Prey Range and Genome Evolution of \textit{Halobacteriovorax marinus} Predatory Bacteria from an Estuary

Brett G. Enos,a Molly K. Anthony,a Joseph A. DeGiorgis,a,b Laura E. Williamsa

\textsuperscript{a}Department of Biology, Providence College, Providence, Rhode Island, USA
\textsuperscript{b}Cellular Dynamics Program, Marine Biological Laboratory, Woods Hole, Massachusetts, USA

\textbf{ABSTRACT} \textit{Halobacteriovorax} strains are saltwater-adapted predatory bacteria that attack Gram-negative bacteria and may play an important role in shaping microbial communities. To understand how \textit{Halobacteriovorax} strains impact ecosystems and develop them as biocontrol agents, it is important to characterize variation in predation phenotypes and investigate \textit{Halobacteriovorax} genome evolution. We isolated \textit{Halobacteriovorax marinus} BE01 from an estuary in Rhode Island using \textit{Vibrio} from the same site as prey. Small, fast-moving, attack-phase BE01 cells attach to and invade prey cells, consistent with the intraperiplasmic predation strategy of the \textit{H. marinus} type strain, SJ. BE01 is a prey generalist, forming plaques on \textit{Vibrio} strains from the estuary, \textit{Pseudomonas} from soil, and \textit{Escherichia coli}. Genome analysis revealed extremely high conservation of gene order and amino acid sequences between BE01 and SJ, suggesting strong selective pressure to maintain the genome in this \textit{H. marinus} lineage. Despite this, we identified two regions of gene content difference that likely resulted from horizontal gene transfer. Analysis of modal codon usage frequencies supports the hypothesis that these regions were acquired from bacteria with different codon usage biases than \textit{H. marinus}. In one of these regions, BE01 and SJ carry different genes associated with mobile genetic elements. Acquired functions in BE01 include the \textit{dnd} operon, which encodes a pathway for DNA modification, and a suite of genes involved in membrane synthesis and regulation of gene expression that was likely acquired from another \textit{Halobacteriovorax} lineage. This analysis provides further evidence that horizontal gene transfer plays an important role in genome evolution in predatory bacteria.

\textbf{IMPORTANCE} Predatory bacteria attack and digest other bacteria and therefore may play a role in shaping microbial communities. To investigate phenotypic and genotypic variation in saltwater-adapted predatory bacteria, we isolated \textit{Halobacteriovorax marinus} BE01 from an estuary in Rhode Island, assayed whether it could attack different prey bacteria, and sequenced and analyzed its genome. We found that BE01 is a prey generalist, attacking bacteria from different phylogenetic groups and environments. Gene order and amino acid sequences are highly conserved between BE01 and the \textit{H. marinus} type strain, SJ. By comparative genomics, we detected two regions of gene content difference that likely occurred via horizontal gene transfer events. Acquired genes encode functions such as modification of DNA, membrane synthesis and regulation of gene expression. Understanding genome evolution and variation in predation phenotypes among predatory bacteria will inform their development as biocontrol agents and clarify how they impact microbial communities.

\textbf{KEYWORDS} horizontal gene transfer, host range, marine ecosystem, mobile genetic element, predation
Predation is an important force shaping microbial communities, which include microbial species that prey on other microbes. Eukaryotic microbial predators have received the majority of attention; however, bacterial predators are found in a wide range of environments and attack bacteria and fungi (1). Predatory bacteria such as *Bdellovibrio bacteriovorus* attack animal and plant pathogens, which makes them a potential biocontrol agent and an alternative to antibiotics (2, 3). To further understand bacterial predation and inform development of predatory bacteria as biocontrol agents, it is important to characterize variation in predation phenotypes, such as prey range, and to examine evolution of predatory bacteria lineages at different scales.

*Halobacteriovorax* is a genus of predatory bacteria belonging to the Deltaproteobacteria. Similar to *Bdellovibrio bacteriovorus*, which is also a member of the Deltaproteobacteria, *Halobacteriovorax* exhibits a biphasic life cycle (4, 5). In the attack phase, small, highly motile predatory bacterial cells search for prey bacteria and attach to the prey cell envelope. The predatory cell then invades the prey periplasm and reshapess the prey cell envelope to form a bdelloplast. In the subsequent growth phase, the predatory cell residing in the periplasm secretes lytic enzymes into the prey cytoplasm. The enzymes digest prey cell contents, and the predatory cell uses the prey components to build its own macromolecules. After depleting the prey cell cytoplasm, the predatory cell divides into multiple progeny, which secrete lytic enzymes to lyse the bdelloplast and release themselves to enter the attack phase.

Because of the similarity in predatory life cycles between *Halobacteriovorax* and *Bdellovibrio bacteriovorus*, *Halobacteriovorax* species were originally classified within the genus *Bdellovibrio*. Analysis of 16S rRNA gene sequences led to an initial reclassification into the genus *Bacteriovorax* (6) and then a subsequent reclassification into the genus *Halobacteriovorax* within the family *Halobacteriovoraceae* (7). *Halobacteriovorax* is adapted to saltwater environments and is distributed worldwide in oceans, estuaries, and saltwater lakes (8). Analysis of gene sequences from *Halobacteriovorax* of different saltwater environments revealed multiple phylogenetic clusters or operational taxonomic units (9). The *H. marinus* type strain, SJ, belongs to cluster III and was isolated over 25 years ago off the coast of St. John’s Island in the Caribbean (5).

