Variation in genome content and predatory phenotypes between *Bdellovibrio* sp. NC01 isolated from soil and *B. bacteriovorus* type strain HD100

Laura E. Williams[a]#, Nicole Cullen[a], Joseph A. DeGiorgis[a,b], Karla J. Martinez[a], Justina Mellone[a], Molly Oser[a], Jing Wang[c], Ying Zhang[c]

[a]Department of Biology, Providence College, Providence, RI, USA

[b]Cellular Dynamics Program, Marine Biological Laboratory, Woods Hole, MA, USA

[c]Department of Cell and Molecular Biology, College of the Environment and Life Sciences, University of Rhode Island, Kingston, RI, USA

#Corresponding author: Laura E. Williams, lwillia7@providence.edu

Keywords: predatory bacteria, prey range, predation efficiency, plaque phenotype

Genome accession number: CP030034

Abstract

Defining phenotypic and associated genotypic variation among *Bdellovibrio* may further our understanding of how this genus attacks and kills different Gram-negative bacteria. We isolated *Bdellovibrio* sp. NC01 from soil. Analysis of 16S rRNA gene sequences and average amino acid identity showed that NC01 belongs to a different species than the type species *bacteriovorus*. By clustering amino acid sequences from completely sequenced *Bdellovibrio* and comparing the resulting ortholog groups to a previously published analysis, we defined a "core genome" of 778 protein-coding genes and identified four protein-coding genes that appeared to be missing only in NC01. To determine how horizontal gene transfer (HGT) may have impacted NC01 genome evolution, we performed genome-wide comparisons of *Bdellovibrio* nucleotide sequences, which indicated that eight NC01 genomic regions were likely acquired by HGT. To investigate how genome variation may impact predation, we compared protein-coding gene content between NC01 and the *B. bacteriovorus* type strain HD100, focusing on genes implicated as important in successful killing of prey. Of these, NC01 is missing ten genes that may play roles in lytic activity during predation. Compared to HD100, NC01 kills fewer tested prey strains and kills *E. coli* ML35 less efficiently. NC01 causes a smaller log reduction in ML35, after which the prey population recovers and the NC01 population decreases. In addition, NC01 forms turbid plaques on lawns of *E. coli* ML35, in contrast to clear plaques formed by HD100. Linking phenotypic variation in interactions between *Bdellovibrio* and Gram-negative bacteria with underlying *Bdellovibrio* genome variation is valuable for understanding the ecological significance of predatory bacteria and evaluating their effectiveness in clinical applications.

35 **Introduction**

36 Since their discovery in the early 1960s, predatory bacteria have been isolated from a
37 wide range of terrestrial and aquatic environments, including some with extreme conditions such
38 as high temperature or salinity (1). The impact of predatory bacteria in these environments is not
39 fully known, but recent studies indicate they may play an important role in shaping microbial
40 communities (2–4). In addition, their ability to attack and kill Gram-negative bacterial pathogens
41 has generated interest in developing predatory bacteria as biocontrol agents to combat drug-
42 resistant infections (5–7). Given their potential ecological significance and possible clinical
43 applications, it is important to define phenotypic and genotypic variation among predatory
44 bacteria. In particular, variation in predatory phenotypes such as prey range and predation
45 efficiency will affect how different strains of predatory bacteria impact microbial communities,
46 and this phenotypic variation may be determined by variation in genome content. To further our
47 understanding of phenotypic and genome variation in predatory bacteria, we isolated, assayed
48 and sequenced a *Bdellovibrio* strain from soil and compared its predatory phenotypes and
49 genome content to other sequenced *Bdellovibrio*.

50 Among the different genera of predatory bacteria, *Bdellovibrio* is the most well-studied.
51 Currently, this genus includes two species that differ in their predation strategy. *B. bacteriovorus*
52 invades the periplasm of Gram-negative prey cells (8), whereas *B. exovorus* attaches to the
53 outside of prey cells but does not enter the cell itself (9). Here, we focus on intraperiplasmic
54 *Bdellovibrio*. This lifecycle includes two main phases: the attack phase and the growth phase
55 (10). During the attack phase, small, highly motile *Bdellovibrio* cells seek out and attach to
56 Gram-negative prey bacteria. *Bdellovibrio* then invades the prey periplasm by using lytic
57 enzymes to open a small hole in the prey outer membrane and cell wall (11). Both attachment

58 and invasion may be mediated by Type IV pili expressed by *Bdellovibrio* (12–14). After
59 *Bdellovibrio* moves into the prey periplasm, it uses peptidoglycan-modifying enzymes to reseal
60 and reshape the prey cell wall, resulting in formation of a rounded structure called a bdelloplast
61 (15). Within the bdelloplast, *Bdellovibrio* initiates the growth phase, during which it secretes
62 lytic enzymes into the prey cytoplasm. Using digested prey cell contents as nutrients,
63 *Bdellovibrio* grows into a single long filament within the prey periplasm, eventually dividing into
64 multiple progeny, which lyse the bdelloplast and emerge as new attack phase cells.

65 By invading and lysing Gram-negative bacteria, some strains of *Bdellovibrio* can reduce
66 prey populations by up to 8 log units over 48 hours *in vitro* (6, 16). However, not all *Bdellovibrio*
67 strains are equally effective at killing the same Gram-negative species. Researchers have
68 observed variation among *Bdellovibrio* strains in which Gram-negative bacteria are susceptible
69 to predation (17, 18) and in how quickly and to what extent a population of Gram-negative prey
70 is reduced (16, 19). These predatory phenotypes, referred to as prey range and predation
71 efficiency, are clearly important in determining the outcomes of interactions between
72 *Bdellovibrio* and Gram-negative bacteria; however, the extent of naturally occurring variation in
73 these phenotypes and the underlying mechanisms governing them are not well understood.
74 Defining how predatory phenotypes differ among *Bdellovibrio* will contribute to our
75 understanding of the impact of predatory bacteria on microbial communities in different
76 environments and inform their application in biocontrol of bacterial pathogens.

77 Here, we describe the isolation of *Bdellovibrio* sp. NC01 from soil and compare its
78 genome and predatory phenotypes to other *Bdellovibrio*, focusing primarily on the *B.*
79 *bacteriovorus* type strain HD100. The intraperiplasmic predator HD100 was the first
80 *Bdellovibrio* strain sequenced (20), and multiple studies have dissected some of the molecular

81 mechanisms governing predation in this strain (21). To investigate variation in genome content,
82 we compared the genome of NC01 to that of HD100 and other sequenced *Bdellovibrio*.
83 Currently, only a few completely sequenced predatory bacteria genomes are available in public
84 databases. In addition to HD100, there are only four other completely sequenced intraperiplasmic
85 *Bdellovibrio* genomes, including *B. bacteriovorus* W (22), which is the most phylogenetically
86 divergent from HD100 (23). The limited genome dataset highlights the importance of sequencing
87 additional *Bdellovibrio*, especially strains that are divergent from HD100. To define variation in
88 predatory phenotypes between NC01 and HD100, we determined prey range differences by
89 testing each strain against a panel of Gram-negative bacteria, and we measured predation
90 efficiency by quantifying viable *E. coli* ML35 during co-culture with each strain.

