism genes are present on a nonhomologous island named VPI-2 or the sialic acid metabolism island (62, 105, 106, 112, 119, 120). The first *V. cholerae* genome to be sequenced was *V. cholerae* N16961 in 2000, an El Tor biotype strain responsible for the ongoing seventh cholera pandemic (121). Two sixth pandemic biotype classical strains O395 and M66-2 are also available. The most striking difference between the classical and El Tor strains is the presence of two putative pathogenicity islands: *Vibrio* seventh pandemic island I (VPI-I) and VSP-II in N16961. VSP-I is proposed to be important for host adaptation through the production of a regulatory cyclic dinucleotide (122). The CTXΦ is absent from M66-2, whereas the classical O395 strain contains an extra copy of the phage in chromosome II, identical to that found in chromosome I (123). Additionally, the *rstR* genes of the CTXΦ in strains N16961 and O395 differ significantly from each other (123). *V. cholerae* MJ-1236 is described as a hybrid strain. It contains a 33-kb excisable element Kappa phage and a 133-kb ICE-like element, both found in chromosome I and absent from N16961. Additionally, genomic island 14 (GI-14), a 19-kb island, is also a common link between MJ-1236 and the classical strain. On the large chromosome, MJ-1236 contains both VPI-1 and VPI-2 as well as VSP-I and VSP-II, but a second copy of VSP-I can be found on the smaller chromosome (124). The core genomic region of the CTX φ phage contains *ctxAB* of the classical strains as well as the El Tor “pre-CTX” region. There is even more variability in the RS2 region of MJ-1236 because it contains a classical *rstR* gene, an El Tor *rstB* gene, and a *rstA* gene that differs from both biotypes.

The seventh cholera pandemic hit Latin America in 1991 and, in 1994, a sucrose-fermenting defective strain, *V. cholerae* ICE224, which caused outbreaks until 1996, was isolated and eventually sequenced (125). When compared with the N16961 genomes, single nucleotide polymorphisms (SNPs) were seen in 122 genes and other Latin American strains shared 48 of these SNPs. Additionally, distinguishing it from N16961, there is a 49-kb insertion of the WASA1 bacteriophage at an attachment site in the membrane alanine aminopeptidase gene (125). The strain 2010EL-1786 was isolated from Haiti in 2010 and determined to be of the El Tor biotype, but it contained the *ctxB* gene of the classical biotype. This isolate, as well as all isolates from Haiti, contained a tcpA allele of a Bangladeshi El Tor strain (CIRS 101) that carries an El Tor CTXΦ, but produces the classical cholera toxin (126).

Two other important *Vibrio* human pathogens are *V. vulnificus*, which causes wound infection and septicemia in immune-compromised patients, and *V. parahaemolyticus*, which causes gastroenteritis via consumption of raw infected seafood (127, 128, 129). *V. vulnificus* is an opportunistic pathogen that can have a greater than 50% mortality rate in susceptible hosts (130). Three complete genome sequences are available and all are
clinical isolates. *V. vulnificus* YJ016 contains a plasmid, pYJ016, not found in the other two *V. vulnificus* strains that are sequenced, CMCP6 and MO6-24/O (131). Comparative genome analysis of *Y. pseudotuberculosis* RIMD2210633 with *V. cholerae* N16961 (131). Comparative genome analysis of YJ016 and CMCP6 identified 14 regions ranging in size from 14 to 117 kb, which had the characteristics of recently horizontally acquired DNA (99). These 14 genomic islands each contain an integrase gene and had aberrant GC content and dinucleotide frequency (99). Most of the genomic islands were absent from a collection of *V. vulnificus* isolates, suggesting that, in this species, each strain may contain its own repertoire of horizontally acquired DNA (99). We identified a genome region among clinical isolates named region XII that contained genes required for the transport and catabolism of glycoaminoglycans, a potential source of carbon and nitrogen in *vivo*. Two metabolic pathways were encoded within this region, lacto-N-biose and chondroitin sulfate, as well as putative transporters for each of these substrates (Fig. 5). Thus, this region may potentially be involved in host interactions that allow it to survive in *vivo* (132). The complete region XII metabolic island is also present in *Vibrio* sp. RC341, five *V. mimicus* isolates, and *V. nigripulchritudo* (Fig. 5).