As a widespread, albeit seasonally fluctuating, member of saltwater ecosystems, *Halobacteriovorax* may play an important role in shaping microbial communities at these sites. One experiment compared the impact of naturally occurring *Halobacteriovorax* versus naturally occurring marine bacteriophage on mortality of *Vibrio parahaemolyticus* added to microcosms of surface water samples (10). *Halobacteriovorax* appeared to cause a larger reduction in *V. parahaemolyticus* cell density than bacteriophage. Studies of other ecosystems, such as the coral microbiome, have also suggested that *Halobacteriovorax* may impact microbial community structure (11).

How *Halobacteriovorax* shapes saltwater microbial communities depends in part on which bacterial species are susceptible to predation by different *Halobacteriovorax* strains. Tests of *Halobacteriovorax* isolates from various saltwater environments indicate that, in general, this genus has a broad prey range (12, 13). For example, predatory bacteria in saltwater aquarium and tidal pool samples attacked a phylogenetically diverse set of prey, including multiple species of *Vibrio*, *Pseudomonas*, and *Escherichia coli* (13). Other studies show that within the genus, some *Halobacteriovorax* isolates may have a narrower prey range: for example, *Halobacteriovorax* isolated from seawater attacked multiple strains of *V. parahaemolyticus* but did not attack two other *Vibrio* species, *E. coli*, or *Salmonella enterica* serovar Typhimurium (14). The prey species used to initially isolate *Halobacteriovorax* from water samples likely biases which predatory strains are recovered and therefore affects our understanding of variation in prey range phenotypes. This was shown when *Halobacteriovorax* strains with broader prey ranges were isolated from a tidal river using *E. coli* or *Salmonella* serovar Typhimurium (14).

To understand the adaptation and evolution of *Halobacteriovorax*, it is important to examine genome evolution across a range of phylogenetic distances. Currently, *H. marinus* SJ is the only complete genome for family *Halobacteriovoraceae* (5). Draft genomes are available for four strains representing four other phylogenetic clusters of...
Halobacteriovorax (15). Overall, genes in Halobacteriovorax show high sequence divergence, illustrated by the large proportion of predicted genes with no significant matches to other genera (5) and a relatively low average amino acid identity among the five Halobacteriovorax genomes (15). Genome evolution in Halobacteriovorax may be affected by horizontal gene transfer (HGT), with multiple regions of the H. marinus SJ genome showing signatures associated with foreign DNA (5). The extent of horizontal gene transfer and its impact on functional capacity are unknown.

To further investigate phenotypic and genotypic variation in Halobacteriovorax, we isolated a strain of H. marinus from an estuary using a Vibrio strain from the same site. We tested the prey range of the isolate against bacteria from the estuary and bacteria from other environments to explore variation in this predation phenotype. Comparative genomics with the closely related H. marinus type strain, SJ, revealed two regions of gene content difference that likely arose via horizontal gene transfer.

RESULTS
Small, fast-moving H. marinus BE01 cells invade prey cells. We isolated a strain of predatory bacteria from an estuary in Rhode Island using a Vibrio strain from the same estuary as prey. The predatory bacterial isolate has two copies of the 16S rRNA gene, one of which is identical to that of H. marinus type strain, SJ, whereas the other copy differs at only one nucleotide position. This supports classification of the isolate as H. marinus, and we further distinguish it as strain BE01. H. marinus BE01 and SJ have very similar cell morphologies. BE01 attack-phase cells are small and highly motile (Fig. 1A; see Movie S1 in the supplemental material). They have a characteristic vibroid (comma-shaped) morphology with a single polar flagellum (Fig. 1B). H. marinus BE01 forms completely clear uniform plaques on lawns of susceptible prey bacteria (Fig. 1C). Observations by 1,000× phase-contrast microscopy show that BE01 invades prey cells. The closely related type strain, SJ, occupies the periplasmic space of Gram-negative prey cells after invasion (5), suggesting that BE01 is also an intraperiplasmic predator.

H. marinus BE01 is a prey generalist. To assess prey range, we challenged H. marinus BE01 with different Gram-negative prey bacteria (see Table S1 in the supplemental material). To test BE01’s ability to attack bacteria that it is likely to encounter in its natural habitat, we isolated multiple Vibrio strains from the estuary site and chose four distinct strains based on 16S rRNA gene sequences (see Fig. S1 in the supplemental material). We also tested whether BE01 could attack Gram-negative isolates from other environments by challenging it with an Acinetobacter strain isolated from a freshwater lake, a Pseudomonas strain isolated from soil, and two strains of E. coli, including ML35, a commonly used prey strain in studies of Bdellovibrio. We considered BE01 able to attack a particular prey strain if plaques formed on a lawn of that strain in a double agar overlay assay. Based on the results presented in Table 1, H. marinus BE01 appears to be a prey generalist, attacking all four Vibrio strains as well as the Pseudomonas strain and both strains of E. coli. Plaque formation was consistent over three biological replicates.

