bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 Prey range and genome evolution of *Halobacteriovorax marinus* predatory bacteria from
2 an estuary
3
4
5 Brett G. Enos[1], Molly K. Anthony[1], Joseph A. DeGiorgis[1,2] and Laura E. Williams[1]*
6 [1]Department of Biology, Providence College, Providence, RI
7 [2]Cellular Dynamics Program, Marine Biological Laboratory, Woods Hole, MA
8
9 benos@friars.providence.edu
10 manthon2@friars.providence.edu
11 jdegiorg@providence.edu
12 *Corresponding author: lwillia7@providence.edu
13

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

14 **Abstract**:

15 **Background:** *Halobacteriovorax* are saltwater-adapted predatory bacteria that attack
16 Gram-negative bacteria and therefore may play an important role in shaping microbial
17 communities. To understand the impact of *Halobacteriovorax* on ecosystems and develop
18 them as biocontrol agents, it is important to characterize variation in predation
19 phenotypes such as prey range and investigate the forces impacting *Halobacteriovorax*
20 genome evolution across different phylogenetic distances.

21 **Results:** We isolated *H. marinus* BE01 from an estuary in Rhode Island using *Vibrio*
22 from the same site as prey. Small, fast-moving attack phase BE01 cells attach to and
23 invade prey cells, consistent with the intraperiplasmic predation strategy of *H. marinus*
24 type strain SJ. BE01 is a prey generalist, forming plaques on *Vibrio* strains from the
25 estuary as well as *Pseudomonas* from soil and *E. coli*. Genome analysis revealed that
26 BE01 is very closely related to SJ, with extremely high conservation of gene order and
27 amino acid sequences. Despite this similarity, we identified two regions of gene content
28 difference that likely resulted from horizontal gene transfer. Analysis of modal codon
29 usage frequencies supports the hypothesis that these regions were acquired from bacteria
30 with different codon usage biases compared to *Halobacteriovorax*. In BE01, one of these
31 regions includes genes associated with mobile genetic elements, such as a transposase not
32 found in SJ and degraded remnants of an integrase occurring as a full-length gene in SJ.
33 The corresponding region in SJ included unique mobile genetic element genes, such as a
34 site-specific recombinase and bacteriophage-related genes not found in BE01. Acquired
35 functions in BE01 include the *dnd* operon, which encodes a pathway for DNA
36 modification that may protect DNA from nucleases, and a suite of genes involved in

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

37 membrane synthesis and regulation of gene expression that was likely acquired from
38 another *Halobacteriovorax* lineage.
39 **Conclusions:** Our results support previous observations that *Halobacteriovorax* prey on a
40 broad range of Gram-negative bacteria. Genome analysis suggests strong selective
41 pressure to maintain the genome in the *H. marinus* lineage represented by BE01 and SJ,
42 although our results also provide further evidence that horizontal gene transfer plays an
43 important role in genome evolution in predatory bacteria.
44

45 Keywords:
46 predation, horizontal gene transfer, host range, marine ecosystem, mobile genetic element
47

48 **Background**:

49 Predation is an important force shaping microbial communities, which include microbial
50 species that prey on other microbes. Eukaryotic microbial predators have received the
51 majority of attention; however, bacterial predators are found in a wide range of
52 environments and attack bacteria and fungi [1]. Predatory bacteria such as *Bdellovibrio*
53 *bacteriovorus* attack animal and plant pathogens, which makes them a potential
54 biocontrol agent and an alternative to antibiotics [2, 3]. To further understand bacterial
55 predation and inform development of predatory bacteria as biocontrol agents, it is
56 important to characterize variation in predation phenotypes such as prey range and
57 examine evolution of predatory bacteria lineages at different scales.
58

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

59 *Halobacteriovorax* is a genus of predatory bacteria belonging to the Deltaproteobacteria.
60 Similar to *Bdellovibrio bacteriovorus*, which is also a member of Deltaproteobacteria,
61 *Halobacteriovorax* exhibits a biphasic life cycle [4, 5]. In the attack phase, small, highly
62 motile predatory bacteria cells search for prey bacteria and attach to the prey cell
63 envelope. The predatory cell then invades the prey periplasm and re-shapes the prey cell
64 envelope to form a bdelloplast. In the subsequent growth phase, the predatory cell
65 residing in the periplasm secretes lytic enzymes into the prey cytoplasm. The enzymes
66 digest prey cell contents, and the predatory cell uses the prey components to build its own
67 macromolecules. After depleting the prey cell cytoplasm, the predatory cell divides into
68 multiple progeny, which secrete lytic enzymes to lyse the bdelloplast and release
69 themselves to enter the attack phase.

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

81

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

82 As a widespread, albeit seasonally fluctuating, member of saltwater ecosystems,
83 *Halobacteriovorax* may play an important role in shaping microbial communities at these
84 sites. One experiment compared the impact of naturally occurring *Halobacteriovorax*
85 versus naturally occurring marine bacteriophage on mortality of *Vibrio parahaemolyticus*
86 added to microcosms of surface water samples [10]. *Halobacteriovorax* appeared to
87 cause a larger reduction in *V. parahaemolyticus* cell density than bacteriophage. Studies
88 of other ecosystems, such as the coral microbiome, have also suggested that
89 *Halobacteriovorax* may impact microbial community structure [11].
90
91 How *Halobacteriovorax* shapes saltwater microbial communities depends in part on
92 which bacterial species are susceptible to predation by different *Halobacteriovorax*
93 strains. Tests of *Halobacteriovorax* isolates from various saltwater environments indicate
94 that, in general, this genus has a broad prey range [12, 13]. For example, predatory
95 bacteria in saltwater aquarium and tidal pool samples attacked a phylogenetically diverse
96 set of prey, including multiple species of *Vibrio*, *Pseudomonas* and *E. coli* [13]. Other
97 studies show that within the genus, some *Halobacteriovorax* isolates may have a
98 narrower prey range; for example, *Halobacteriovorax* isolated from seawater attacked
99 multiple strains of *V. parahaemolyticus* but did not attack two other *Vibrio* species, *E.*
100 *coli* or *Salmonella* serovar Typhimurium [14]. The prey species used to initially isolate
101 *Halobacteriovorax* from water samples likely biases which predatory strains are
102 recovered and therefore affects our understanding of variation in prey range phenotypes.
103 This was shown when *Halobacteriovorax* with broader prey ranges were isolated from a
104 tidal river using *E. coli* or *Salmonella* serovar Typhimurium [14].

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

105

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

117

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

124

125 **Results**:

126 *Small, fast-moving Halobacteriovorax marinus BE01 cells invade prey cells*

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

127 We isolated a strain of predatory bacteria from an estuary in Rhode Island using a *Vibrio*
128 strain from the same estuary as prey. The predatory bacteria isolate has two copies of the
129 16S rRNA gene, one of which is identical to that of *Halobacteriovorax marinus* type
130 strain SJ, whereas the other copy differs at only one nucleotide position. This supports
131 classification of the isolate as *H. marinus*, and we further distinguish it as strain BE01. *H.*
132 *marinus* BE01 and *H. marinus* SJ have very similar cell morphologies. BE01 attack
133 phase cells are small and highly motile (Figure 1a and Additional File 1). They have a
134 characteristic vibroid (comma-shaped) morphology with a single polar flagellum (Figure
135 1b). *H. marinus* BE01 forms completely clear, uniform plaques on lawns of susceptible
136 prey bacteria (Figure 1c). Observations by 1000x phase-contrast microscopy show that
137 BE01 invades prey cells. The closely related type strain SJ occupies the periplasmic
138 space of Gram-negative prey cells after invasion [5], suggesting that BE01 is also an
139 intraperiplasmic predator.
140
141 *Halobacteriovorax marinus BE01 is a prey generalist*
142 To assess prey range, we challenged *H. marinus* BE01 with different Gram-negative prey
143 bacteria (Additional File 2). To test BE01's ability to attack bacteria that it is likely to
144 encounter in its natural habitat, we isolated multiple *Vibrio* strains from the estuary site
145 and chose four distinct strains based on 16S rRNA gene sequences (Additional File 3).
146 We also tested whether BE01 could attack Gram-negative isolates from other
147 environments by challenging it with an *Acinetobacter* strain isolated from a freshwater
148 lake, a *Pseudomonas* strain isolated from soil and two strains of *E. coli*, including ML35,
149 a commonly used prey strain in studies of *Bdellovibrio*. We considered BE01 able to