91 **Methods**

92 **Isolation and identification of environmental bacteria for use as prey**

93 We isolated bacteria from an urban freshwater stream in Providence, RI (lat 41.835, lon -
94 71.44299) and from soil in a residential area of Massachusetts (lat 42.0877, lon -71.2309) for use
95 as prey. We also obtained *E. coli* ML35 from Mark Martin (University of Puget Sound, Puget
96 Sound, WA) and *Escherichia* 0057 from Brett Pellock (Providence College, Providence, RI).
97 To identify these isolates, we sequenced the 16S rRNA gene following a previously
98 described procedure (24). Briefly, we used primers 63F (25) and 1378R (26), which amplify the
99 16S rRNA gene from many members of Domain Bacteria. Sanger sequencing of purified PCR
100 products was performed by GeneWiz (South Plainfield, NJ) using the same primers. We trimmed
101 and assembled reads using Phred/Phrap/Consed (27–29), then classified the assembled 16S
102 rRNA gene sequences using BLAST (30), the SILVA Incremental Aligner (31) and the
103 Ribosomal Database Project classifier (32). Supplementary Table 1 shows the full classification

104 results. The RDP classifier and the SILVA Incremental Aligner classify sequences to the genus
105 level, but not the species level; therefore, species names are not provided for these isolates.

106 **Isolation and identification of *Bdellovibrio* sp. NC01 from bioswale soil**

107 In September 2015, we collected 50 g of soil from a bioswale on the Providence College
108 campus (lat 41.84277, lon -71.43944). To isolate predatory bacteria from the soil sample, we
109 adapted standard methods (33, 34). We combined the soil sample with 500 ml of sterile HM
110 buffer (25 mM HEPES adjusted to pH 7.4 and supplemented with 3 mM calcium chloride
111 dihydrate and 2 mM magnesium chloride hexahydrate). We then filtered this mixture through
112 sterile cheesecloth into a sterile container to remove large particulate matter. We vortexed the
113 filtrate to release bacterial cells from the small particles of soil that passed through the
114 cheesecloth. To pellet the soil particles, we centrifuged the filtrate at 500 x g (1753 rpm in a
115 Sorvall GSA rotor) for 10 minutes, after which we retained the supernatant containing suspended
116 bacterial cells.

117 To enrich for predatory bacteria in the soil sample supernatant, we used *Pseudomonas* sp.
118 NC02 as prey, because it was also isolated from a soil environment, and its genome has been
119 sequenced (35). We cultured *Pseudomonas* sp. NC02 overnight in 120 ml tryptic soy broth
120 (TSB) at 28°C and 200 rpm, then centrifuged the culture at 5854 x g (6000 rpm in a Sorvall GSA
121 rotor) for 10 minutes, washed the resulting pellet with 100 ml sterile HM buffer, centrifuged
122 again and resuspended the resulting pellet in 2.5 ml HM buffer. We combined 1 ml of
123 *Pseudomonas* sp. NC02 resuspension with 20 ml of soil sample supernatant and incubated this
124 enrichment at 28°C and 200 rpm.

125 Once we observed small, fast-moving cells under 1000x phase-contrast microscopy, we
126 filtered the enrichment using a 0.45 μm filter and performed 10-fold serial dilutions of the filtrate

127 in sterile HM buffer. Dilutions were then plated using a double agar overlay method. For each
128 dilution of the filtrate to be plated, we combined 100 μl of the dilution, 250 μl of *Pseudomonas*
129 sp. NC02 at 10^8 cfu/ml and 4.5 ml of molten HM top agar (0.6% agar). After vortexing, we
130 poured this mixture onto HM agar plates (1.5% agar). We allowed the top agar to solidify at
131 room temperature, then incubated plates at 28°C until we observed plaques in the lawn of
132 *Pseudomonas* sp. NC02. To begin the process of obtaining a pure isolate, we used a sterile
133 Pasteur pipette to extract single, well-defined plaques from the top agar. We combined each
134 plaque with 500 μl of sterile HM buffer, incubated the plaque suspension at room temperature
135 for at least 5 minutes, then vortexed it vigorously.

136 To culture predatory bacteria from the plaque suspensions, we made lysates combining
137 300 μl of each plaque suspension, 1.5 ml of an overnight culture of *Serratia* 0043 and 20 ml
138 sterile HM buffer. *Serratia* 0043 was also isolated from a soil environment, and switching to this
139 prey species during the isolation procedure selected for predatory isolates that were capable of
140 attacking both *Pseudomonas* and *Serratia*, which belong to different orders. We incubated
141 lysates at 28°C and 200 rpm. After 48 h, one lysate showed small, fast-moving cells, and we
142 used this lysate to begin two additional rounds of serial dilution, plating and plaque picking to
143 ensure a pure isolate. We used *Serratia* 0043 as prey during both of these rounds. After the final
144 round, we filtered the plaque suspension using a 0.45 μm filter, then combined 500 μl of filtrate
145 with 500 μl of sterile 50% glycerol and stored the stock at -80°C.

146 To identify NC01, we extracted the 16S rRNA gene sequence from the NC01 genome
147 based on the annotation generated by PGAP (see below for genome sequencing and annotation
148 methods). Using SINA 1.2.11 (31), we aligned this sequence against 16S rRNA gene sequences
149 extracted from nine genomes of obligate predatory bacteria belonging to *Bacteriovorax*,

150 *Bdellovibrio* and *Halobacteriovorax* (*Bacteriovorax stolpii* DSM12778 CP025704, *Bdellovibrio*
151 *bacteriovorus* 109J CP007656, *Bdellovibrio bacteriovorus* HD100 NC_005363, *Bdellovibrio*
152 *bacteriovorus* SSB218315 CP020946, *Bdellovibrio bacteriovorus* Tiberius NC_019567,
153 *Bdellovibrio bacteriovorus* W NZ_CP002190, *Bdellovibrio exovorus* JSS NC_020813,
154 *Halobacteriovorax marinus* BE01 CP017414, *Halobacteriovorax marinus* SJ NC_016620). We
155 then inferred a phylogenetic tree from the alignment with the online server RAxML Blackbox
156 (36), using 100 bootstrap replicates to estimate the confidence of the topology.