The H. marinus BE01 genome is highly similar to SJ, but lacks plasmid. Table 2 shows general statistics for the chromosomes of H. marinus BE01 (CP017414) and SJ (NC_016620). The chromosome sequences of these two strains are very similar in size and identical in GC content. Average nucleotide identity (ANI) between the two strains is 98.2% when calculated by JSpecies using nucmer (16) and 98.0% when calculated at http://enve-omics.ce.gatech.edu/ani/ (17). Initially, we annotated the BE01 chromosome using the Prokaryotic Genome Annotation Pipeline (PGAP) at GenBank and compared it to the existing GenBank annotation of SJ. PGAP classified more protein-coding genes as encoding hypothetical proteins in BE01 compared to SJ (2,398 versus 1,571). Some of these classifications in the BE01 chromosome appear overly conservative; for example, BIY24_00015 in strain BE01 is annotated as a hypothetical protein, although the amino acid sequence is 99% identical to BMS_0003 in strain SJ, which is annotated as DNA recombination protein RecF on the basis of conserved protein domain families. We therefore submitted both BE01 and SJ chromosome sequences to the Rapid Annotation using Subsystem Technology (RAST) server for annotation.
RAST classified a similar number of protein-coding genes as genes encoding hypothetical proteins in the two strains (Table 2), and the proportion of hypothetical proteins was closer to the PGAP annotation of SJ. We supplemented the RAST annotation with Infernal annotation (21) to detect RNA-coding sequences and proceeded with our analyses using the RAST/H11001 Infernal annotations, which can be found as text files at the Figshare repository (https://figshare.com/projects/Supporting_data_for_Halobacteriovorax_BE01_paper/24229).

The genome of *H. marinus* SJ includes a small (1,973-bp) plasmid with a single coding sequence (5). To determine whether *H. marinus* BE01 harbors a plasmid, we used megablast to identify the top hits for each of the 93 contigs generated by *de novo* assembly using PacBio reads. With the exception of the contig corresponding to the BE01 chromosome, all contigs aligned with at least 97% similarity to *E. coli* sequences in the nonredundant GenBank database. This is expected because we did not separate...
Table 1: Prey range of Halobacteriovorax marinus BE01

<table>
<thead>
<tr>
<th>Genus</th>
<th>Strain ID</th>
<th>Environment</th>
<th>Plaque formation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vibrio</td>
<td>0024</td>
<td>Estuary</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>0026</td>
<td>Estuary</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>0027</td>
<td>Estuary</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>0028</td>
<td>Estuary</td>
<td>Yes</td>
</tr>
<tr>
<td>Acinetobacter</td>
<td>0036</td>
<td>Freshwater</td>
<td>No</td>
</tr>
<tr>
<td>Pseudomonas</td>
<td>0042</td>
<td>Soil</td>
<td>Yes</td>
</tr>
<tr>
<td>Escherichia</td>
<td>0057</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ML3S</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Predatory bacterial cells from E. coli prey cells before extracting genomic DNA. The megablast results demonstrate that de novo assembly was able to handle the mixed pool of reads (H. marinus predatory bacteria and E. coli prey) and assemble reads from each species into separate contigs. Based on the megablast results, we conclude that the H. marinus BE01 genome consists of one chromosome and no plasmids.

Conservation of synteny and amino acid sequences between H. marinus genomes. Using RAST, we identified 3,048 bidirectional best hits between BE01 and SJ. To check the accuracy of the RAST analysis, we used the Reciprocal Smallest Distance algorithm (22), which detected 3,040 orthologs. By plotting the position of the RAST bidirectional best hits on each chromosome, we observed extremely high conservation of gene order between the two H. marinus strains (Fig. 2A). We did not detect any major inversions, translocations, or duplications. Most bidirectional best hits between BE01 and SJ have high amino acid sequence identity. In particular, 86% of bidirectional best hits (2,610/3,048) have at least 98% amino acid identity, and 94% (2,865/3,048) have at least 96% amino acid identity (Fig. 3). Only 27 bidirectional best hits have <70% identity at the amino acid sequence level, and many of these genes occur in one of the two major regions of difference detected in the synteny plot (Fig. 2B). Such high conservation of gene order and amino acid sequence across the chromosome suggests that the lineage of H. marinus represented by these two strains is experiencing strong purifying selection.

Differences in gene content between H. marinus genomes. The synteny plot of bidirectional best hits revealed two major regions of difference in gene content between H. marinus BE01 and H. marinus SJ (Fig. 2). One of these regions (region B in Fig. 2B) is bounded by a hypothetical protein (BE01_721 and SJ_717; see Table S2 in the supplemental material for the corresponding PGAP locus tags) and a TonB-dependent outer membrane receptor (BE01_770 and SJ_792). In BE01, region B encompasses 48 genes, 32 of which (67%) are unique to BE01, whereas in SJ, this region encompasses 74 genes, 56 of which (76%) are unique to SJ. In BE01, 7 of the 48 genes are unidirectional best hits against the SJ genome, with <65% amino acid identity, and 9 of the 48 genes are bidirectional best hits, with >60% amino acid identity.