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

150 attack a particular prey strain if plaques formed on a lawn of that strain in a double agar
151 overlay assay. Based on the results presented in Table 1, *H. marinus* BE01 appears to be
152 a prey generalist, attacking all four *Vibrio* as well as *Pseudomonas* and both strains of *E.*
153 *coli*. Plaque formation was consistent over three biological replicates.
154
155 Table 1. Prey range of *Halobacteriovorax marinus* BE01

Genus	Strain ID	Environment	Plaque formation?
Vibrio	0024	Estuary	Yes
Vibrio	0026	Estuary	Yes
Vibrio	0027	Estuary	Yes
Vibrio	0028	Estuary	Yes
Acinetobacter	0036	Freshwater	No
Pseudomonas	0042	Soil	Yes
Escherichia	0057		Yes
Escherichia	ML35		Yes

156
157 *Halobacteriovorax marinus BE01 genome is highly similar to SJ, but lacks plasmid*
158 Table 2 shows general statistics for the chromosomes of *H. marinus* BE01 (CP017414)
159 and SJ (NC_016620). The chromosome sequences of these two strains are very similar in
160 size and identical in GC content. Average nucleotide identity (ANI) between the two
161 strains is 98.2% when calculated by JSpecies using nucmer [16] and 98.0% when
162 calculated at http://enve-omics.ce.gatech.edu/ani/ [17]. Initially, we annotated the BE01
163 chromosome using the Prokaryotic Genome Annotation Pipeline (PGAP) at GenBank
164 and compared it to the existing GenBank annotation of SJ. PGAP classified more protein-
165 coding genes as hypothetical proteins in BE01 compared to SJ (2,398 versus 1,571).
166 Some of these classifications in the BE01 chromosome appear overly conservative; for
167 example, BIY24_00015 in strain BE01 is annotated as a hypothetical protein although the
168 amino acid sequence is 99% identical to BMS_0003 in strain SJ, which is annotated as

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

169 DNA recombination protein RecF on the basis of conserved protein domain families. We
170 therefore submitted both BE01 and SJ chromosome sequences to the Rapid Annotation
171 using Subsystem Technology (RAST) server for annotation [18–20]. RAST classified a
172 similar number of protein-coding genes as hypothetical proteins in the two strains (Table
173 2), and the proportion of hypothetical proteins was closer to the PGAP annotation of SJ.
174 We supplemented the RAST annotation with Infernal annotation [21] to detect RNA-
175 coding sequences and proceeded with our analyses using the RAST+Infernal annotations,
176 which can be found as text files at the figshare repository
177 (https://figshare.com/projects/Supporting_data_for_Halobacteriovorax_BE01_paper/242
178 29).
179
180 Table 2. Chromosome statistics

	Halobacteriovorax marinus BE01		*Halobacteriovorax marinus* SJ	
	PGAP	RAST+Infernal	PGAP	RAST+Infernal
Genome size (bp)	3,393,238		3,435,933	
GC content (%)	36.7		36.7	
Genes	3,253	3,288	3,307	3,350
Protein-coding	3,201	3,238	3,254	3,300
Hypothetical proteins	2,398	1,255	1,571	1,310
tRNA	36	36	36	36
rRNA	6	6	6	6
Other RNA	4	8	4	8

181
182 The genome of *H. marinus* SJ includes a small (1,973 bp) plasmid with a single coding
183 sequence [5]. To determine whether *H. marinus* BE01 harbors a plasmid, we used
184 megablast to identify the top hits for each of the 93 contigs generated by *de novo*
185 assembly using PacBio reads. With the exception of the contig corresponding to the
186 BE01 chromosome, all contigs aligned with at least 97% similarity to *E. coli* sequences in

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

187 the non-redundant GenBank database. This is expected because we did not separate
188 predatory bacteria cells from *E. coli* prey cells before extracting genomic DNA. Based on
189 the megablast results, we conclude that the *H. marinus* BE01 genome consists of one
190 chromosome and no plasmids.
191

192 *Conservation of synteny and amino acid sequences between H. marinus genomes*

193 Using RAST, we identified 3,048 bidirectional best hits between BE01 and SJ. To check
194 the accuracy of the RAST analysis, we used the Reciprocal Smallest Distance algorithm
195 [22], which detected 3,040 orthologs. By plotting the position of the RAST bidirectional
196 best hits on each chromosome, we observed extremely high conservation of gene order
197 between the two *H. marinus* strains (Figure 2a). We did not detect any major inversions,
198 translocations or duplications. Most bidirectional best hits between BE01 and SJ have
199 high amino acid sequence identity. In particular, 86% of bidirectional best hits
200 (2,610/3,048) have at least 98% amino acid identity, and 94% (2,865/3,048) have at least
201 96% amino acid identity (Figure 3). Only 27 bidirectional best hits have <70% identity at
202 the amino acid sequence level, and many of these genes occur in one of the two major
203 regions of difference detected in the synteny plot (Figure 2b). Such high conservation of
204 gene order and amino acid sequence across the chromosome suggests that the lineage of
205 *Halobacteriovorax marinus* represented by these two strains is experiencing strong
206 purifying selection.
207

208 *Differences in gene content between H. marinus genomes*

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

209 The synteny plot of bidirectional best hits revealed two major regions of difference in
210 gene content between *H. marinus* BE01 and *H. marinus* SJ (Figure 2). One of these
211 regions (region B in Figure 2b) is bounded by a hypothetical protein (BE01_721 and
212 SJ_717, see Additional File 4 for corresponding PGAP locus tags) and a TonB-dependent
213 outer membrane receptor (BE01_770 and SJ_792). In BE01, region B encompasses 48
214 genes, 32 of which (67%) are unique to BE01, whereas in SJ, this region encompasses 74
215 genes, 56 of which (76%) are unique to SJ. In BE01, seven of the 48 genes are
216 unidirectional best hits against the SJ genome, with <65% amino acid identity, and nine
217 of the 48 genes are bidirectional best hits, with >60% amino acid identity.
218
219 Regarding functions annotated in region B in the BE01 genome, 23 of the 48 genes
220 (48%) are hypothetical proteins with no predicted function. Three genes are annotated
221 with functions related to horizontal gene transfer. BE01_722 is annotated as a mobile
222 element protein, with hits to a COG (COG3464) and a PFAM domain (pfam01610) for a
223 transposase family. Two consecutive genes BE01_727-728 are both annotated as
224 integrases. BE01_727 is a bidirectional best hit to SJ_739, whereas BE01_728 is a
225 unidirectional best hit for the same SJ gene. BLASTX analysis of the nucleotide sequence
226 spanning these two genes and the intergenic regions suggests that the two genes are
227 pseudogenes of the full-length integrase. Accumulation of mutations has degraded the
228 gene, leaving two frameshifted ORFs that align to different regions of the full-length SJ
229 integrase sequence with 67% and 57% amino acid identity by BLASTP.
230