157 To further investigate the relationships among NC01 and other *Bdellovibrio*, we used an
158 average amino acid identity (AAI) calculator (http://enve-omics.ce.gatech.edu/) (37) to analyze
159 the similarity among protein-coding sequences from NC01, 109J, HD100 and W, as annotated by
160 RAST (see below for genome sequencing and annotation methods).

161 **Negative stain electron microscopy of *Bdellovibrio* sp. NC01**

162 To obtain samples of *Bdellovibrio* sp. NC01 for negative stain electron microscopy, we
163 added a small amount of the -80°C freezer stock to 20 ml of HM buffer mixed with 1.5 ml of an
164 overnight culture of *E. coli* ML35 grown in TSB. After 48 h of incubation at 28°C and 200 rpm,
165 we stained samples with 1% uranyl acetate (pH 4.2) following a previously described procedure
166 (24) and imaged the resulting specimens using a JEOL CX 2000 transmission electron
167 microscope.

168 **Genome sequencing and annotation of *Bdellovibrio* sp. NC01**

169 To obtain genomic DNA, we combined a small amount of the -80°C stock, 1.5 ml of an
170 overnight culture of *E. coli* ML35 and 25 ml of sterile HM buffer. We chose ML35 as prey
171 because it is a standard prey strain used in cultivation of *Bdellovibrio*, and the genome sequence
172 is available for screening reads if needed (38). After 72 hours of incubation at 28°C and 200 rpm,

173 we centrifuged the lysate at 10000 x g (9100 rpm in a Thermo Scientific A27 rotor) for 10
174 minutes to pellet bacterial cells. Because PacBio sequencing technology requires at least 10 μg
175 of genomic DNA, we did not filter the lysate to prevent loss of predatory bacteria cells. We then
176 used the Wizard Genomic DNA purification kit to extract genomic DNA, following the
177 manufacturer's instructions beginning with resuspension of the pellet in 600 μl of the kit's nuclei
178 lysis solution. For the final rehydration step, we rehydrated the DNA pellet overnight at 4°C.
179 Quantification by Qubit estimated a final concentration of 340 ng/μl genomic DNA.

180 For long-read sequencing, library construction and sequencing were performed at the
181 Institute for Genome Sciences at the University of Maryland Baltimore on a Pacific Biosciences
182 RSII instrument using P6-C4 chemistry. Quality control analysis of the reads showed that <10%
183 of reads aligned by BLAST to *E. coli* or *Shigella*, indicating that representation of the prey
184 genome in the read dataset is low. To analyze and assemble the reads, we launched an instance
185 of SMRT Portal 2.3.0 using Amazon EC2. A single SMRT cell yielded 79,088 postfilter
186 polymerase reads (N_{50} 17,028) and 94,031 subreads (N_{50} 13,572). Within SMRT Portal, we
187 assembled the reads with HGAP3 (39), using 4.5 Mbp for the genome size. This generated 186
188 contigs. The largest contig was 4,001,137 bp and aligned to *Bdellovibrio* in the database using
189 BLASTN. The other 185 contigs were each <48,000 bp in length, and all of them aligned to *E.*
190 *coli* by BLASTN, which is expected due to the inclusion of prey genomic DNA in the
191 sequencing library.

192 To close the genome, we used BLASTN to identify overlap in the ends of the contig and
193 trimmed the sequence to generate a closed contig of 3,975,771 bp. We also adjusted the start site
194 so that the first nucleotide of the closed genome sequence corresponds to the first nucleotide in
195 the start codon of *dnaA*. To polish this sequence, we used 250 bp paired-end Illumina MiSeq data

196 generated at the Rhode Island Genomics and Sequencing Center. We filtered and trimmed the
197 raw read dataset to retain only those read pairs for which the quality score of every base was ≥13
198 and the read length was ≥65 nt. This yielded 1,256,135 read pairs. Using bwa-mem (40),
199 samtools (41) and Pilon (42), we aligned these reads to the closed genome sequence and
200 corrected 1 SNP, 84 small insertions totaling 85 bases and 4 small single-base deletions. We
201 manually examined changes flagged by Pilon using the alignment viewer Tablet (43), generating
202 a final, polished *Bdellovibrio* sp. NC01 genome sequence of 3,975,853 bp. This sequence has
203 been deposited in GenBank as CP030034.

204 In addition to annotation with NCBI's Prokaryotic Genome Annotation Pipeline (PGAP)
205 version 4.5, we also annotated the sequence using ClassicRAST (44, 45), Infernal (46), and
206 tRNAScan-SE (47, 48).

207 **Comparative genomics of *Bdellovibrio* sp. NC01 and other sequenced *Bdellovibrio***

208 We used a custom software pipeline (49) to cluster amino acid sequences from
209 *Bdellovibrio* sp. NC01, all five completely sequenced *B. bacteriovorus* genomes, and *B.*
210 *exovorus* JSS into ortholog groups. NCBI accession numbers for these genomes are listed above.
211 Ortholog groups were identified among all genomes based on the identification of bi-directional
212 best hits in the protein-coding genes using BLASTP for the pairwise sequence comparisons. The
213 pipeline requires an alignment of two amino acid sequences to have ≥30% identity, ≥70%
214 coverage, and an e-value of <0.001. To define the "core genome" of the seven *Bdellovibrio*
215 analyzed here, we identified the set of ortholog groups that each included at least one protein-
216 coding gene from each genome.

217 We compared the core genome defined by our analysis with the set of 795 conserved
218 *Bdellovibrio* genes identified in (23). To do this, we extracted NCBI protein accession numbers

219 associated with *B. exovorus* JSS from Additional File 4a, because JSS was the only genome for
220 which all 795 genes were listed. We then used R and bash scripts to determine whether these
221 accession numbers appeared in the ortholog groups identified as the core genome in our analysis.
222 To define the set of HD100 protein-coding genes implicated in predation by previous
223 studies, we extracted the HD100 locus tags listed in Supporting Information Figure S1 of (50)
224 and all HD100 locus tags with a value <0.5 in the column labelled Mini Tn-seq ECPL avg *W* in
225 Data Set S4 of (51). For each of these HD100 locus tags, we used R and bash scripts to
226 determine the corresponding GenBank protein ID (as listed in the NC_005363 annotation on
227 GenBank) and identify the ortholog group in our analysis that included this protein ID. We then
228 determined whether the ortholog group also included a NC01 protein-coding gene. If not, we
229 considered the gene missing in NC01.
230 For genome-wide pairwise comparisons of nucleotide sequences, we used the BLAST
231 Ring Image Generator (BRIG) (52).
232 **Phenotype assays to test prey range differences and predation efficiency**
233 We obtained *Bdellovibrio bacteriovorus* HD100 from Mark Martin (University of Puget
234 Sound, Puget Sound, WA) for predatory phenotype comparisons. To conduct phenotype assays,
235 we adapted standard methods (33, 34, 53). To culture *Bdellovibrio* NC01 and HD100, we added
236 a small amount of -80°C stocks to 20 ml HM buffer combined with 1.5 ml of an overnight
237 culture of *E. coli* ML35. After 72 hours of incubation at 28°C and 200 rpm, we filtered the lysate
238 using a 0.45 μm filter to separate predatory cells from prey cells and cell debris. For each
239 phenotype assay described below, we followed this protocol to obtain attack phase *Bdellovibrio*
240 cells.