We performed systematic BLAST analysis for each of the open reading frames (ORFs) of *V. parahaemolyticus* RIMD2210633 compared with each of the ORFs from the genome sequences of *V. cholerae* N16961, *V. vulnificus* YJ016 and CMCP6, and *V. fischeri* ES114 (97). We identified 24 regions, gaps in the genome atlas, of greater than 10 kb that were unique to RIMD2210633. These regions included an integron, f237 phage, two T3SSs, a T6SS, and 7 *V. parahaemolyticus* genomic islands named VPaI-1 to VPaI-7 (97). Comparative genomic analysis of RIMD2210633 with *V. parahaemolyticus* AQ3810, an O3:K6 isolate recovered in 1983, identified five regions unique to RIMD2210633: VPaI-1, VPaI-2, VPaI-4, VPaI-5, and VPaI-7 (97, 133). RIMD2210633 encoded an additional T6SS as well as an additional T3SS encoded on VPaI-7. Our data showed that there was considerable genomic flux and that the highly virulent clone of RIMD2210633 arose from an O3:K6 isolate that acquired at least seven novel regions (97, 133). We recently sequenced an environmental isolate of *V. parahaemolyticus* strain UCM_V493. This strain lacked all seven islands including the T3SS on chromosome II and the T6SS on chromosome I (78).

*V. splendidus* is a species associated with invertebrate mortalities worldwide. The complete genome sequence for strain LGP32 is available along with partial genome sequences for strains MED222 and 12B01 (78). Genome comparison studies identified 409 genes unique to LGP32, and many of these genes were present in regions that had characteristics of genomic islands (78). An RS-like phage and 9 transposons were identified. Unusually one of the most divergent acquired genetic elements, the integron, was missing; in strain 12B01 and in strain LGP31 it only contained a limited number of cassettes (78).

The genome sequence of *V. anguillarum* strain 775 was compared with draft sequences of strains 96F and rv22 and *V. ordalii* (4). *V. anguillarum* is known to cause vibriosis in nearly 50 different fish species as well as mollusks and crustaceans worldwide. Two serogroups, O1 and O2, of the 28 recognized serogroups are associated with virulent strains. The analysis of the genome sequence identified 10 genomic islands ranging in size from 4 to 140 kb that encoded integrases and transposases, and three regions were adjacent to tRNA loci (4). Although *V. anguillarum* and *V. ordalii* are closely related, their genome size differed significantly; *V. anguillarum* has a 4-MB genome compared with *V. ordalii* with a 3.4-MB genome. Analysis of *V. anguillarum* strain 86F identified a T3SS, which was related to the T3SS identified on chromosome II of *V. parahaemolyticus* RIMD2210933, and to that found in *V. cholerae* non-choleraigenic strains and *V. mimicus* (4). The T3SS was missing for the other strains examined, although remnants were present in strain 775 and *V. ordalii*, suggesting that it was deleted from these strains (4).