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

231 The presence of genes associated with mobile genetic elements led us to examine the
232 genes unique to BE01 in this region, which may be the result of horizontal gene transfer
233 events. We found two sets of genes indicative of HGT. One set of five genes (BE01_733-
234 737) encodes *dnd* genes involved in phosphorothioation of DNA. The *dnd* operon is not
235 found in *H. marinus* SJ, but it is found in multiple divergent bacterial lineages, with
236 phylogenetic evidence suggesting horizontal transfer [23, 24]. We attempted to identify a
237 likely source of the BE01 *dnd* operon, but each Dnd protein aligned with <55% identity
238 to protein sequences in the database and had a different bacterial species as the top hit in
239 BLASTP analysis, thereby providing no clear evidence of the donor species.
240
241 In addition to the *dnd* operon, we identified a set of nine genes (BE01_761-769) that may
242 have been acquired from another *Halobacteriovorax* lineage. By BLASTP analysis, each
243 of the amino acid sequences has 37-64% identity (query coverage ≥97%) to sequences in
244 *Halobacteriovorax* sp. BAL6_X, which belongs to a different phylogenetic cluster than
245 SJ and BE01. The nine genes are in the same order and orientation in BE01 and BAL6_X
246 and include three genes involved in fatty acid and phospholipid metabolism and two
247 genes encoding proteins with similarity to RNA polymerase sigma factor RpoE and an
248 anti-sigma factor.
249
250 We also examined genes unique to SJ in region B to identify possible HGT events
251 experienced by this strain. Forty-six of the 56 unique SJ genes were annotated as
252 hypothetical proteins, with no predicted function. The remaining ten genes included six
253 genes associated with mobile genetic elements, including a site-specific recombinase

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

254 (SJ_729), an RNA-directed DNA polymerase (SJ_745) and four consecutive genes
255 encoding a phage transcriptional regulator (SJ_755) and a Type I restriction-modification
256 system (SJ_756-758). Overall, analysis of region B in BE01 and SJ suggests that it may
257 be a "hotspot" for incorporation of mobile genetic elements in this lineage of *H. marinus*.
258
259 The other major region of difference (region A in Figure 2b) is bounded by chaperone
260 protein DnaK (BE01_310 and SJ_313) on one end. On the other end, this region is
261 bounded by different mannosyltransferases (BE01_364 or SJ_375), which are not each
262 other's bidirectional best hit. In BE01, region A encompasses 53 genes, 29 of which
263 (55%) are unique to BE01, whereas in SJ, this region encompasses 61 genes, 25 of which
264 (41%) are unique to SJ. In BE01, 12 of the 53 genes are unidirectional best hits against
265 the SJ genome, with ≤40% amino acid identity, and 12 are bidirectional best hits, only
266 two of which have amino acid identity >50%. In contrast to region B, which contains
267 mostly unique gene content in each *H. marinus* strain, region A appears to have
268 experienced more recombination and divergence of shared gene content (Figure 2b).
269
270 Regarding functions annotated in region A in the BE01 genome, 15 of the 53 genes
271 (28%) are hypothetical proteins, with no predicted function. Among the remaining genes,
272 we identified 22 genes with transferase activity, either annotated as transferases by RAST
273 or classified as a transferase by analysis with InterProScan (GO term or detailed domain
274 signature match). Ten of these transferase genes are unique to BE01. We also identified
275 six genes involved in polysaccharide biosynthesis, three of which are unique to BE01.

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

276 We did not detect genes associated with mobile genetic elements, such as transposases or
277 integrases, in this region in BE01.
278
279 Overall, these two regions encompass 61 of the 147 total unique genes (41%) in BE01
280 and 81 of the 187 total unique genes (43%) in SJ. A large proportion of unique genes
281 across the whole chromosome are annotated as hypothetical proteins (70% in BE01 and
282 73% in SJ). These ORFs may not encode functional proteins, therefore the number of
283 unique protein-coding genes may be even lower. This emphasizes the high degree of
284 shared gene content between *H. marinus* BE01 and *H. marinus* SJ, with the two regions
285 described above encompassing the majority of unique or highly divergent genes.
286
287 *Modal codon usage indicates horizontal gene transfer in regions of difference*
288 To further explore the possibility of horizontal gene transfer in this lineage of *H. marinus*,
289 we analyzed modal codon usage frequencies in both BE01 and SJ. Codon usage bias, in
290 which certain codons are preferred for a particular amino acid, differs among bacterial
291 species. Within a bacterial chromosome, regions with a codon usage bias that differs
292 from that of the rest of the chromosome may have been horizontally transferred, although
293 this is not the only explanation [25]. Here, we analyzed modal codon usage, which
294 describes the codon usage frequencies of the largest number of genes in a given sequence
295 [26]. We compared the entire chromosomes of BE01 and SJ and found a very small
296 distance between the modes of codon usage frequencies (Table 3). This is expected based
297 on the high average nucleotide identity between these two strains.
298

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

299 To test our hypothesis that the regions of gene content difference discussed above were
300 acquired by horizontal gene transfer from a bacterial species with a different codon usage
301 bias, we performed within-genome pairwise comparisons to determine the distances
302 between the modes of codon usage frequencies for the two regions and the rest of the
303 chromosome (excluding these two regions). Each pairwise comparison included a test of
304 the null hypothesis that the two sequences shared the same modal codon usage
305 frequencies. This is done within the software package by combining all the genes from
306 both sequences in a pairwise comparison and then generating two new random sets of
307 genes with the same number of genes as the original. The software then calculates the
308 modal codon usage frequencies of ten of these "shuffled" sequence pairs, generating an
309 average and standard deviation. The average distance of these ten shuffled sequence pairs
310 can be considered the expected distance if the two sequences shared the same modal
311 codon usage frequencies [26].
312
313 Using this approach, we found that distances between the modes of codon usage
314 frequencies for each region and the rest of the chromosome were larger than expected in
315 both BE01 and SJ (Table 3). This supports the hypothesis that these regions were
316 acquired via horizontal gene transfer from bacterial species with different codon usage
317 biases compared to *H. marinus*. Highly expressed genes may also have different codon
318 usage biases; however, given the annotated functions of genes in these regions, it is
319 unlikely that this explains the distances observed. We also tested the distance between the
320 modes of codon usage frequencies for the two regions themselves. These distances were

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

321 also larger than expected in both BE01 and SJ (Table 3). This suggests that these regions
322 were not acquired from a single bacterial species.
323
324 Table 3. Modal codon usage of *H. marinus* chromosomes and regions of gene content
325 difference within each chromosome

Comparison	Sequence 1	Sequence 2	Distance between Sequence Modes	Distance between Modes of Shuffled Sequences
Whole chromosomes	BE01	SJ	0.0241	0.0402 ± 0.0049
BE01 regions	Region A	Chromosome (excluding regions A and B)	0.3589	0.1047 ± 0.0117
	Region B	Chromosome (excluding regions A and B)	0.2976	0.1149 ± 0.0171
	Region A	Region B	0.2607	0.1429 ± 0.0257
SJ regions	Region A	Chromosome (excluding regions A and B)	0.3756	0.0985 ± 0.0092
	Region B	Chromosome (excluding regions A and B)	0.2237	0.0913 ± 0.0133
	Region A	Region B	0.3053	0.1385 ± 0.0166

326
327 **Discussion**:
328 Based on prey range tests, *H. marinus* BE01 appears to be a prey generalist. It is capable
329 of attacking *Vibrio* species isolated from the same estuary site as well as *Pseudomonas*
330 from soil and two strains of *E. coli*, including ML35. We have used the same strain of