Regarding functions annotated in region B in the BE01 genome, 23 of the 48 genes (48%) are hypothetical proteins with no predicted function. Three genes are annotated with functions related to horizontal gene transfer. BE01_722 is annotated as coding for a mobile

Table 2: Chromosome statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>H. marinus BE01</th>
<th>H. marinus SJ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PGAP</td>
<td>RAST+Infernal</td>
</tr>
<tr>
<td>Genome size (bp)</td>
<td>3,393,238</td>
<td>3,393,238</td>
</tr>
<tr>
<td>GC content (%)</td>
<td>36.7</td>
<td>36.7</td>
</tr>
<tr>
<td>Genes (no.)</td>
<td>3,253</td>
<td>3,288</td>
</tr>
<tr>
<td>Protein-coding genes (no.)</td>
<td>3,201</td>
<td>3,238</td>
</tr>
<tr>
<td>Hypothetical proteins (no.)</td>
<td>2,398</td>
<td>1,255</td>
</tr>
<tr>
<td>rRNA (no.)</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>tRNA (no.)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Other RNAs (no.)</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>
element protein, with hits to a COG (COG3464) and a PFAM domain (pfam01610) for a transposase family. Two consecutive genes, BE01_727 and -728, are both annotated as coding for integrases. BE01_727 is a bidirectional best hit to SJ_739, whereas BE01_728 is a unidirectional best hit for the same SJ gene. BLASTX analysis of the nucleotide sequence spanning these two genes and the intergenic regions suggests that the two genes are pseudogenes of the full-length integrase. Accumulation of mutations has degraded the gene, leaving two frameshifted ORFs that align to different regions of the full-length SJ integrase sequence with 67% and 57% amino acid identity by blastp.

The presence of genes associated with mobile genetic elements led us to examine the genes unique to BE01 in this region, which may be the result of horizontal gene transfer (HGT) events. We found two sets of genes indicative of HGT. One set of five genes (BE01_733 to -737) contains dnd genes involved in phosphorothioation of DNA. The dnd operon is not found in H. marinus SJ, but it is found in multiple divergent bacterial lineages, with phylogenetic evidence suggesting horizontal transfer (23, 24). We attempted to identify a likely source of the BE01 dnd operon, but each Dnd protein

![Synteny plot of bidirectional best hits between H. marinus BE01 and SJ.](image)

FIG 2  Synteny plot of bidirectional best hits between H. marinus BE01 and SJ. Bidirectional best hits identified by RAST are plotted based on their gene number on each chromosome. Individual genes are denoted with symbols corresponding to the similarity between BE01 and SJ amino acid sequences. Panel A shows the entire chromosomes, whereas panel B highlights the two major regions of difference in gene content (labeled A and B). id, identity.
aligned with <55% identity to protein sequences in the database and had a different bacterial species as the top hit in blastp analysis, thereby providing no clear evidence of the donor species.

In addition to the *dnd* operon, we identified a set of nine genes (BE01_761 to -769) that may have been acquired from another *Halobacteriovorax* lineage. By blastp analysis, each of the amino acid sequences has 37 to 64% identity (query coverage, \( \geq 97\% \)) to sequences in *Halobacteriovorax* sp. strain BAL6_X, which belongs to a different phylogenetic cluster than SJ and BE01. The nine genes are in the same order and orientation in BE01 and BAL6_X and include three genes involved in fatty acid and phospholipid metabolism and two genes encoding proteins with similarity to RNA polymerase sigma factor RpoE and an anti-sigma factor.

We also examined genes unique to SJ in region B to identify possible HGT events experienced by this strain. Forty-six of the 56 unique SJ genes were annotated as coding for hypothetical proteins, with no predicted function. The remaining 10 genes included six genes associated with mobile genetic elements, including a site-specific recombinase gene (SJ_729), an RNA-directed DNA polymerase gene (SJ_745), and four consecutive genes encoding a phage transcriptional regulator (SJ_755) and a type I restriction-modification system (SJ_756-758). Overall, analysis of region B in BE01 and SJ suggests that it may be a “hot spot” for incorporation of mobile genetic elements in this lineage of *H. marinus*.

The other major region of difference (region A in Fig. 2B) is bounded by chaperone protein DnaK (BE01_310 and SJ_313) on one end. On the other end, this region is bounded by different mannosyltransferases (BE01_364 or SJ_375), which are not each other’s bidirectional best hit. In BE01, region A encompasses 53 genes, 29 of which (55%) are unique to BE01, whereas in SJ, this region encompasses 61 genes, 25 of which (41%) are unique to SJ. In BE01, 12 of the 53 genes are unidirectional best hits against the SJ genome, with \( \leq 40\% \) amino acid identity, and 12 are bidirectional best hits, only two of which have amino acid identity of \( > 50\% \). In contrast to region B, which contains mostly unique gene content in each *H. marinus* strain, region A appears to have experienced more recombination and divergence of shared gene content (Fig. 2B).

Regarding functions annotated in region A in the BE01 genome, 15 of the 53 genes (28%) are hypothetical proteins with no predicted function. Among the remaining genes, we identified 22 genes with transferase activity, either annotated as transferases by RAST or classified as a transferase by analysis with InterProScan (GO term or detailed...
domain signature match). Ten of these transferase genes are unique to BE01. We also identified six genes involved in polysaccharide biosynthesis, three of which are unique to BE01. We did not detect genes associated with mobile genetic elements, such as transposases or integrases, in this region in BE01.

Overall, these two regions encompass 61 of the 147 total unique genes (41%) in BE01 and 81 of the 187 total unique genes (43%) in SJ. A large proportion of unique genes across the whole chromosome are annotated as coding for hypothetical proteins (70% in BE01 and 73% in SJ). These ORFs may not encode functional proteins; therefore, the number of unique protein-coding genes may be even lower. This emphasizes the high degree of shared gene content between *H. marinus* BE01 and SJ, with the two regions described above encompassing the majority of unique or highly divergent genes.