*V. salmonicida*, also known as *Aliivibrio salmonicida*, is a halophilic psychrophile responsible for cold water vibriosis (CV) in farmed Atlantic salmon and rainbow trout as well as captive Atlantic cod. Similar to all *Vibrio*, *V. salmonicida* contains two chromosomes as well as four circular plasmids, which encompass 2.7% of the entire genome and 111 protein-coding sequences (85). Plasmid pVSAL840 contains the tra locus, associated with plasmid conjugation, which shows high synteny with pYJ016 of *V. vulnificus* YJ016. pVSAL43 and pVSAL54 are predicted to encode acetyltransferases enabling antigenic variation at the cell surface and thus possible improved protection from the host immune...
**FIGURE 5** Genome context and arrangement of region XII cluster among *Vibrio* species. Open reading frames (ORFs) are indicated as arrows, the direction of which shows the direction of transcription; numbers underneath ORFs represent locus tags. ORFs of similar color represent homologous pathway genes among the different species examined. Asterisk represents incomplete genome sequence. The region is present in all *V. mimicus* strains examined. doi:10.1128/microbiolspec.VE-0009-2014.f5
response. Plasmid pVSAL320 contains an iron ABC transporter that may be involved in the nonsiderophore uptake of ferrous iron (85). These plasmids are not necessary to cause CV in salmon (134), but, based on encoded proteins, it appears they may play a role in the pathogenicity of *V. salmonicida*. Another defining characteristic of *V. salmonicida* is the extremely large number of insertion sequence (IS) elements found on both chromosomes and plasmids, totaling 288. For comparison, the closest relative of *V. salmonicida*, *V. fischeri*, contains only one IS element. A whole genome comparison with *V. fischeri* reveals breaks in synteny in *V. salmonicida* in regions flanked by IS elements. *V. fischeri* contains more than 70% orthologous coding sequence (CDSs) to *V. salmonicida*. *V. fischeri* is a known symbiont of the squid *Euprymna scolopes* and is required for bioluminescence in the squid light organ. *V. fischeri* ES114 is able to colonize *E. scolopes*, whereas *V. fischeri* MJ11, a Japanese fish species isolate, is unable to colonize squid. Comparative genomic studies revealed that *E. coli* ES114 is able to colonize *V. parahaemolyticus* and *V. vulnificus* strains, whereas *V. fischeri* MJ11, a Japanese fish species isolate, is unable to colonize squid. Comparative genomic studies revealed that *RscS*, a two-component sensor kinase necessary for colonization of the squid, is absent from the MJ11 strain (135). *Photobacterium profundum* SS9 was isolated from the Sulu Trough at 2,500 meters and is a piezophile that can withstand high pressure (136). In addition to the 2 bipartite chromosomes, *P. profundum* SS9 contains an 80-kb plasmid. *P. profundum* SS9 is 25% larger than *V. parahaemolyticus* and *V. vulnificus* (136). SS9 contains two complete operons for F1F0 ATP synthase, 3 cbb3 cytochrome oxidase genes, and an unusual di-heme cytochrome c gene, all believed to be necessary for metabolic activity at high pressures (136).

Within all these genomes, signatures of horizontally acquired DNA can be found either in the form of extra-chromosomal elements, such as plasmids, or as chromosomally encoded genetic elements, such as IS elements, transposons, genomic islands, prophages, and integrons. The number and type of horizontally acquired elements differ between species and also within species. Many of these elements contain functions that are critical for a specific lifestyle and many others have an as yet unknown function. One of the most well-studied and intriguing mobile genetic element is the integron, and, in the next section, we will describe the integron genome context and structure that is found within the completed genome sequences of *Vibrio*.

**SUPERINTEGRONS IN VIBRIO**

Integrons are assembly platforms that incorporate exogenous gene cassettes by site-specific recombination and convert them to functional genes. To date, three classes of integrons (*intI1* to *intI3*) implicated in multi-resistance phenomena have been characterized and are classified based on the homology of their integrase protein, which ranges from 43 to 58% (137). These multiresistant integrons (MRIs) are defective in self-transposition, but they are often found in association with ISs, transposons, and/or conjugative plasmids that can serve as vehicles for their mobility. Therefore, the MRIs are also called mobile integrons. All integrons characterized to date are composed of three key elements necessary for the capture of exogenous genes: a gene (*intI*) encoding an integrase belonging to the tyrosine-recombinase family; a primary recombination site (*attI*); and an outward-oriented strong promoter (PC) that directs transcription of the captured genes (138, 139).