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

331 *Pseudomonas* to isolate *Bdellovibrio* from soil (unpublished observations), and *E. coli*
332 ML35 is often used to culture *Bdellovibrio* isolated from both freshwater and soil (for
333 example, [27]). Our finding that *H. marinus* BE01 can attack these strains contrasts with
334 reported observations that saltwater-adapted predatory bacteria generally do not attack
335 the same prey species as *Bdellovibrio* [5]. Characterizing variation in predation
336 phenotypes such as prey range is important for understanding diversity and adaptation in
337 predatory bacteria. In this case, we used a *Vibrio* strain from the same site to isolate *H.*
338 *marinus* BE01 rather than known strains such as *V. parahaemolyticus* P-5. This is a
339 useful strategy for increasing the diversity of predatory bacteria recovered from different
340 environments.
341
342 Comparison of the genomes of *H. marinus* BE01 and SJ clearly demonstrate that these
343 strains are very closely related. The high conservation of synteny, gene content and
344 nucleotide sequence is striking given the geographic distance between the sampling sites
345 and the amount of time between sample collections. In particular, the lack of genome
346 rearrangements and the large proportion of amino acid sequences with at least 96%
347 identity between the two strains suggest strong selective pressure to maintain the genome
348 within this lineage of *H. marinus*.
349
350 Despite this selective pressure, there are two genomic regions with evidence of past
351 horizontal gene transfer events. Comparison of modal codon usage frequencies within
352 these regions and the rest of the *H. marinus* chromosomes supports a hypothesis that
353 these regions were acquired from bacterial species with different codon usage biases

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

354 compared to *Halobacteriovorax*. One of the regions (region B) has a high proportion of
355 unique gene content in both *H. marinus* BE01 and *H. marinus* SJ, including genes
356 associated with mobile genetic elements such as transposons and bacteriophage.
357
358 Within region B, one set of nine genes was likely acquired from another
359 *Halobacteriovorax* lineage. The donor may belong to *Halobacteriovorax* phylogenetic
360 cluster X, since the top BLASTP hit for each of the BE01 genes is the sequenced
361 representative of this cluster [15]. However, the current database is limited to only five
362 *Halobacteriovorax* genomes. Without more genome data, it is unclear whether this suite
363 of genes was exchanged directly between BAL6_X and BE01 and then diverged (thereby
364 explaining <65% amino acid identity), or if it has been exchanged widely among other
365 *Halobacteriovorax*, accumulating mutations in each new host. In the latter scenario,
366 sequencing of additional *Halobacteriovorax* isolates from multiple phylogenetic clusters
367 may identify a lineage with genes more similar in sequence to those of BE01. It is also
368 unclear how these genes affect BE01's functional capacity. Based on their annotations,
369 they may affect membrane synthesis and regulation of gene expression.
370
371 Region B in BE01 also includes the *dnd* operon, which encodes a pathway for DNA
372 modification. The *dnd* operon is found in multiple bacterial lineages. Phylogenies of
373 individual *dnd* genes and investigations of genomic context suggest acquisition via
374 horizontal gene transfer [23, 24]. Operons that include *dndA* typically have two
375 orientations: *dndA* divergently transcribed from *dndBCDE* or *dnaAEDCB* transcribed in
376 the same direction [28]. The operon in BE01 has the latter orientation, which may be less

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

377 common based on genome surveys [28]. Although our analysis did not reveal a likely
378 source of the *dnd* operon in *H. marinus* BE01, this operon has been reported in coastal
379 Vibrionaceae and metagenome data of ocean samples [23]. The *dnd* genes in BE01 are
380 intact, albeit highly divergent from *dnd* genes in the GenBank database. In other bacteria,
381 Dnd proteins are involved in phosphorothioation, in which sulfur replaces a non-bridging
382 oxygen molecule in the phosphate of the DNA backbone [29, 30]. Some researchers have
383 suggested that phosphorothioation may protect genomic DNA from degradation by
384 nucleases [29]. Predatory bacteria such as *Halobacteriovorax* rely on nucleases to digest
385 prey DNA. If functional in BE01, the *dnd* operon may provide a horizontally acquired
386 self-defense mechanism for *H. marinus* BE01 to protect its DNA from its own nucleases.
387
388 Similar to comparative genomics studies of *Bdellovibrio* [31–33], our analysis implicates
389 an important role for horizontal gene transfer in the evolution of saltwater-adapted
390 *Halobacteriovorax*. The extent to which predatory bacteria acquire genes from prey
391 bacteria during predation is an interesting open question. Intraperiplasmic predators such
392 as *Halobacteriovorax* and *Bdellovibrio bacteriovorus* secrete nucleases into the prey
393 cytoplasm to digest genomic DNA; however, it is possible that partially digested
394 fragments could be incorporated into the genome of the predatory bacteria cell during
395 intraperiplasmic growth. It is also possible that partially digested fragments are released
396 upon lysis of the prey cell by predatory progeny, enabling predatory bacteria cells in
397 close proximity to take up the fragments and incorporate them into their chromosome.
398

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

399 In addition to unique genes suggestive of horizontal gene transfer, we also examined
400 shared gene content between *H. marinus* BE01 and SJ. The high amino acid identity
401 observed between these two strains is contrasted by the divergence of these sequences
402 compared to database sequences, as described by Crossman and colleagues [5] in their
403 analysis of the SJ genome. We also observed a high proportion of hypothetical proteins
404 or proteins of unknown function. RAST annotations identified 39% of BE01 protein-
405 coding genes and 40% of SJ protein-coding genes as hypothetical proteins. This
406 emphasizes the need for characterization of these predicted genes to determine whether
407 they encode a protein, and, if so, the function of that protein.
408
409 With a broad prey range such as observed here with *H. marinus* BE01,
410 *Halobacteriovorax* species may exert a significant impact on microbial community
411 structure in ecosystems such as estuaries. How these predatory bacteria affect nutrient
412 cycling and fit into food web interactions is a key question for understanding these
413 ecosystems [34]. In addition, predatory bacteria and *Halobacteriovorax* in particular have
414 shown promise as an alternative to antibiotics in the control of bacterial pathogens [35].
415 Characterization of phenotypic and genotypic variation in a diverse range of
416 *Halobacteriovorax* provides important information to advance development of these
417 bacteria as biocontrol agents.
418
419 **Conclusions**:
420 The results of prey range tests of *H. marinus* BE01 are consistent with previous
421 observations that *Halobacteriovorax* prey on a broad range of Gram-negative bacteria.

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

422 Comparative genomics between *H. marinus* BE01 and the closely related *H. marinus*
423 type strain SJ suggest strong selective pressure to maintain the genome in this *H. marinus*
424 lineage. Despite this selective pressure, presence of mobile genetic element genes and
425 atypical modal codon usage frequencies suggest that horizontal gene transfer impacted
426 two genomic regions in this *Halobacteriovorax* lineage. HGT events provide these strains
427 with unique functional capacities that may impact their adaptation.
428
429 **Methods:**
430 *Isolation and classification of environmental bacteria from estuary for use as prey*
431 We isolated bacteria from Mount Hope Bay, an estuary in Bristol, RI (41.69717, -
432 71.24578) for use as potential prey. We collected water from 1 meter below the surface in
433 sterile sample bottles and then filtered 100 ml through a 0.45 µm 47 mm membrane filter.
434 We placed the filter in a 50 mm Petri dish on an absorbent pad presoaked with either 2 ml
435 of sea water yeast extract (SWYE) broth [36] or Luria-Bertani (LB) broth (also known as
436 lysogeny broth) with 3% NaCl. We incubated the filters at 29°C, then picked colonies
437 and streaked them onto plates of the same growth medium used to presoak the filter. We
438 performed four rounds of streak plates to ensure pure isolates. Using a similar approach,
439 *Acinetobacter* #0036 was isolated from a freshwater lake, and *Pseudomonas* #0042 was
440 isolated from soil.
441
442 To classify these isolates, we performed PCR targeting the 16S rRNA gene. We used
443 primers 63F [37] and 1378R [38] with KAPA HiFi (high fidelity) DNA polymerase. PCR
444 cycle conditions were: 95°C for 5 minutes followed by 30 cycles of 95°C for 30 seconds,