**Modal codon usage indicates horizontal gene transfer in regions of difference.**

To further explore the possibility of horizontal gene transfer in this lineage of *H. marinus*, we analyzed modal codon usage frequencies in both BE01 and SJ. Codon usage bias, in which certain codons are preferred for a particular amino acid, differs among bacterial species. Within a bacterial chromosome, regions with a codon usage bias that differs from that of the rest of the chromosome may have been horizontally transferred, although this is not the only explanation (25). Here, we analyzed modal codon usage, which describes the codon usage frequencies of the largest number of genes in a given sequence (26). We compared the entire chromosomes of BE01 and SJ and found a very small distance between the modes of codon usage frequencies (Table 3). This is expected based on the high average nucleotide identity between these two strains.

To test our hypothesis that the regions of gene content difference discussed above were acquired by horizontal gene transfer from a bacterial species with a different codon usage bias, we performed within-genome pairwise comparisons of modal codon usage. For example, to compare region A in *H. marinus* BE01 to the rest of the BE01 chromosome (excluding regions A and B), we calculated the modes of codon usage frequencies for region A and the rest of the chromosome and then determined the distance between these two modes (fourth column in Table 3). To test the null hypothesis that region A and the rest of the chromosome shared the same modal codon usage frequencies, the software used a “shuffled sequence” approach. This involved combining all the genes from region A and the rest of the chromosome into a single pool. From this pool, the software generated two new random sets of genes: one with the same number of genes as region A and another with the same number of genes as the rest of the chromosome. The software repeated this nine additional times, calculating the modal codon usage frequencies and the distance between the modes for each pair of “shuffled” sequences. After 10 rounds, the software calculated an average distance and standard deviation (fifth column in Table 3), which provides a measure of the expected distance if region A and the rest of the chromosome shared the same modal codon usage frequencies (26).

Using this approach, we found that distances between the modes of codon usage frequencies for each region and the rest of the chromosome were larger than expected in both BE01 and SJ (Table 3). This supports the hypothesis that these regions were

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Sequence 1</th>
<th>Sequence 2</th>
<th>Distance between sequence modes</th>
<th>Distance between modes of shuffled sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole chromosomes</td>
<td>BE01</td>
<td>SJ</td>
<td>0.0241</td>
<td>0.0402 ± 0.0049</td>
</tr>
<tr>
<td>BE01 regions</td>
<td>Region A</td>
<td>Chromosome (excluding regions A and B)</td>
<td>0.3589</td>
<td>0.1047 ± 0.0117</td>
</tr>
<tr>
<td></td>
<td>Region B</td>
<td>Chromosome (excluding regions A and B)</td>
<td>0.2976</td>
<td>0.1149 ± 0.0171</td>
</tr>
<tr>
<td></td>
<td>Region A</td>
<td>Region B</td>
<td>0.2607</td>
<td>0.1429 ± 0.0257</td>
</tr>
<tr>
<td>SJ regions</td>
<td>Region A</td>
<td>Chromosome (excluding regions A and B)</td>
<td>0.3756</td>
<td>0.0985 ± 0.0092</td>
</tr>
<tr>
<td></td>
<td>Region B</td>
<td>Chromosome (excluding regions A and B)</td>
<td>0.2237</td>
<td>0.0913 ± 0.0133</td>
</tr>
<tr>
<td></td>
<td>Region A</td>
<td>Region B</td>
<td>0.3053</td>
<td>0.1385 ± 0.0166</td>
</tr>
</tbody>
</table>
acquired via horizontal gene transfer from bacterial species with different codon usage biases compared to _H. marinus_. Highly expressed genes may also have different codon usage biases; however, given the annotated functions of genes in these regions, it is unlikely that this explains the distances observed. We also tested the distance between the modes of codon usage frequencies for the two regions themselves. These distances were also larger than expected in both BE01 and SJ (Table 3). This suggests that these regions were not acquired from a single bacterial species.

**DISCUSSION**

Based on prey range tests, _H. marinus_ BE01 appears to be a prey generalist. It is capable of attacking _Vibrio_ species isolated from the same estuary site as well as _Pseudomonas_ from soil and two strains of _E. coli_, including ML35. We have used the same strain of _Pseudomonas_ to isolate _Bdellovibrio_ from soil (unpublished observations), and _E. coli_ ML35 is often used to culture _Bdellovibrio_ isolated from both freshwater and soil (for example, see reference 27). Our finding that _H. marinus_ BE01 can attack these strains contrasts with reported observations that saltwater-adapted predatory bacteria generally do not attack the same prey species as _Bdellovibrio_ (5). Characterization of variation in predation phenotypes such as prey range is important for understanding diversity and adaptation in predatory bacteria. In this case, we used a _Vibrio_ strain from the same site to isolate _H. marinus_ BE01 rather than known strains such as _V. parahaemolyticus_ P-5. This is a useful strategy for increasing the diversity of predatory bacteria recovered from different environments.

Comparison of the genomes of _H. marinus_ BE01 and SJ clearly demonstrates that these strains are very closely related. The high conservation of synteny, gene content, and nucleotide sequence is striking given the geographic distance between the sampling sites and the amount of time between sample collections. In particular, the lack of genome rearrangements and the large proportion of amino acid sequences with at least 96% identity between the two strains suggest strong selective pressure to maintain the genome within this lineage of _H. marinus_.