The superintegron (SI) is a distinctive class of integrons. Like MRI, the SI is also composed of an integrase (*intI*) gene and other key elements. The *intIA* gene product has 45 to 50% identity with the three known integrases, IntI1 to IntI3 (137). In comparison with MRIs, which contain only 5 to 8 gene cassettes, SI contains a large array of gene cassettes (137, 139). For example, in *V. cholerae* N16961, the SI is 126 kb in size and contains 179 gene cassettes (215 ORFs) (140, 141). Another striking difference between MRIs and the SI is that the SI does not seem to be mobile, because it is located on the chromosome and not associated with mobile DNA elements (139, 142).

The SI is one of the largest genetic elements identified so far in the *V. cholerae* genome as well as in other vibrios. In comparison with other genetic elements, the SI is dynamic in terms of capturing or releasing gene cassettes (141). The gene cassettes can exist as circular molecules that are unable to replicate or as an integrated element at the *attI* site in an integron (143). The *attI* site has a 7-bp core site (CS) with the consensus of GTTAAACA or GTTAAAAT (143). The gene cassettes found in integron platforms normally have only two functional components, a gene or noncoding DNA and an imperfect inverted repeat located at the 3′ end of the gene known as 59-base element (59-be) or *attC* site (141, 143). The *attC* site is a diverse family of sequences that vary in length (57 to 141 bp) and functions as a recognition site for the site-specific integrase. The most highly conserved feature of the *attC* site is a 7-bp core site (CS) at the 3′ end with the consensus of GTTAAACA or GTTAAAAT (143, 143). The CS of an *attC* is perfectly complementary to its ICS.
Like attC sites in MRIs, the gene cassettes in SI are also followed by short (123 to 126 bp) repeat sequences commonly referred as XXR, where the first X stands for genus, the second X represents species, and R represents repeat. For example, VCR indicates V. cholerae repeat (144, 145). In general, cassette genesis occurs through recombination between attl and VCR sequences catalyzed by IntIA, the site-specific recombinase (139, 141). MRIs play a leading role in the acquisition and spread of antibiotic resistance genes among Gram-negative bacteria, especially among Enterobacteriaceae and Pseudomonas (141, 146). In contrast, the SI gene cassettes encode proteins for metabolic activities, virulence, DNA-modification enzymes, as well as antimicrobial agents; however, most of their functions are still unknown (139, 141).

NCBI microbial genome database shows numerous ongoing projects for whole genome sequencing of diverse Vibrio species and among them are 20 strains of 11 different species whose whole genome sequences have been completed (Table 4). We performed a PSI-BLAST (Position-Specific Iterated BLAST) search using the V. cholerae SI integrase IntIA (VCA0291) as a seed sequence on these 20 strains of 11 different species. We found that all 20 strains possessed a copy of IntIA, which is the primary indicator of the presence of an SI (Table 4). We constructed a phylogenetic tree based on the intIA gene and found that the sequences clustered the Vibrio strains in a species-specific manner (Fig. 6). The intIA sequence was 100% conserved in all 7 sequenced V. cholerae genomes. Similarly, three V. vulnificus strains shared identical intIA sequences and formed a cluster which branched with intIA from V. parahaemolyticus strains. Two undefined Vibrio species strains EX25 and EJY3 clustered with the V. parahaemolyticus clade. Vibrio sp. EX25 was very close to V. parahaemolyticus and Vibrio sp. EJY3 was more similar to V. harveyi. V. anguillarum, V. furnissii, and V. splendidus intIA genes were more distantly related to each other (Fig. 6). The two V. fischeri and one V. salmonicida intIA sequences clustered together, and were distinct from the other vibrios examined (Fig. 6). The ancestry and evolutionary relationship between the intIA gene sequences were compared with a