241 To test prey range differences and plaque appearance, we plated 10-fold serial dilutions
242 of filtered lysate using a double agar overlay method as described above for the isolation
243 procedure. Each tested prey strain was grown in 120 ml TSB overnight at 28°C and 200 rpm,
244 with the exception of *Escherichia* strains, which were grown at 37°C. After centrifugation and
245 washing, we resuspended prey cell pellets in 5 ml of sterile HM buffer. Based on cfu/ml counts,
246 this yielded concentrations of at least 10^8 cfu/ml, which ensured a dense prey lawn for
247 visualizing plaques. For some prey strains, we added 500 µl of prey resuspension to top agar to
248 improve our ability to detect and observe plaques in the prey lawn. We incubated plates at 28°C
249 for 8 days before scoring plaque formation and performed three replicates for each prey strain. In
250 observations of plaque appearance, we photographed plates at multiple time points over the
251 course of 8 days.
252 To test predation efficiency, we adapted a protocol described in (19). We centrifuged
253 filtered lysate at 8635 x g (8500 rpm in a Thermo Scientific A27 rotor) for 10 minutes. The
254 resulting pellet of predatory bacteria was resuspended in sterile HM buffer to yield ~10^8 cells/ml
255 based on direct cell counts using a Petroff-Hausser counting chamber and 1000x phase-contrast
256 microscopy. To initiate the predation efficiency assay, we combined 500 µl of the *Bdellovibrio*
257 cell resuspension with 12.5 ml of an overnight culture of *E. coli* ML35 diluted to OD_{600} 0.15
258 (range: 0.135 – 0.165) in sterile HM buffer. Dilution of the overnight culture yielded ~10^8 *E. coli*
259 cells/ml based on cfu/ml counts. As a control, we used 500 µl of sterile HM buffer in place of
260 *Bdellovibrio* cell resuspension. Using cfu/ml counts, we quantified viable *E. coli* ML35 at the
261 beginning of the assay (time zero) and then after 6, 24, 48 and 72 hours of incubation at 28°C
262 and 200 rpm. Based on the initial concentrations of *Bdellovibrio* and *E. coli* as estimated by

263 direct cell counts and cfu/ml counts, the predator:prey ratio at the beginning of the assay was 1:5
264 for NC01 and 1:4 for HD100.

265 To test survival of *Bdellovibrio* in the absence of prey, we used the same protocol as the
266 predation efficiency assay, but we combined 500 µl of *Bdellovibrio* cell resuspension with 12.5
267 ml of sterile HM buffer instead of a dilution of prey overnight culture. After 0, 24, 48 and 72
268 hours of incubation at 28°C and 200 rpm, we quantified viable *Bdellovibrio* by plating serial
269 dilutions of flask contents using the double agar overlay method described above for testing prey
270 range differences. After incubating these plates for 48 hours at 28°C, we counted plaques formed
271 on the *E. coli* ML35 prey lawn and then obtained a final plaque count after another 24 hours of
272 incubation.

273 Because we used both direct cell counts and plaque counts to quantify *Bdellovibrio*, we
274 validated the methods by comparing initial *Bdellovibrio* concentrations for the survival assay
275 without prey as estimated by direct cell counts using the Petroff-Hausser counting chamber and
276 plaque counts using serial dilution and double agar overlay. If all cells counted in the Petroff-
277 Hausser counting chamber are viable, these estimates should be the same. For both NC01 (n=4)
278 and HD100 (n=3), the two quantification methods showed good agreement (Supplementary
279 Figure 1). The average initial concentration estimated by direct cell counts was 1.05 x 10^7
280 cells/ml for NC01 and 2.3 x 10^7 cells/ml for HD100, whereas the average initial concentration
281 estimated by plaque counts (reported as plaque-forming units (pfu)/ml) at time zero was 1.17 x
282 10^7 pfu/ml for NC01 and 2.71 x 10^7 pfu/ml for HD100. These data show that direct cell counts
283 and plaque counts provide comparable results for quantification of *Bdellovibrio*.

284 **Results**

285 **Bioswale soil harbors predatory bacteria belonging to *Bdellovibrio***

286 Using Gram-negative soil bacteria isolated from a residential area as prey, we performed
287 an enrichment followed by three rounds of plaque picking to obtain a pure isolate of predatory
288 bacteria from a sample of soil collected at a bioswale on the Providence College campus. The
289 bioswale is a manmade landscape feature designed to collect and filter stormwater runoff from
290 nearby buildings. We designated the bioswale soil isolate as NC01. Using 1000x phase-contrast
291 microscopy of lysates combining NC01 and Gram-negative prey, we observed small, fast-
292 moving attack phase NC01 cells that attached to prey cells (Figure 1a). We also observed
293 rounded structures characteristic of bdelloplasts (Figure 1b), suggesting that NC01 uses a cell
294 invasion strategy similar to *Bdellovibrio bacteriovorus*. Negative stain electron microscopy
295 showed that NC01 attack phase cells have a comma-shaped morphology with a single polar
296 flagellum (Figure 1c).
297 To identify NC01, we constructed a phylogenetic tree of 16S rRNA gene sequences from
298 this isolate, all five *Bdellovibrio bacteriovorus* with complete genomes available in GenBank,
299 and representatives of *Bacteriovorax*, *Bdellovibrio exovorus* and *Halobacteriovorax* with
300 complete genomes available in GenBank (Figure 2). In this phylogenetic tree, NC01 clearly
301 clusters with other *Bdellovibrio*. Pairwise alignment of 16S rRNA gene sequences from NC01
302 and *Bdellovibrio bacteriovorus* HD100, which is the type strain of the species, showed 97%
303 similarity across 1,516 nucleotides. This value is above the 95% similarity cutoff typically used
304 to define bacterial genera (54), which supports identifying NC01 as *Bdellovibrio*. However, it is
305 below the 98.7% similarity cutoff typically used to define species (55), which suggests that
306 NC01 belongs to a different species than *bacteriovorus*.
307 To explore this further, we compared the average amino acid identity (AAI) of
308 *Bdellovibrio* strains NC01, W, HD100 and 109J (Table 1). HD100 and 109J, which are both

309 identified as *B. bacteriovorus*, cluster in a well-supported clade in the 16S rRNA phylogenetic
310 tree, and their amino acid sequences are highly similar (average AAI 99.2%). However, NC01
311 amino acid sequences are <70% similar on average to HD100 and 109J amino acid sequences,
312 supporting the conclusion that NC01 belongs to a different species than these two strains. NC01
313 amino acid sequences are even less similar on average to those of strain W, suggesting that
314 NC01 does not belong to the same species as this strain. Based on these analyses, we chose not
315 to designate a species for NC01, and we refer to this isolate as *Bdellovibrio* sp. NC01.