Despite this selective pressure, there are two genomic regions with evidence of past horizontal gene transfer events. Comparison of modal codon usage frequencies within these regions and the rest of the _H. marinus_ chromosomes supports a hypothesis that these regions were acquired from bacterial species with different codon usage biases compared to _Halobacteriovorax_. One of the regions (region B) has a high proportion of unique gene content in both _H. marinus_ BE01 and _H. marinus_ SJ, including genes associated with mobile genetic elements, such as transposons and bacteriophage.

Within region B, one set of nine genes was likely acquired from another _Halobacteriovorax_ lineage. The donor may belong to _Halobacteriovorax_ phylogenetic cluster X, since the top blastp hit for each of the BE01 genes is the sequenced representative of this cluster (15). However, the current database is limited to only five _Halobacteriovorax_ genomes. Without more genome data, it is unclear whether this suite of genes was exchanged directly between BAL6_X and BE01 and then diverged (thereby explaining <65% amino acid identity), or if it has been exchanged widely among other _Halobacteriovorax_ strains, accumulating mutations in each new host. In the latter scenario, sequencing of additional _Halobacteriovorax_ isolates from multiple phylogenetic clusters may identify a lineage with genes more similar in sequence to those of BE01. It is also unclear how these genes affect BE01’s functional capacity. Based on their annotations, they may affect membrane synthesis and regulation of gene expression.

Region B in BE01 also includes the _dnd_ operon, which encodes a pathway for DNA modification. The _dnd_ operon is found in multiple bacterial lineages. Phylogenies of individual _dnd_ genes and investigations of genomic context suggest acquisition via horizontal gene transfer (23, 24). Operons that include _dndA_ typically have two orientations: _dndA_ divergently transcribed from _dndBCDE_ or _dnaAEDCB_ transcribed in the same direction (28). The operon in BE01 has the latter orientation, which may be less common based on genome surveys (28). Although our analysis did not reveal a likely source of the _dnd_ operon in _H. marinus_ BE01, this operon has been reported in coastal _Vibrionaceae_ and metagenome
data of ocean samples (23). The Dnd protein sequences in BE01 are highly divergent from Dnd proteins in the GenBank database; therefore, it is unclear whether the dnd operon is functional in H. marinus BE01. In other bacteria, Dnd proteins are involved in phosphorothioation, in which sulfur replaces a nonbridging oxygen molecule in the phosphate of the DNA backbone (29, 30). Some researchers have suggested that phosphorothioation may protect genomic DNA from degradation by nucleases (29). Predatory bacteria such as Halobacteriovorax rely on nucleases to digest prey DNA. If functional in BE01, the dnd operon may provide a horizontally acquired self-defense mechanism for H. marinus BE01 to protect its DNA from its own nucleases.

Similar to comparative genomics studies of Bdellovibrio (31–33), our analysis implicates an important role for horizontal gene transfer in the evolution of saltwater-adapted Halobacteriovorax. The extent to which predatory bacteria acquire genes from prey bacteria during predation is an interesting open question. Intraperiplasmic predators such as Halobacteriovorax and Bdellovibrio bacteriovorans secrete nucleases into the prey cytoplasm to digest genomic DNA. In B. bacteriovorans 109J, extensive degradation of host genomic DNA occurs within 60 min after prey cell invasion, yielding fragments with an average size of 700 bp (34). However, it is possible that partially digested fragments could be incorporated into the genome of the predatory bacteria cell during intraperiplasmic growth. It is also possible that partially digested fragments are released upon lysis of the prey cell by predatory progeny, enabling predatory bacterial cells in close proximity to take up the fragments and incorporate them into their chromosome.

In addition to unique genes suggestive of horizontal gene transfer, we also examined shared gene content between H. marinus BE01 and SJ. The high amino acid identity observed between these two strains is contrasted by the divergence of these sequences compared to database sequences, as described by Crossman and colleagues in their analysis of the SJ genome (5). We also observed a high proportion of hypothetical proteins or proteins of unknown function. RAST annotations identified 39% of BE01 protein-coding genes and 40% of SJ protein-coding genes as hypothetical proteins. This emphasizes the need for characterization of these predicted genes to determine whether they encode a protein and, if so, the function of that protein.

With a broad prey range such as observed here with H. marinus BE01, Halobacteriovorax species may exert a significant impact on microbial community structure in ecosystems such as estuaries. How these predatory bacteria affect nutrient cycling and fit into food web interactions is a key question for understanding these ecosystems (35). In addition, predatory bacteria and Halobacteriovorax strains in particular have shown promise as an alternative to antibiotics in the control of bacterial pathogens (36). Characterization of phenotypic and genotypic variation in a diverse range of Halobacteriovorax strains provides important information to advance development of these bacteria as biocontrol agents.