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain ID</th>
<th>Size (Kb)</th>
<th>(GC SI/genome)</th>
<th>No. XRR</th>
<th>No. ORF&lt;sup&gt;c&lt;/sup&gt;</th>
<th>5′-Flanking ORF</th>
<th>3′-Flanking ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. cholerae</td>
<td>N16961</td>
<td>125</td>
<td>42/48</td>
<td>177</td>
<td>216</td>
<td>VC0290 (L20)</td>
<td>VC0507 (Tnp)</td>
</tr>
<tr>
<td></td>
<td>IEC224</td>
<td>125</td>
<td>42/47</td>
<td>177</td>
<td>218</td>
<td>O3Y_14818 (L20)</td>
<td>O3Y_15923 (Tnp)</td>
</tr>
<tr>
<td></td>
<td>MJ-1236</td>
<td>119</td>
<td>42/47</td>
<td>137</td>
<td>163</td>
<td>VCD_000985 (L20)</td>
<td>VCD_000819 (Tnp)</td>
</tr>
<tr>
<td></td>
<td>2010DEL-1786</td>
<td>99</td>
<td>42/48</td>
<td>136</td>
<td>160</td>
<td>Vch1786_II0036 (L20)</td>
<td>Vch1786_II0197 (Tnp)</td>
</tr>
<tr>
<td></td>
<td>LMA3894-4</td>
<td>37</td>
<td>43/48</td>
<td>53</td>
<td>50</td>
<td>VCLMA_B0258 (L20)</td>
<td>VCLMA_B0309 (HP)</td>
</tr>
<tr>
<td></td>
<td>O395</td>
<td>119</td>
<td>41/48</td>
<td>139</td>
<td>188</td>
<td>VC395_A0327 (L20)</td>
<td>VC395_A0516 (Tnp)</td>
</tr>
<tr>
<td></td>
<td>M66-2</td>
<td>98</td>
<td>42/48</td>
<td>138</td>
<td>169</td>
<td>VCM66_A0289 (L20)</td>
<td>VCM66_A0466 (Tnp)</td>
</tr>
<tr>
<td>V. vulnificus</td>
<td>CMP6</td>
<td>151</td>
<td>41/47</td>
<td>218</td>
<td>152</td>
<td>VV1_2400 (L20)</td>
<td>VV1_2550 (Tnp)</td>
</tr>
<tr>
<td></td>
<td>YJ016</td>
<td>138</td>
<td>41/47</td>
<td>187</td>
<td>203</td>
<td>VV1942 (L20)</td>
<td>VV1738 (Tnp)</td>
</tr>
<tr>
<td></td>
<td>MO6-24/O</td>
<td>129</td>
<td>41/47</td>
<td>195</td>
<td>179</td>
<td>VVO6_01279 (L20)</td>
<td>VVO6_01459 (HP)*</td>
</tr>
<tr>
<td>Vibrio sp.</td>
<td>Ex25</td>
<td>114</td>
<td>41/45</td>
<td>95</td>
<td>127</td>
<td>VEA_003168 (CHP)*</td>
<td>VEA_003296 (GSH transferase)*</td>
</tr>
<tr>
<td>V. parahaemolyticus</td>
<td>RIMD2210633</td>
<td>47</td>
<td>40/45</td>
<td>68</td>
<td>77</td>
<td>VP1866 (CHP)*</td>
<td>VP1878 (Tnp)</td>
</tr>
<tr>
<td>Vibrio sp.</td>
<td>EJY3</td>
<td>190</td>
<td>43/45</td>
<td>31</td>
<td>173</td>
<td>VEJY3_09680 (acyltransferase)</td>
<td>VEJY3_08800 (HP)</td>
</tr>
<tr>
<td>V. harveyi (campbellii)</td>
<td>BAA-1116</td>
<td>1</td>
<td>43/45&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1</td>
<td>VIBHAR_02002 (GPD)*</td>
<td>VIBHAR_02004 (HP)*</td>
</tr>
<tr>
<td>V. anguillarum</td>
<td>775</td>
<td>68</td>
<td>41/45</td>
<td>88</td>
<td>120</td>
<td>VAA_01071 (L20)</td>
<td>VAA_00312 (Tnp)</td>
</tr>
<tr>
<td>V. furnissii</td>
<td>NCTC 11218</td>
<td>39</td>
<td>42/51</td>
<td>63</td>
<td>35</td>
<td>vfu_B01510 (ATPase ABC tr.)*</td>
<td>vfu_B000012 (AAFGH)*</td>
</tr>
<tr>
<td>V. splendidus</td>
<td>LGP32</td>
<td>18</td>
<td>40/44</td>
<td>17</td>
<td>24</td>
<td>VS_1891 (Cytochrome P450)</td>
<td>VS_1866 (HP)*</td>
</tr>
<tr>
<td>V. salmonicida</td>
<td>LFI1238</td>
<td>22</td>
<td>36/39</td>
<td>22</td>
<td>27</td>
<td>VSAI_J0533 (HP)*</td>
<td>VSAI_J0561 (Tnp)</td>
</tr>
<tr>
<td>V. fischeri</td>
<td>ES114</td>
<td>29</td>
<td>35/38</td>
<td>37</td>
<td>36</td>
<td>VFA0664 (HP)*</td>
<td>VFA0628 (MOBHM)*</td>
</tr>
<tr>
<td></td>
<td>MJ11</td>
<td>102</td>
<td>35/38</td>
<td>130</td>
<td>126</td>
<td>VFMI1_J0760 (HP)*</td>
<td>VFMI1_J0633 (HP)*</td>
</tr>
</tbody>
</table>