316 **Analysis of ortholog groups identifies protein-coding genes that are missing in *Bdellovibrio***
317 **sp. NC01 but conserved in other sequenced *Bdellovibrio***

318 The genome of *Bdellovibrio* sp. NC01 is larger than that of both HD100 and W (Table 2),
319 with a lower GC content than that of HD100, but slightly higher than that of W. We compared
320 the NC01 genome to all five publicly available complete *Bdellovibrio bacteriovorus* genomes
321 and the *Bdellovibrio exovorus* JSS genome using a software pipeline that clusters protein-coding
322 genes into ortholog groups based on amino acid sequence similarity (49). Because the RAST and
323 PGAP annotation tools generate slightly different annotations of the same genome sequence, we
324 performed separate analyses of ortholog groups using either RAST or PGAP annotations of the
325 seven genomes in our dataset. To determine the "core genome" in each analysis, we identified
326 ortholog groups that included at least one protein-coding gene from each genome. The core
327 genome based on RAST annotations included 1,364 ortholog groups, whereas the core genome
328 based on PGAP annotations included 1,394 ortholog groups.

329 Recently, Oyedara and colleagues conducted comparative genomics of the same six
330 publicly available complete *Bdellovibrio* genomes and the genome of *Bdellovibrio* sp.
331 SKB1291214, which is not complete (23). The core genome shared by these seven *Bdellovibrio*

332 strains included 795 protein-coding genes. We compared these to the PGAP core genome
333 described above and identified 778 protein-coding genes that were conserved in all eight
334 *Bdellovibrio* genomes. Of the 17 protein-coding genes that were identified as part of the core
335 genome by Oyedara and colleagues but not by our analysis, four were missing only in NC01.
336 The remaining 13 were missing in some of the other *Bdellovibrio* genomes according to our
337 analysis, likely due to differences in the parameters, such as alignment length, used by the two
338 software programs to cluster genes into ortholog groups.

339 To further investigate the four protein-coding genes identified as missing only in NC01,
340 we used BLAST to align representative amino acid sequences from the corresponding ortholog
341 groups against the full set of amino acid sequences annotated in NC01 by PGAP and RAST.
342 These analyses confirmed that two of the four genes are missing in NC01. The two genes are
343 predicted to encode a transporter permease (Bd1145/WP_011163662) and a protein with a
344 conserved domain involved in synthesis of cytochrome c oxidase (Bd3223/WP_011165579). The
345 two remaining genes are predicted to encode a hypothetical protein and the flagellar biosynthesis
346 protein FliL. For each of these genes, BLAST alignments yielded evidence of a possible ortholog
347 within the NC01 genome.

348 In particular, BLASTP alignment showed 78% sequence similarity between the protein-
349 coding gene annotated as a hypothetical protein (Bd1976/WP_011164428) and a 493 aa
350 sequence in NC01 that was annotated by both PGAP (DOE51_09405) and RAST (peg.1859).
351 The NC01 amino acid sequence is 161 aa longer than the representative amino acid sequence and
352 other sequences in the ortholog group due to a stop site difference. It was clustered in an ortholog
353 group by itself in the analyses of PGAP and RAST protein-coding genes, likely due to the
354 alignment length cutoff. Because the NC01 amino acid sequence shares high sequence similarity

355 with members of the hypothetical protein ortholog group over two-thirds of its length, it may
356 have derived from the same ancestral gene.

357 Finally, the flagellar biosynthesis protein FliL was identified as missing only in NC01
358 because it was annotated as a pseudogene by PGAP (DOE51_16425). However, RAST
359 annotated it as a functional protein-coding gene (peg.3231). BLASTP alignment showed 89%
360 sequence similarity between a representative FliL sequence (Bd3329/WP_011165675) and the
361 NC01 amino acid sequence annotated by RAST, but over only 69% of the representative
362 sequence. This is due to differences in predicted start sites and therefore length between the
363 NC01 amino acid sequence (177 aa) and the representative amino acid sequence (236 aa).
364 TBLASTN alignment of the representative amino acid sequence against the region of the NC01
365 nucleotide sequence annotated as a pseudogene by PGAP (3428266-3428813) confirmed that
366 there is no alternative start site in this region, and the NC01 protein-coding gene is shorter. These
367 analyses explain why the flagellar biosynthesis protein FliL was annotated as a pseudogene by
368 PGAP and therefore not included in the ortholog group generated using PGAP annotations.
369 However, the NC01 protein-coding gene was included in the FliL ortholog group when using
370 RAST annotations.

371 **Comparative genomics identifies eight regions of unique gene content possibly acquired by**
372 ***Bdellovibrio* sp. NC01 via horizontal gene transfer**

373 Using the BLAST Ring Image Generator (BRIG), we conducted pairwise comparisons of
374 NC01 to HD100 and W based on nucleotide sequence alignments with BLASTN. We chose
375 HD100 and W based on the 16S rRNA gene phylogenetic tree and average AAI comparisons to
376 represent phylogenetic diversity and genome variation among sequenced intraperiplasmic
377 *Bdellovibrio*. BRIG analysis identified eight NC01 genomic regions of at least 15,000

378 nucleotides in which no alignments of >200 nucleotides and e-value <0.1 to HD100 or W were
379 reported. These eight regions, which may represent unique NC01 gene content, are marked on
380 Figure 3. The plot of GC content shows that four of the eight regions (regions 1, 4, 5, and 7)
381 encompass substantial decreases in GC content compared to the typical variation along the
382 genome. Such deviation in GC content may be an indicator of horizontal gene transfer (HGT)
383 (56).