MATERIALS AND METHODS

Isolation and classification of environmental bacteria from estuary for use as prey. We isolated bacteria from Mount Hope Bay, an estuary in Bristol, RI (41.6971°N, −71.24578°W) for use as potential prey. We collected water from 1 m below the surface in sterile sample bottles and then filtered 100 ml through a 0.45-μm-pore 47-mm membrane filter (Pall Corporation, Ann Arbor, MI). We placed the filter in a 50-mm petri dish on an absorbent pad presoaked with either 2 ml of sea water yeast extract (SWYE) broth, (37) or Luria-Bertani (LB) broth (also known as lysogeny broth; Becton, Dickinson and Company, Sparks, MD) with 3% NaCl (Amresco Life Science, Solon, OH). We incubated the filters at 29°C and then picked colonies and streaked them onto plates of the same growth medium used to presoak the filter. We performed four rounds of streak plates to ensure pure isolates. Using a similar approach, Acinetobacter strain 0036 was isolated from a freshwater lake, and Pseudomonas strain 0042 was isolated from soil. We obtained E. coli ML35 from Mark Martin (University of Puget Sound, Puget Sound, WA) and E. coli 0057 from Brett Pellock (Providence College, Providence, RI).

To classify these isolates, we performed PCR targeting the 16S rRNA gene. We used primers 63F (38) and 1378R (39) with KAPA hi-fi (high fidelity) DNA polymerase (KAPA Biosystems, Wilmington, MA). The PCR cycle conditions were 95°C for 5 min, followed by 30 cycles of 95°C for 30 s, 50°C for 30 s, and 72°C for 2 min and a final extension step of 72°C for 10 min. After confirming the presence of a PCR product by gel electrophoresis, we purified PCR products using the Ultra Clean PCR cleanup kit (Mo Bio, Inc., Carlsbad, CA) and quantified them on a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA). Sanger sequencing used the same primers as amplification and was performed by...
After 48 h of incubation at 26°C and 200 rpm, we placed Formvar-coated EM grids on 30 of the prey strains in 35 ml of SWYE broth for 24 h for plaques but did not detect any. We observed plaques after 3 to 4 days. We picked plaques and then incubated the plates at 25°C. To check for possible bacteriophage, we examined the plates after 24 h for plaques but did not detect any. We observed plaques after 3 to 4 days. We picked plaques and made a lysate for each by placing a plaque in 20 ml of 100% IO with 1.5 ml of an overnight culture of E. coli strain 0057. After 2 days of incubation, we examined the lysates for active predatory bacterial cells. We selected the three lysates that appeared to have the highest phase-contrast microscopy to check for small, highly motile cells. After detecting predatory bacteria, we filtered the lysate through a 0.45-µm filter and repeated the double agar overlay technique to obtain individual plaques on a lawn of Vibrio strain 0024. The double agar overlay and plaque picking procedure was performed a total of three times to ensure a pure isolate of predatory bacteria. The lysate made from the final plaque-picking procedure was filtered through a 0.45-µm-pore filter. We combined 500 µl of this filtrate (containing cells of the pure isolate of predatory bacteria) with 500 µl sterile 50% glycerol (Sigma-Aldrich, St. Louis, MO) and stored this stock at −80°C.

**Prey range tests.** To obtain active *H. marinus* BE01 for prey range tests, we added a small amount of the −80°C stock to 15 ml of 100% IO mixed with 1 ml of an E. coli strain 0057 overnight culture. We chose to use *E. coli* because prior work reported viable Vibrio cells passing through 0.45-µm filters as minicells (14), which could confound the results of the prey range tests. We incubated the lysate at 26°C and 200 rpm. After 5 days, we filtered the lysate using a 0.45-µm filter to separate predatory bacteria from prey bacteria and cell debris. Swabs of the filter on LB plates confirmed that no viable *E. coli* cells passed through the filter. We performed 1:10 serial dilutions of the filtrate in 100% IO. To test prey range, we used the double agar overlay method described above to observe plaque formation. We cultured prey strains in 35 ml of SWYE broth for Vibrio prey strains or tryptic soy broth (TSB; Becton, Dickinson and Company) for all other prey strains. We centrifuged cultures at 6000 rpm for 10 min, washed the pellets in 100% IO, and then resuspended the pellets in 4 ml 100% IO. All prey resuspensions were at least 10^6 CFU/ml. For the prey range tests, we plated the 10−3 to 10−6 dilutions of the filtrate. We incubated plates at 26°C and checked for plaques daily, starting on day 3 until day 7. Plaque formation on any of these days was scored as positive for the prey range test. We repeated this procedure twice for each prey strain to obtain three biological replicates.

**EM.** To obtain BE01 samples for electron microscopy (EM), we added a small amount of the −80°C stock of *H. marinus* BE01 to 20 ml of 100% IO mixed with 1.5 ml of *E. coli* strain 0057 overnight culture. After 48 h of incubation at 26°C and 200 rpm, we placed Formvar-coated EM grids on 30 µl droplets of bacterial sample for 30 s to allow the bacteria to adhere to the Formvar surface. We then transferred grids to 50-µl drops of 1% uranyl acetate in water for 1 min. The grids were lifted from the drops of uranyl acetate, and the excess stain was wicked off with Whatman filter paper. The stained sample-coated grids were air dried for 10 min, and the resulting specimens were imaged with a JEOL CX 2000 transmission electron microscope (JEOL, Peabody, MA).