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

445 50°C for 30 seconds and 72°C for 2 minutes and a final extension step of 72°C for 10
446 minutes. After confirming presence of a PCR product by gel electrophoresis, we purified
447 PCR products using the Ultra Clean PCR Clean-Up Kit (Mo Bio) and quantified them on
448 a NanoDrop spectrophotometer. Sanger sequencing used the same primers as
449 amplification and was performed by GeneWiz (South Plainfield, NJ). We used
450 Phred/Phrap/Consed [39–41] to trim and assemble the reads, and we classified sequences
451 using BLAST [42], the SILVA Incremental Aligner [43] and the Ribosomal Database
452 Project classifier [44]. Additional File 2 shows the complete results of the classifications.
453
454 *Isolation of Halobacteriovorax marinus strain BE01*
455 To isolate predatory bacteria, we collected a water sample (31 ppt salinity measured with
456 a refractometer) from the same estuary site as described above following the same
457 procedure. We combined 20 ml with 1 ml of *Vibrio* strain #0024 at 10^9 cfu/ml and then
458 incubated this enrichment at 26°C and 200 rpm. Enrichments were examined daily for 2
459 to 4 days by 1000x phase contrast microscopy for the presence of small, highly motile
460 cells. Once we observed the presence of predatory bacteria, we filtered the enrichment
461 through a 0.45 μm filter. We performed a 10-fold serial dilution of the filtrate in sterile
462 100% Instant Ocean (IO). Dilutions were plated using a double agar overlay method.
463 Specifically, we added 1 ml of *Vibrio* strain #0024 at 10^9 cfu/ml to test tubes containing
464 3.3 ml of molten Pp20 top agar (1 g polypeptone peptone and 19.5 g agar dissolved in 1 L
465 of 70% IO). We vortexed to mix, then added 5 ml of the filtrate dilution to be plated and
466 vortexed again. We poured this mixture onto Pp20 plates (1 g polypeptone peptone and
467 15 g agar dissolved in 1 L of 70% IO), allowed the top agar to solidify at room

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

468 temperature and then incubated plates at 25°C. To check for possible bacteriophage, we
469 examined the plates after 24 hours for plaques, but did not detect any. We observed
470 plaques after 3-4 days. We picked plaques and made a lysate for each by placing a plaque
471 in 20 ml of 100% IO with 1.5 ml of a *Vibrio* #0024 overnight culture. We incubated the
472 lysates at 26°C and 200 rpm. After at least 24 hours of incubation, we used 1000x phase
473 contrast microscopy to check for small, highly motile cells. After detecting predatory
474 bacteria cells, we filtered the lysate through a 0.45 μm filter and repeated the double agar
475 overlay technique to obtain individual plaques on a lawn of *Vibrio* #0024. The double
476 agar overlay and plaque picking procedure was performed a total of three times to ensure
477 a pure isolate of predatory bacteria. The lysate made from the final plaque picking
478 procedure was filtered through a 0.45 μm filter. We combined 500 μl of this filtrate
479 (containing cells of the pure isolate of predatory bacteria) with 500 μl sterile 50%
480 glycerol and stored this stock at -80°C.
481
482 *Prey range tests*
483 To obtain active *H. marinus* BE01 for prey range tests, we added a small amount of the
484 -80°C stock to 15 ml of 100% IO mixed with 1 ml of an *E. coli* #0057 overnight culture.
485 We chose to use *E. coli* because prior work reported viable *Vibrio* passing through 0.45
486 μm filters as minicells [14], which could confound the results of the prey range tests. We
487 incubated the lysate at 26°C and 200 rpm. After three days, we filtered the lysate using a
488 0.45 μm filter to separate predatory bacteria from prey bacteria and cell debris. Swabs of
489 the filtrate on LB plates confirmed that no viable *E. coli* passed through the filter. We
490 performed 1:10 serial dilutions of the filtrate in 100% IO. To test prey range, we used the

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

491 double agar overlay method described above to observe plaque formation. We cultured
492 prey strains in 35 ml of SWYE broth for *Vibrio* prey strains or tryptic soy broth (TSB) for
493 all other prey strains. We centrifuged cultures at 6000 rpm for 10 minutes, washed pellets
494 in 100% IO and then resuspended pellets in 4 ml 100% IO. All prey resuspensions were
495 at least 10^8 cfu/ml. For the prey range tests, we plated the 10^{-3} to 10^{-6} dilutions of the
496 filtrate. We incubated plates at 26°C and checked for plaques daily starting on day three
497 until day seven. Plaque formation on any of these days was scored as positive for the prey
498 range test. We repeated this procedure twice for each prey strain to obtain three
499 biological replicates.
500
501 *Electron microscopy*
502 To obtain BE01 samples for electron microscopy, we added a small amount of the -80°C
503 stock of *Halobacteriovorax marinus* BE01 to 20 ml of 100% IO mixed with 1.5 ml of an
504 *E. coli* #0057 overnight culture. After 48 hours of incubation at 26°C and 200 rpm, we
505 placed formvar coated EM grids on 30 uL droplets of bacterial sample for 30 seconds to
506 allow the bacteria to adhere to the formvar surface. We then transferred grids to 50 uL
507 drops of 1% uranyl acetate in water for 1 minute. The grids were lifted from the drops of
508 uranyl acetate and the excess stain wicked off with Whatman filter paper. The stained
509 sample coated grids were air dried for 10 minutes and resulting specimens imaged with a
510 Jeol CX 2000 transmission electron microscope.
511
512 *Sequencing and assembly of Halobacteriovorax marinus BE01 genome*

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

513 To obtain genomic DNA for sequencing, we cultivated BE01 using *E. coli* #0057 as prey.
514 We chose to use *E. coli* because there is extensive genome information available which
515 would allow us to screen reads to remove prey bacteria reads if necessary. To make
516 lysates for genomic DNA preparation, we added a small amount of the -80°C stock of
517 BE01 to 20 ml of 100% IO mixed with 1.5 ml of an *E. coli* #0057 overnight culture.
518 After two days of incubation, we examined the lysates for active predatory bacteria cells.
519 We selected three lysates that appeared to have the highest ratio of predator to prey and
520 pooled these. Because PacBio technology requires at least 10 µg of genomic DNA for
521 library construction, we did not filter the lysates to avoid any potential loss of predatory
522 bacteria cells. To extract genomic DNA, we used the Wizard Genomic DNA Purification
523 Kit (Promega) with the protocol for Gram-positive and Gram-negative bacteria. We
524 centrifuged the pooled lysates at 9100 rpm for 10 minutes and resuspended the pellet in
525 600 µl of Nuclei Lysis Solution. We then continued with the manufacturer's instructions.
526 At the final step, we left the genomic DNA at 4°C overnight. By Qubit 2.0, the genomic
527 DNA was at 299 µg/ml.
528
529 Library construction and sequencing were performed at the Institute for Genome
530 Sciences at the University of Maryland Baltimore on a Pacific Biosciences RSII
531 instrument using P6-C4 chemistry. We launched an Amazon EC2 instance of SMRT
532 Portal 2.3.0 to analyze and assemble the data. Two SMRT cells generated 93,922 post-
533 filter polymerase reads (20,025 bp N50) and 151,636 subreads (10,161 bp N50). We
534 performed *de novo* assembly using the RS_HGAP_Assembly.3 protocol [45] with default
535 settings except for genome size, which we changed to 3.5 Mbp. This generated 93 contigs