384 To assess whether the genes encoded in these eight NC01 genomic regions are present
385 but divergent in HD100 and W, we used TBLASTN to compare NC01 amino acid sequences
386 annotated in each region by RAST to the HD100 and W genomes translated into all six reading
387 frames. RAST annotated 11 to 50 protein-coding sequences within each of the eight NC01
388 genomic regions. By searching for these sequences along the entire HD100 and W genomes, this
389 analysis detects similar protein-coding sequences even if the genes have undergone
390 rearrangement. For the majority of the protein-coding sequences in each of the eight NC01
391 regions, TBLASTN did not produce any alignments to HD100 or W with >70% query coverage.
392 For those protein-coding sequences that did align to HD100 and/or W with >70% query
393 coverage, identity was low (<40%) for most of the sequences. Less than five protein-coding
394 sequences in each region aligned to HD100 or W at >70% coverage and >40% identity. This
395 suggests that the protein-coding sequences in the eight NC01 genomic regions are missing from
396 HD100 and W, instead of divergent.

397 In addition to comparisons against HD100 and W, we also used BLASTP to align amino
398 acid sequences annotated in each region by RAST against the non-redundant GenBank database.
399 For each sequence, we examined the top 250 alignments for any sequences identified as
400 *Bdellovibrio*. This analysis expands the search space beyond complete *Bdellovibrio* genomes to

401 any deposited sequences that were denoted by the submitters as *Bdellovibrio*. In two of the
402 regions (regions 1 and 4), only one-third of amino acid sequences aligned to *Bdellovibrio*
403 sequences. In every other region except region 8, 52-64% of the amino acid sequences aligned to
404 *Bdellovibrio* sequences. In region 8, three-fourths of the amino acid sequences aligned to
405 *Bdellovibrio* sequences. These results suggest that most of the protein-coding sequences in
406 regions 1 and 4 may be unique to *Bdellovibrio* sp. NC01, because they do not align to other
407 *Bdellovibrio* sequences in the database. In the other regions, the larger proportion of shared gene
408 content between NC01 and *Bdellovibrio* in the database may be the result of HGT events that
409 occurred before the divergence of NC01 and these strains.

410 The hypothesis that the eight NC01 genomic regions were acquired by horizontal gene
411 transfer is also supported by predicted functions of the protein-coding sequences encoded within
412 the regions. RAST annotation identified functions related to mobile elements in five of the eight
413 regions. These include phage tail fibers (regions 1, 5 and 6), site-specific recombinases or
414 integrases (regions 1, 4, and 7), a phage-related DNA primase (region 4), and a plasmid
415 conjugative transfer protein (region 7). In addition, Prophinder analysis of the complete NC01
416 genome identified a portion of region 5 (1,632,099 – 1,632,833) as a possible prophage. Three
417 regions (regions 2, 3 and 8) did not encode any obvious mobile element functions.

418 Apart from mobile element-related genes, most of the protein-coding sequences in the
419 eight NC01 genomic regions are annotated as hypothetical proteins by RAST, therefore it is
420 unclear what functional capacities may have been gained by *Bdellovibrio* sp. NC01 via HGT and
421 whether those functions impact predation. Some of the genes were annotated with predicted
422 functions, which enabled us to assess gene content in the eight regions in terms of broad
423 functional categories. Four of the eight regions (regions 1, 3, 6 and 8) encode predicted functions

424 related to transcriptional regulation, and five regions (regions 1, 3, 5, 6, and 7) encode predicted
425 functions related to transport. Regarding enzymatic activity, region 6 encodes a predicted cell
426 wall endopeptidase, which is interesting given *Bdellovibrio*'s interactions with the Gram-
427 negative prey cell wall during its life cycle.

428 **Differential gene content between *Bdellovibrio* sp. NC01 and HD100 includes protein-**
429 **coding genes that may be important for predation**

430 In a pairwise comparison of protein-coding gene content between NC01 and HD100, we
431 found that ~32% of NC01 protein-coding genes are not found in HD100 and ~27% of HD100
432 protein-coding genes are not found in NC01 (Table 3). To assess broad functional classification
433 of these differentially present genes, we used RAST annotations of the genomes, primarily
434 because, in our experience, RAST predicts functions for more protein-coding genes, whereas
435 PGAP classifies more protein-coding genes as "hypothetical proteins". Additionally, both
436 genomes were annotated using the same version of RAST, which ensures consistency in gene
437 calls and functional predictions. Of the 1,213 genes identified as present in NC01 and absent in
438 HD100, 516 were assigned a predicted function by RAST, and 697 were annotated as
439 hypothetical proteins. Of the 1,026 genes identified as present in HD100 and absent in NC01,
440 465 were assigned a predicted function by RAST, and 561 were annotated as hypothetical
441 proteins. Protein-coding genes annotated as "hypothetical proteins" may impact predation
442 phenotype differences between NC01 and HD100; however, the lack of predicted function
443 precludes any hypotheses about the roles of the gene products in predation.

444 For the differentially present protein-coding genes assigned a predicted function by
445 RAST, we used InterProScan (57) to assess how the amino acid sequences were grouped into
446 families based on functional or structural classifications. Gene Ontology (GO) classifications are

447 often used for this purpose; however, only half of the amino acid sequences were assigned GO
448 terms. By contrast, Pfam classifications were assigned to 461 of the amino acid sequences
449 identified as present in NC01 and absent in HD100 (89.3% of those annotated with a predicted
450 function) and 408 of the amino acid sequences identified as present in HD100 and absent in
451 NC01 (87.7% of those annotated with a predicted function). Overall, differentially present amino
452 acid sequences belonged to a range of Pfam families, with 349 unique functional or structural
453 classifications for those present in NC01 and absent in HD100 and 336 unique functional or
454 structural classifications for those present in HD100 and absent in NC01.

455 Previous studies of *B. bacteriovorus* have analyzed gene expression and constructed
456 mutant libraries to identify protein-coding genes likely important for predation. Lambert and
457 colleagues identified 240 protein-coding genes that were up-regulated in HD100 after 30 minutes
458 of interaction between *Bdellovibrio* and *E. coli* prey (50), whereas Duncan and colleagues used a
459 library of *B. bacteriovorus* 109J mutants generated by transposon insertions to identify 104
460 protein-coding genes that are important for *Bdellovibrio* fitness when preying on planktonic *E.*
461 *coli* (51). One protein-coding gene (Bd1399) was identified by both studies, yielding a combined
462 list of 343 unique protein-coding genes implicated in successful predation of *E. coli* by
463 *Bdellovibrio*. To determine whether these genes are present in NC01, we used NCBI annotations
464 of the genomes, which allowed us to match the HD100 locus tags referenced in these studies to
465 NCBI amino acid sequence accession numbers used in the identification of ortholog groups.