**Sequencing and assembly of *H. marinus* BE01 genome.** To obtain genomic DNA for sequencing, we cultivated BE01 using *E. coli* strain 0057 as prey. We chose to use *E. coli* because there is extensive genome information available that would allow us to screen reads to remove prey bacterial reads if necessary. To make lysates for genomic DNA preparation, we added a small amount of the −80°C stock of BE01 to 20 ml of 100% IO mixed with 1.5 ml of an overnight culture of *E. coli* strain 0057. After 2 days of incubation, we examined the lysates for active predatory bacterial cells. We selected the three lysates that appeared to have the highest ratio of predator to prey and pooled these. Because PacBio technology requires at least 10 µg of genomic DNA for library construction, we did not filter the lysates to avoid any potential loss of predatory bacterial cells. To extract genomic DNA, we used the Wizard Genomic DNA purification kit (Promega, Madison, WI) with the protocol for Gram-positive and Gram-negative bacteria. We centrifuged the pooled lysates at 9,100 rpm for 10 min and resuspended the pellet in 600 µl of the kit’s lysis solution for nuclei. We then continued with the manufacturer’s instructions. At the final step, we left the genomic DNA at 4°C overnight. By Qubit 2.0 (Thermo Fisher Scientific), the genomic DNA was at 299 µg/ml.

Library construction and sequencing were performed at the Institute for Genome Sciences at the University of Maryland Baltimore on a Pacific Biosciences RSII instrument (Pacific Biosciences, Menlo Park, CA) using P6-C4 chemistry. We launched an Amazon EC2 instance of SMRT portal 2.3.0 to analyze and assemble the data. Two SMRT cellswith generated 93,922 postfilter polymerase reads (N50 20,025 bp) and 151,636 subreads (N50 10,161 bp). We performed de novo assembly using the RS_HGAP_Assembly3 framework.
protocol (46) with default settings, except for genome size, which we changed to 3.5 Mbp. This generated 93 contigs in the polished assembly. The largest contig was 3,413,657 bp and aligned to *H. marinus* SJ by blastn. We used BLAST2Go (47) to align the 92 smaller contigs against the nonredundant database (restricted to *Bacteria*) with megablast to determine their sources.

To close the large *Halobacteriovorax* contig, we used Gepard (48) to identify overlaps between the ends of the contig, which indicated that the contig could be circularized. We used blastn alignments to specifically determine the overlap regions, which resulted in trimming 20,805 bp from the beginning of the contig. We then edited the trimmed contig so that the first nucleotide corresponded to the first nucleotide of the *dnaA* protein-coding sequence. To check the accuracy of the draft sequence at this stage, we aligned the PacBio reads against this draft sequence using the R5_Resequencing.1 protocol in SMRT Portal. The consensus sequence from this alignment had 342 differences compared to the draft sequence used as a reference.

To polish the sequence, we generated 150-bp paired-end Illumina reads. Library construction and sequencing (equivalent to 5% of a channel) were performed at IGS on an Illumina HiSeq (Illumina, San Diego, CA). We filtered the resulting reads so that every base in each read pair was \( \geq Q25 \). This yielded 6,604,606 read pairs. We aligned these read pairs to the recalled draft sequence using bwa-mem (49), yielding \( 507 \times \) average coverage (with a minimum of \( 51 \times \)). We used samtools (50) to convert the alignment to a sorted and indexed bam file. Finally, we used Pilon (51) to identify corrections based on the Illumina data, which amounted to 372 small insertions. The corrected sequence generated by Pilon was deposited in GenBank as the complete chromosome of *H. marinus* BE01.

**Genome annotation and analysis of gene content.** We annotated the *H. marinus* BE01 genome initially using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 3.3. Because of the unusually high proportion of hypothetical proteins identified by PGAP (see Results), we submitted both BE01 and SJ chromosome sequences to RAST (18–20) in January 2017. We used classic RAST with the gene caller and FIGfam release 70. We separately annotated RNA-coding genes using Infernal 1.1.2 (21). Files of the RAST+Infernal annotations and the output from RAST bidirectional best hit analysis are available at the Figshare repository (https://figshare.com/projects/Supporting_data_for_Halobacteriovorax_BE01_paper/24229). R code used to generate the synteny plot and the plot of amino acid identity for bidirectional best hits are available at the Figshare repository.

**Modal codon usage analysis.** To compare the modal codon usage frequencies, we used a freely available software package downloaded from http://www.life.illinois.edu/gary/programs/codon_usage.html (26).

**Accession number(s).** The genome sequence generated and analyzed during the present study is available under BioProject no. PRJNA343955, BioSample no. SAMN05806433, and GenBank accession no. CP017141.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at https://doi.org/10.1128/mSphere.00508-17.

**FIG S1**, PDF file, 0.1 MB.

**TABLE S1**, XLSX file, 0.1 MB.

**TABLE S2**, XLSX file, 0.1 MB.

**MOVIE S1**, MOV file, 9.5 MB.

**ACKNOWLEDGMENTS**

This research was supported by an Institutional Development award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant no. P20GM103430 and funding from Providence College. Neither funder played a role in study design, data analysis or interpretation, or writing the manuscript.

We thank Nicole Cullen for isolating *Pseudomonas* strain 0042 and Sean O’Donnell for isolating *Acinetobacter* strain 0036. We thank Lisa Sadzewicz and Luke Tallon at the Institute for Genome Sciences at the University of Maryland Baltimore for sequencing services. We are grateful to Mark Martin for providing *E. coli* ML35, Brett Pellock for providing *E. coli* 0057, and Cameron Thrash for guidance on growth media. L.E.W. thanks her network of scientific colleagues on Twitter for much useful advice and support.

**REFERENCES**


21. Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homol-