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

536 in the polished assembly. The largest contig was 3,413,657 bp and aligned to
537 *Halobacteriovorax marinus* SJ by BLASTN. We used BLAST2Go [46] to align the other
538 92 contigs against the non-redundant database (restricted to Bacteria) with megablast to
539 determine if any of the smaller contigs aligned to *Halobacteriovorax*.
540
541 To close the large *Halobacteriovorax* contig, we used Gepard [47] to identify overlaps
542 between the ends of the contig, which indicated that the contig could be circularized. We
543 used BLASTN alignments to specifically determine the overlap regions, which resulted
544 in trimming 20,805 bp from the beginning of the contig. We then edited the trimmed
545 contig so that the first nucleotide corresponded to the first nucleotide of the *dnaA* protein-
546 coding sequence. To check the accuracy of the draft sequence at this stage, we aligned
547 the PacBio reads against this draft sequence using the RS_Resequencing.1 protocol in
548 SMRT Portal. The consensus sequence from this alignment had 542 differences
549 compared to the draft sequence used as a reference.
550
551 To polish the sequence, we generated 150 bp paired-end Illumina reads. Library
552 construction and sequencing (equivalent to 5% of a channel) were performed at IGS on
553 an Illumina HiSeq. We filtered the resulting reads so that every base in each read pair
554 was ≥Q25. This yielded 6,604,606 read pairs. We aligned these read pairs to the recalled
555 draft sequence using bwa-mem [48], yielding 507x average coverage (with a minimum of
556 51x). We used samtools [49] to convert the alignment to a sorted and indexed bam file.
557 Finally, we used Pilon [50] to identify corrections based on the Illumina data, which
558 amounted to 372 small insertions. The corrected sequence generated by Pilon was

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

559 deposited in GenBank as the complete chromosome of *Halobacteriovorax marinus*
560 BE01.
561
562 *Genome annotation and analysis of gene content*
563 We annotated the *H. marinus* BE01 genome initially using the NCBI Prokaryotic
564 Genome Annotation Pipeline version 3.3. Because of the unusually high proportion of
565 hypothetical proteins identified by PGAP (see Results), we submitted both BE01 and SJ
566 chromosome sequences to RAST [18–20] in January 2017. We used classic RAST with
567 the RAST gene caller and FIGfam Release70. We separately annotated RNA-coding
568 genes using Infernal 1.1.2 [21]. Files of the RAST+Infernal annotations and the output
569 from RAST bidirectional best hit analysis are available at the figshare repository
570 (https://figshare.com/projects/Supporting_data_for_Halobacteriovorax_BE01_paper/242
571 29). R code used to generate the synteny plot and the plot of amino acid identity for
572 bidirectional best hits is available at the figshare repository
573 (https://figshare.com/projects/Supporting_data_for_Halobacteriovorax_BE01_paper/242
574 29).
575
576 *Modal codon usage analysis*
577 To compare the modal codon usage frequencies, we used a freely available software
578 package downloaded from http://www.life.illinois.edu/gary/programs/codon_usage.html
579 [26].
580
581 **Additional files:**

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

582 Additional File 1

583 Quicktime Movie .MOV

584 1000x phase-contrast microscopy of *Halobacteriovorax marinus* BE01 cells attacking *E.*

585 *coli* ML35.

586

587 Additional File 2

588 Excel spreadsheet

589 Classification of bacterial isolates used in prey range tests based on analysis of 16S rRNA

590 gene sequences (>1000 bp) with three databases.

591

592 Additional File 3

593 PDF

594 Neighbor-joining phylogeny of 16S rRNA gene sequences of four *Vibrio* isolates used in

595 prey range tests, with *Enterovibrio norvegicus* (RDP accession LK391520) as outgroup.

596 Distance matrix was constructed using dnadist in Phylip with Jukes-Cantor model of

597 nucleotide substitution.

598

599 Additional File 4

600 Excel spreadsheet

601 For individual genes discussed in the text, the RAST locus tag is matched with the

602 corresponding PGAP locus tag to enable interested readers to quickly find relevant genes

603 in the GenBank annotations.

604

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

605 **Declarations**

606 **Ethics approval and consent to participate:** Not applicable

607 **Consent for publication:** Not applicable

608 **Availability of data and material:** The genome sequence generated and analyzed during
609 the current study is available as BioProject PRJNA343955, BioSample SAMN05806433
610 and GenBank accession CP017414. Other data and R code are available at the figshare
611 repository
612 (https://figshare.com/projects/Supporting_data_for_Halobacteriovorax_BE01_paper/242
613 29).

614 **Competing interests:** The authors declare that they have no competing interests.

615 **Funding:** This research was supported by an Institutional Development Award (IDeA)
616 from the National Institute of General Medical Sciences of the National Institutes of
617 Health under grant number P20GM103430 and funding from Providence College.
618 Neither funder played a role in study design, data analysis or interpretation, or writing the
619 manuscript.

620 **Authors' contributions:** BGE designed and performed experiments, analyzed data and
621 wrote the manuscript. MKA designed and performed experiments. JAG performed
622 negative staining and EM imaging. LEW designed experiments, analyzed data and wrote
623 and edited the manuscript. All authors read and approved the final manuscript.

624 **Acknowledgments:** We thank Nicole Cullen for isolating *Pseudomonas* #0042 and Sean
625 O'Donnell for isolating *Acinetobacter* #0036. We thank Lisa Sadzewicz and Luke Tallon
626 at the Institute for Genome Sciences at the University of Maryland Baltimore for
627 sequencing services. We are grateful to Mark Martin for providing *E. coli* ML35, Brett

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

628 Pellock for providing *E. coli* #0057 and Cameron Thrash for guidance on growth media.

629 LEW thanks science Twitter for much useful advice and support.

630 **Authors' information:** BGE and MKA conducted this research as undergraduate

631 researchers at Providence College.

632

633 **References**

634 1. Jurkevitch E, Davidov Y. Phylogenetic diversity and evolution of predatory
635 prokaryotes. In: Predatory Prokaryotes. Berlin: Springer-Verlag; 2006. p. 11–56.

636 2. Shatzkes K, Singleton E, Tang C, Zuena M, Shukla S, Gupta S, et al. Predatory
637 Bacteria Attenuate Klebsiella pneumoniae Burden in Rat Lungs. mBio. 2016;7:e01847-
638 16.

639 3. Willis AR, Moore C, Mazon-Moya M, Krokowski S, Lambert C, Till R, et al.
640 Injections of Predatory Bacteria Work Alongside Host Immune Cells to Treat Shigella
641 Infection in Zebrafish Larvae. Curr Biol. 2016;26:3343–51.

642 4. Sockett RE. Predatory Lifestyle of Bdellovibrio bacteriovorus. Annu Rev Microbiol.
643 2009;63:523–39.

644 5. Crossman LC, Chen H, Cerdeño-Tárraga A-M, Brooks K, Quail MA, Pineiro SA, et al.
645 A small predatory core genome in the divergent marine Bacteriovorax marinus SJ and the
646 terrestrial Bdellovibrio bacteriovorus. ISME J. 2013;7:148–60.

647 6. Baer ML, Ravel J, Piñeiro SA, Guether-Borg D, Williams HN. Reclassification of salt-
648 water Bdellovibrio sp. as Bacteriovorax marinus sp. nov. and Bacteriovorax litoralis sp.
649 nov. Int J Syst Evol Microbiol. 2004;54:1011–6.

650 7. Koval SF, Williams HN, Stine OC. Reclassification of Bacteriovorax marinus as
651 Halobacteriovorax marinus gen. nov., comb. nov. and Bacteriovorax litoralis as
652 Halobacteriovorax litoralis comb. nov.; description of Halobacteriovoraceae fam. nov. in
653 the class Deltaproteobacteria. Int J Syst Evol Microbiol. 2015;65:593–7.

654 8. Pineiro SA, Stine OC, Chauhan A, Steyert SR, Smith R, Williams HN. Global survey
655 of diversity among environmental saltwater Bacteriovoracaceae. Environ Microbiol.
656 2007;9:2441–50.

657 9. Piñeiro SA, Williams HN, Stine OC. Phylogenetic relationships amongst the saltwater
658 members of the genus Bacteriovorax using rpoB sequences and reclassification of
659 Bacteriovorax stolpii as Bacteriolyticum stolpii gen. nov., comb. nov. Int J Syst Evol
660 Microbiol. 2008;58:1203–9.