466 Of the 343 HD100 protein-coding genes implicated as important for predation on *E. coli*,
467 90 are missing in NC01 based on lack of a NC01 amino acid sequence in the ortholog group
468 including the HD100 amino acid sequence. This includes 80 from the list of genes identified by
469 Lambert and colleagues and 11 from the list identified by Duncan and colleagues, with Bd1399

470 included in both sets. Most of these differentially present genes are annotated as hypothetical
471 proteins or proteins of unknown function. However, 14 are annotated with a predicted function.
472 Ten of these may play important roles in predation, including four protein-coding genes
473 predicted to function in peptidoglycan metabolism (Bd0816/WP_011163367: D-alanyl-D-
474 alanine carboxypeptidase, Bd1358/WP_011163852: peptidoglycan-binding protein,
475 Bd1673/WP_011164147: LTA synthase family protein, Bd3575/WP_011165900: lytic
476 transglycosylase domain-containing protein), four annotated as serine proteases
477 (Bd0629/WP_011163194: trypsin-like serine protease, Bd2332/WP_011164755: serine protease,
478 Bd2334/WP_011164757: serine protease, Bd3238/WP_011165592: serine protease precursor),
479 and two protein-coding genes predicted to function in lipid metabolism or transport
480 (Bd0976/WP_011163513: helicase with a phospholipase domain, Bd3162/WP_011165523:
481 kinase). These ten genes may be involved in lytic activity related to the *Bdellovibrio* predatory
482 lifecycle. Our analysis indicates that these genes are missing from NC01, although it is possible
483 that the functions are fulfilled by other protein-coding genes that do not belong to the same
484 ortholog groups as the HD100 protein-coding genes.

485 ***Bdellovibrio* sp. NC01 kills fewer tested prey strains than HD100 and is less efficient at**
486 **killing *E. coli* ML35**

487 To assess variation in predatory phenotypes between *Bdellovibrio* sp. NC01 and *B.*
488 *bacteriovorus* HD100, we assayed prey range differences and predation efficiency. To determine
489 differences in prey range, we tested NC01 and HD100 against a panel of eight Gram-negative
490 bacteria isolated from different environments, including a freshwater stream and soil from a
491 residential area (Table 4). Plaque formation on a bacterial lawn indicates that the predatory
492 isolate is able to attack and lyse that particular Gram-negative prey strain. NC01 formed plaques

493 on lawns of five of the eight prey strains tested, whereas HD100 formed plaques on lawns of all
494 eight prey strains. NC01 was not able to attack and lyse two prey strains belonging to
495 *Acinetobacter* and *Raoultella* that were isolated from a freshwater stream. In addition, although
496 NC01 formed plaques on *E. coli* ML35, it did not attack and lyse *E. coli* 0057.

497 To measure predation efficiency, we quantified viable *E. coli* ML35 by calculating cfu/ml
498 at different time points over 72 hours of co-culture with NC01 or HD100 (Figure 4). In the
499 absence of predatory bacteria, we did not observe any reduction in viable *E. coli* ML35,
500 demonstrating that co-culture conditions do not adversely affect the ML35 population. After 6
501 hours of co-culture, NC01 did not appreciably reduce viable *E. coli* ML35 compared to the
502 control, whereas HD100 reduced viable ML35 by approximately 1 log unit. After 24 hours of co-
503 culture, there was a substantial difference in the impact of NC01 and HD100 on the *E. coli*
504 population. NC01 reduced viable ML35 by approximately 3 log units compared to the control,
505 whereas HD100 reduced viable ML35 by 6 log units. Over the next 48 hours, HD100 continued
506 to restrict the viable ML35 population, but the population recovered in co-cultures with NC01,
507 increasing by 2 log units. ML35 population recovery in the presence of NC01 was consistently
508 observed in all assay replicates, and cfu/ml plates showed no evidence of contaminant bacteria.

509 We quantified NC01 and HD100 after 72 hours of co-culture with *E. coli* ML35 to
510 determine if recovery of the ML35 population is associated with changes in predatory bacteria
511 population density. Using direct cell counts to estimate *Bdellovibrio* cells/ml at the start of co-
512 culture and plaque counts to estimate pfu/ml after 72 hours of co-culture, we observed an
513 increase in the HD100 population of 1 log unit, whereas the NC01 population decreased by 1 log
514 unit (Figure 5). When compared in validation tests, direct cell counts and plaque counts yielded
515 comparable results (Supplementary Figure 1), therefore the observed decrease in NC01

516 population is not due to a difference in methods for determining predator population density.
517 These data demonstrate that for both NC01 and HD100, active predatory bacteria are present
518 after 72 hours of co-culture with *E. coli* ML35, but, in contrast to HD100, the NC01 population
519 density decreases over the course of the predation efficiency assay.

520 To determine how predator population density during co-culture with *E. coli* ML35
521 compares to *Bdellovibrio* survival without prey, we quantified NC01 and HD100 over 72 hours
522 of incubation in HM buffer alone. From similar starting concentrations, the viable NC01 and
523 HD100 populations both decreased by 2 and 1.5 log units, respectively. Overall, the combined
524 data on *Bdellovibrio* population density in the presence and absence of prey show that NC01
525 population density is higher in co-culture with *E. coli* ML35 than in the absence of prey, which
526 indicates that some NC01 attack phase cells are successfully invading ML35 prey cells and
527 replicating to produce progeny. However, the data also show that the decrease in the NC01
528 population over the course of the predation efficiency assay is similar to that experienced by both
529 NC01 and HD100 in starvation conditions. The death of NC01 cells and lack of production of
530 new NC01 progeny occurs in the predation efficiency assay despite the presence of viable *E. coli*
531 ML35.

532 **Plaques formed by *Bdellovibrio* sp. NC01 on *E. coli* ML35 prey lawns are not as clear as**
533 **those formed by HD100**

534 In addition to quantitative differences in the ability of *Bdellovibrio* sp. NC01 and *B.*
535 *bacteriovorus* HD100 to kill *E. coli* ML35, we also observed qualitative differences in the
536 appearance of plaques formed by each strain on lawns of ML35 (Figure 6). In contrast to HD100,
537 which forms completely clear, regular plaques, NC01 forms plaques of a similar size, but with
538 incomplete clearance of the prey lawn within the plaques. Some NC01 plaques remain

539 completely turbid over eight days of incubation, whereas others retain a turbid center but begin
540 to clear on the edges of the plaque.