<sup>a</sup>(CHP), conserved (hypothetical protein); MOBHMT, 3-methyl-2-oxobutanooate hydroxymethyltransferase; GPD, phage late control gene D protein; ATPase ABC tr., ATPase of ABC transporter with duplicated ATPase domains; AAFGH, arginase/agmatinase/formiminoglutamate hydrolase; GSH, glutathione.

<sup>b</sup>XXR: 1st X = genus, 2nd X = species, R = repeat.

<sup>c</sup>Including intIA gene.

<sup>d</sup>Solitary intIA represents SI.

<sup>e</sup>Only 2 VHRs were identified, one in plasmid pRIBHAR (CP000791) and the other in chromosome II (CP000790).
FIGURE 6 Phylogenetic tree based on the Neighbor-Joining method (13) using intIA and concatenated sequences of rpoB, mdh and pyrC genes of 20 Vibrio strains of 11 different species whose whole genome sequences are completed. The locus tag of each gene was used whenever appropriate. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches (30). The evolutionary distances were computed by using the Jukes-Cantor method (52) and evolutionary analyses were conducted in MEGA5 (120). Here, Black circle and diamond indicate the strains that bear the superintegron in Chr II and Chr I, respectively. doi:10.1128/microbiolspec.VE-0009-2014.f6

Post-Genomic Analysis of Members of the Family Vibrionaceae
concatenated tree of three housekeeping genes for the same set of strains. The topology of intIA for all the Vibrio species analyzed was congruent with the HK tree (Fig. 6).

Next, we examined the insertion site of the SI and the SI content among the different Vibrio species. As observed by Rowe-Magnus and colleagues, the intIA gene among V. cholerae strains were all located downstream of the same rplT gene that codes for the ribosomal L20 protein (Table 4) (144, 145). Although the V. vulnificus clade branched with the V. parahaemolyticus clade, their chromosomal locations were different. The SI from the V. vulnificus strains had the same chromosomal location as V. cholerae at the rplT locus. However, within the V. parahaemolyticus clade, the SI was present at a different chromosomal location. V. parahaemolyticus and Vibrio sp. EJ25 had their SI inserted downstream of a conserved hypothetical protein VP1866 and VEA_003168, whose homologue is VC1310 in V. cholerae. The SI in V. harveyi strains was located downstream of VIBHAR_02002, which encodes a phage late control gene D protein, GPD, and in Vibrio sp. EJY3 the SI was inserted at VEJY3_09680, which encodes an acyltransferase involved in lipid metabolism. The intIA gene in V. fischeri strains was located downstream of a hypothetical protein with no homologue present in V. cholerae N16961. The intIA gene in V. furnissii was located downstream of vfu_B01510, which encodes an ABC transporter ATPase. In V. cholerae, V. anguillarum, and V. furnissii, the SI is located in chromosome II, whereas in all other species examined the SI is present on chromosome I. As suggested previously, the conserved chromosomal location of the SI within the relevant subclades can be helpful in identifying the SI structure from other species within the same subclade (144).