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

661 10. Williams HN, Lymperopoulou DS, Athar R, Chauhan A, Dickerson TL, Chen H, et
662 al. Halobacteriovorax, an underestimated predator on bacteria: potential impact relative to
663 viruses on bacterial mortality. ISME J. 2016;10:491–9.

664 11. Welsh RM, Zaneveld JR, Rosales SM, Payet JP, Burkepile DE, Thurber RV.
665 Bacterial predation in a marine host-associated microbiome. ISME J. 2016;10:1540–4.

666 12. Pineiro SA, Sahaniuk GE, Romberg E, Williams HN. Predation Pattern and
667 Phylogenetic Analysis of Bdellovibrionaceae from the Great Salt Lake, Utah. Curr
668 Microbiol. 2004;48:113–7.

669 13. Schoeffield AJ, Williams HN. Efficiencies of Recovery of Bdellovibrios from
670 Brackish- Water Environments by Using Various Bacterial Species as Prey. Appl
671 Environ Microbiol. 1990;56:230–6.

672 14. Richards GP, Fay JP, Uknalis J, Olanya OM, Watson MA. Purification and Host
673 Specificity of Predatory Halobacteriovorax Isolates from Seawater. Appl Environ
674 Microbiol. 2016;82:922–7.

675 15. Chen H, Brinkac LM, Mishra P, Li N, Lymperopoulou DS, Dickerson TL, et al. Draft
676 genome sequences for the obligate bacterial predators Bacteriovorax spp. of four
677 phylogenetic clusters. Stand Genomic Sci. 2015;10:11.

678 16. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. JSpeciesWS: a web server
679 for prokaryotic species circumscription based on pairwise genome comparison.
680 Bioinformatics. 2016;32:929–31.

681 17. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM.
682 DNA-DNA hybridization values and their relationship to whole-genome sequence
683 similarities. Int J Syst Evol Microbiol. 2007;57 Pt 1:81–91.

684 18. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST
685 Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.

686 19. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and
687 the Rapid Annotation of microbial genomes using Subsystems Technology (RAST).
688 Nucleic Acids Res. 2014;42 Database issue:D206-214.

689 20. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, et al. RASTtk: a
690 modular and extensible implementation of the RAST algorithm for building custom
691 annotation pipelines and annotating batches of genomes. Sci Rep. 2015;5:8365.

692 21. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches.
693 Bioinformatics. 2013;29:2933–5.

694 22. Wall DP, Deluca T. Ortholog detection using the reciprocal smallest distance
695 algorithm. Methods Mol Biol Clifton NJ. 2007;396:95–110.

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

696 23. Wang L, Chen S, Vergin KL, Giovannoni SJ, Chan SW, DeMott MS, et al. DNA
697 phosphorothioation is widespread and quantized in bacterial genomes. Proc Natl Acad
698 Sci. 2011;108:2963–8.

699 24. Ho WS, Ou H-Y, Yeo CC, Thong KL. The dnd operon for DNA phosphorothioation
700 modification system in Escherichia coli is located in diverse genomic islands. BMC
701 Genomics. 2015;16:199.

702 25. Friedman R, Ely B. Codon usage methods for horizontal gene transfer detection
703 generate an abundance of false positive and false negative results. Curr Microbiol.
704 2012;65:639–42.

705 26. Davis JJ, Olsen GJ. Modal codon usage: assessing the typical codon usage of a
706 genome. Mol Biol Evol. 2010;27:800–10.

707 27. Snyder AR, Williams HN, Baer ML, Walker KE, Stine OC. 16S rDNA sequence
708 analysis of environmental Bdellovibrio-and-like organisms (BALO) reveals extensive
709 diversity. Int J Syst Evol Microbiol. 2002;52:2089–94.

710 28. He X, Ou H-Y, Yu Q, Zhou X, Wu J, Liang J, et al. Analysis of a genomic island
711 housing genes for DNA S-modification system in Streptomyces lividans 66 and its
712 counterparts in other distantly related bacteria. Mol Microbiol. 2007;65:1034–48.

713 29. Wang L, Chen S, Xu T, Taghizadeh K, Wishnok JS, Zhou X, et al.
714 Phosphorothioation of DNA in bacteria by dnd genes. Nat Chem Biol. 2007;3:709–10.

715 30. Liu G, Ou H-Y, Wang T, Li L, Tan H, Zhou X, et al. Cleavage of Phosphorothioated
716 DNA and Methylated DNA by the Type IV Restriction Endonuclease ScoMcrA. PLOS
717 Genet. 2010;6:e1001253.

718 31. Gophna U, Charlebois RL, Doolittle WF. Ancient lateral gene transfer in the
719 evolution of Bdellovibrio bacteriovorus. Trends Microbiol. 2006;14:64–9.

720 32. Pan A, Chanda I, Chakrabarti J. Analysis of the genome and proteome composition of
721 Bdellovibrio bacteriovorus: indication for recent prey-derived horizontal gene transfer.
722 Genomics. 2011;98:213–22.

723 33. Hobley L, Lerner TR, Williams LE, Lambert C, Till R, Milner DS, et al. Genome
724 analysis of a simultaneously predatory and prey-independent, novel Bdellovibrio
725 bacteriovorus from the River Tiber, supports in silico predictions of both ancient and
726 recent lateral gene transfer from diverse bacteria. BMC Genomics. 2012;13:670.

727 34. Chauhan A, Cherrier J, Williams HN. Impact of sideways and bottom-up control
728 factors on bacterial community succession over a tidal cycle. Proc Natl Acad Sci.
729 2009;106:4301–6.

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

730 35. Richards GP, Fay JP, Dickens KA, Parent MA, Soroka DS, Boyd EF. Predatory
731 Bacteria as Natural Modulators of Vibrio parahaemolyticus and Vibrio vulnificus in
732 Seawater and Oysters. Appl Environ Microbiol. 2012;78:7455–66.

733 36. Jurkevitch E. Isolation and Classification of Bdellovibrio and Like Organisms. In:
734 Current Protocols in Microbiology. John Wiley & Sons, Inc.; 2005.
735 doi:10.1002/9780471729259.mc07b01s26.

736 37. Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ, et al. Design and
737 Evaluation of Useful Bacterium-Specific PCR Primers That Amplify Genes Coding for
738 Bacterial 16S rRNA. Appl Environ Microbiol. 1998;64:795–9.

739 38. Heuer H, Krsek M, Baker P, Smalla K, Wellington EM. Analysis of actinomycete
740 communities by specific amplification of genes encoding 16S rRNA and gel-
741 electrophoretic separation in denaturing gradients. Appl Environ Microbiol.
742 1997;63:3233–41.

743 39. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error
744 probabilities. Genome Res. 1998;8:186–94.

745 40. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces
746 using phred. I. Accuracy assessment. Genome Res. 1998;8:175–85.

747 41. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing.
748 Genome Res. 1998;8:195–202.

749 42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search
750 tool. J Mol Biol. 1990;215:403–10.

751 43. Pruesse E, Peplies J, Glöckner FO. SINA: Accurate high-throughput multiple
752 sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823–9.

753 44. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid
754 assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ
755 Microbiol. 2007;73:5261–7.

756 45. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al.
757 Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing
758 data. Nat Methods. 2013;10:563–9.

759 46. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a
760 universal tool for annotation, visualization and analysis in functional genomics research.
761 Bioinformatics. 2005;21:3674–6.

762 47. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating
763 dotplots on genome scale. Bioinformatics. 2007;23:1026–8.

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

764 48. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler
765 transform. Bioinformatics. 2009;25:1754–60.

766 49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
767 Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

768 50. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An
769 Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly
770 Improvement. PLOS ONE. 2014;9:e112963.