The intIA gene transcription is opposite to the SI gene cassettes and, hence, gene cassettes are always predicted to be present upstream of intIA (139, 141). Therefore, we extracted from each genome 200 kb of DNA upstream of the intIA gene. The presence of VXR (Vibrio species repeat) was determined in this 200-kb sequence by in silico hybridization using VCR/VXR sequence as a probe, as well as by the XXR program developed by Mazel and his colleagues (144). As summarized in Table 4, SIs from all 7 V. cholerae strains were highly diverse in terms of size, number of ORFs, and VXR. The rearrangement of gene cassettes is an important feature of SI diversity in V. cholerae as well as in other vibrios (139, 140, 141, 144). Interestingly, we found 99% identical SI structures in two strains of V. cholerae, N16961 and IEC224, which were isolated from different geographic locations and time points, from Bangladesh in 1975 and from Brazil in 1994, respectively. There were anomalies in total ORF numbers such as N16961, which contained 216, and IEC224, which contained 218, but this was due to the difference in ORF annotation. The presence of identical SIs, as well as their conserved phylogenetic grouping, indicates a shared evolutionary origin.

Among the completed genome sequences of Vibrio species, the size of the SI varied from 22 kb in V. salmonicida to 151 kb in V. vulnificus CMCP6. V. harveyi strain ATCC BAA-1116 did not contain any SI gene cassette, only a solitary intIA gene. Repeated attempts to identify a VXR (VHR) sequence in 200 kb± intIA gene were unsuccessful. Two sequences that fulfilled the VHR criteria were found, one in plasmid pBIBHAR (CP000791) and the other in chromosome II. Draft genome sequence of a V. harveyi strain CAIM 1792 showed the presence of SI sequence of 87.2 kb in chromosome I (147). Therefore, the presence of a solitary intIA gene in ATCC BAA-1116 strain may indicate the early stage of SI evolution or the loss of an entire gene cassette region. The NCBI database does not contain any completed genome of V. campbellii; we did find an SI element of 68.9 kb in size in V. campbellii strain DS40M4. The percent GC content of a region can be used as an indicator of evolutionary origin. We examined the %GC content of the SI elements compared with the entire genome and found that the SI %GC content was always lower (Table 4). To date, only a handful of gene cassette functions have been demonstrated such as antibiotic resistance, virulence, and metabolic activity (139, 141, 144, 145). The majority of genes within SIs do not have any counterpart in the GenBank database, and their function and role in cell fitness and survival remain to be determined.

ACKNOWLEDGMENTS

Research on Vibrio species in the Boyd group is supported by a National Science Foundation grant IOS-0918429 and a National Science Foundation CAREER award DEB-0844409. J.B.L. was supported by the Chemistry-Biology Interface graduate program and a University of Delaware graduate student fellowship. W.B. W. was supported in part by a Dissertation fellowship, and J.J.K. was supported by a BOYCAST Indian government fellowship.

REFERENCES


Post-Genomic Analysis of Members of the Family Vibrionaceae


113. Chowdhury N, Norris J, McAlister E, Lau SY, Thomas GH, Boyd EF. 2012. The VC1777-VC1779 proteins are members of a sialic acid-specific subfamily of TRAP transporters (SiaPQM) and constitute the sole route of sialic acid uptake in the human pathogen Vibrio cholerae. Microbiology 158:2158–2167.