771

772 **Figure legends**

773 Figure 1. *Halobacteriovorax marinus* BE01 microscopy and plaque formation. (a) 1000x
774 phase-contrast microscopy of small, comma-shaped BE01 cells (arrows) and larger *E.*
775 *coli* ML35 cells. (b) BE01 cells stained with uranyl acetate and imaged with electron
776 microscopy. Scale bar is 500 nm. (c) Plaques formed by BE01 on a lawn of *Vibrio* using
777 double agar overlay.

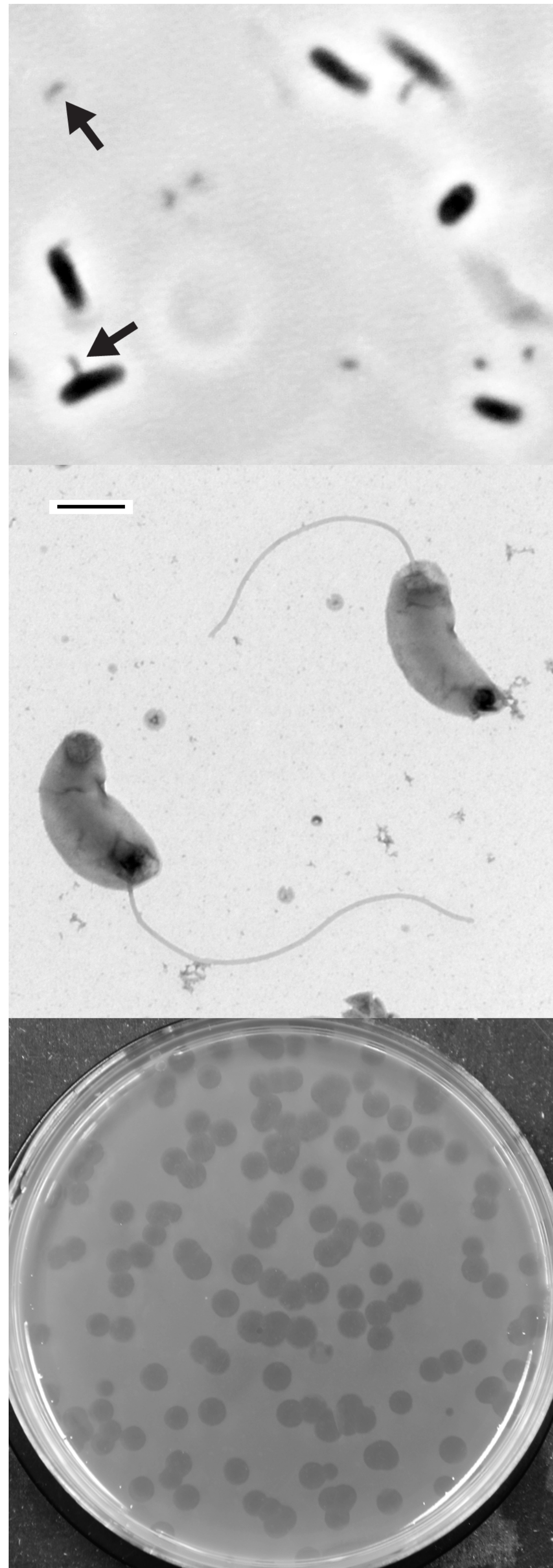
778

779 Figure 2. Synteny plot of bidirectional best hits between *H. marinus* BE01 and SJ.
780 Bidirectional best hits identified by RAST are plotted based on their gene number on
781 each chromosome. Individual genes are denoted with symbols corresponding to the
782 similarity between BE01 and SJ amino acid sequences. (a) shows the entire
783 chromosomes, whereas (b) highlights the two major regions of difference in gene content
784 (labeled A and B).

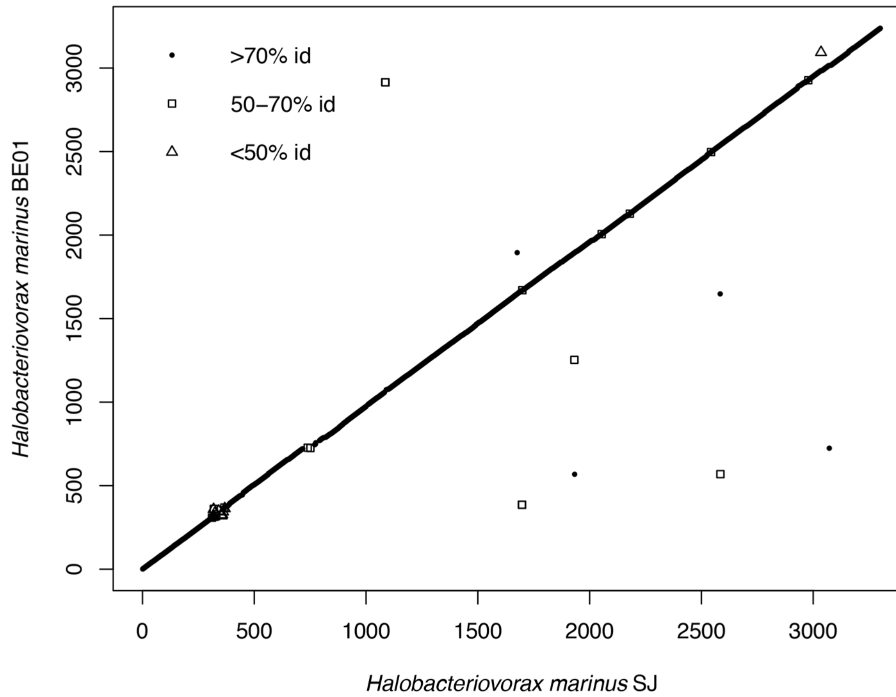
785

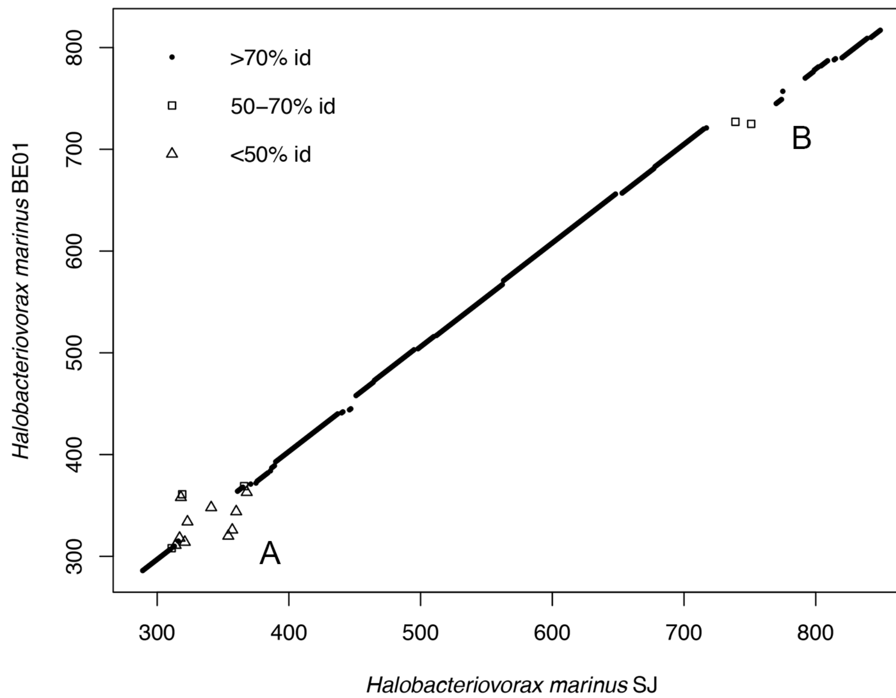
786 Figure 3. Amino acid identity of bidirectional best hits. Bidirectional best hits identified
787 by RAST are plotted based on their position on the SJ chromosome and the similarity
788 between BE01 and SJ amino acid sequences. Horizontal lines indicate 100%, 98% and
789 96% amino acid identity.

bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.





bioRxiv preprint doi: https://doi.org/10.1101/180265; this version posted August 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.