541 **Discussion**

542 We isolated *Bdellovibrio* from soil collected at a manmade landscape feature designed for
543 stormwater management on a college campus. By comparing 16S rRNA gene sequences and
544 analyzing average amino acid identity (AAI) among multiple sequenced *Bdellovibrio*, we
545 determined that the *Bdellovibrio* isolate was sufficiently divergent from *B. bacteriovorus* type
546 strain HD100 and other *B. bacteriovorus* strains that it should be identified as a different species.
547 The only other species currently described within *Bdellovibrio* is *exovorus*; however, this species
548 uses an epibiotic predation strategy that differs from the cell invasion strategy observed by
549 microscopy for the soil *Bdellovibrio* isolate described here. Pending further characterization, we
550 designated the isolate as *Bdellovibrio* sp. NC01.

551 Because *Bdellovibrio* sp. NC01 is phylogenetically divergent from HD100 and other
552 intraperiplasmic *Bdellovibrio* with sequenced genomes, this isolate provides a valuable
553 opportunity to assess variation in genome content and predatory phenotypes among *Bdellovibrio*.
554 We clustered amino acid sequences from NC01, all five publicly available completely sequenced
555 *Bdellovibrio bacteriovorus*, and *Bdellovibrio exovorus* JSS into ortholog groups. We then
556 compared our analysis with a previous analysis of conserved protein-coding genes reported in
557 (23), which included the same six publicly available *Bdellovibrio* genomes and an additional
558 partially sequenced genome of a *Bdellovibrio* strain isolated from soil. This comparison
559 identified 778 protein-coding genes conserved among all eight *Bdellovibrio* genomes.
560 Intraperiplasmic *Bdellovibrio* genomes range in size from 3 - 4 Mb with 2,760 - 3,780 annotated
561 protein-coding genes, whereas the genome of the epibiotic predator *Bdellovibrio exovorus* JSS is

562 smaller at 2.66 Mb with 2,597 annotated protein-coding genes. Our analysis of the *Bdellovibrio*
563 "core genome" indicates that approximately 20-30% of the protein-coding genes encoded by a
564 given *Bdellovibrio* genome are conserved among intraperiplasmic and epibiotic *Bdellovibrio.*
565 Analysis of the *Bdellovibrio* core genome also identified four protein-coding genes that
566 appeared to be missing in *Bdellovibrio* sp. NC01 but conserved in the other seven *Bdellovibrio*.
567 We confirmed the absence of two protein-coding genes using manual BLASTP alignments. One
568 encodes a protein of unknown function. The other is annotated as a transporter permease
569 belonging to a conserved domain family that includes AcrB in *E. coli.* AcrB is the inner
570 membrane component of an efflux pump that transports small molecules such as bile salts and
571 some antibiotic compounds out of the cell (58). In addition, AcrB is involved in contact-
572 dependent growth inhibition, in which a bacterial cell attaches to another cell and inhibits its
573 growth. Mutants that do not express functional *acrB* are resistant to growth inhibition (59).
574 Given these functional roles, the absence of the transporter permease in *Bdellovibrio* sp. NC01
575 may impact this strain's susceptibility to small inhibitory molecules.
576 For the remaining two protein-coding genes that appeared to be missing only in NC01,
577 manual BLAST alignments provided evidence of an ortholog encoded by the NC01 genome. In
578 particular, the PGAP annotation of NC01 flagged a gene encoding the flagellar biosynthesis
579 protein FliL as a pseudogene; therefore, it was not included in an ortholog group in our analysis.
580 However, RAST annotated this gene as a functional protein-coding gene, although the resulting
581 protein product is 59 aa shorter than other *Bdellovibrio* FliL due to a difference in the location of
582 the start site. Given the shortened length of this NC01 amino acid sequence, it is unclear whether
583 it is a fully functional flagellar biosynthesis protein. Negative stain electron microscopy of NC01
584 attack phase cells shows a single, polar flagellum; therefore, NC01 is clearly able to assemble

585 flagella. In addition to the gene discussed here, RAST annotated two other protein-coding genes
586 as flagellar biosynthesis protein FliL. All three of the resulting amino acid sequences are
587 classified into the same Pfam family (pfam03748); therefore, the NC01 genome may encode
588 functionally redundant genes.

589 Genome-wide comparisons of NC01 to HD100 and W using BLASTN identified eight
590 regions of unique gene content along the NC01 genome, and TBLASTN confirmed that the
591 protein-coding sequences within these regions are missing from the genomes of HD100 and W.
592 Multiple lines of evidence support the hypothesis that these eight regions were acquired by
593 *Bdellovibrio* sp. NC01 via horizontal gene transfer. Four of the regions have substantially lower
594 GC content than the genome average, which may indicate horizontal acquisition from a bacterial
595 genome with a lower average GC content than that of NC01 (56). Five of the regions encode
596 genes with predicted functions related to mobile elements such as phage and plasmids, and one
597 of these regions was identified as a possible prophage by phage finding sequence analysis
598 software.

599 Alignment by BLASTP of amino acid sequences encoded within two of the eight regions
600 against the GenBank database suggested that these two regions may have been acquired recently
601 by *Bdellovibrio* sp. NC01, because most of the sequences did not align to *Bdellovibrio* sequences
602 (complete or partial) in the database. This indicates that, given the current representation of
603 *Bdellovibrio* in the database, gene content in these two regions is mostly unique to NC01. For the
604 other six regions, the majority of amino acid sequences encoded in these regions aligned to
605 *Bdellovibrio* sequences in the database. This suggests that these *Bdellovibrio* strains may be
606 closely related to NC01, and the HGT events corresponding to acquisition of these six regions
607 occurred in the ancestral lineage before NC01 diverged from these strains. Addition of more

608 complete *Bdellovibrio* genomes to the database will improve our ability to reconstruct
609 evolutionary events such as horizontal gene transfer.

610 The analysis of *Bdellovibrio* sp. NC01 gene content reported here further illustrates the
611 role of horizontal gene transfer as an important force in predatory bacteria genome evolution.
612 Comparative genomics has yielded evidence of both ancient and recent HGT in *Bdellovibrio* and
613 in the saltwater-adapted predatory genus *Halobacteriovorax* (24, 60–62). However, lack of
614 functional annotation of many *Bdellovibrio* protein-coding sequences prevents a thorough
615 assessment of how HGT events impact predatory phenotypes. Within the eight NC01 genomic
616 regions likely acquired via HGT, many of the protein-coding sequences are annotated as
617 hypothetical proteins, which provides no insight into their function and their potential roles in the
618 predatory life cycle. To fully understand how *Bdellovibrio* genome variation relates to variation
619 in predatory phenotypes, it is essential to assign functions to protein-coding genes.

620 As the type strain, *B. bacteriovorus* HD100 has been the focus of multiple studies
621 investigating molecular mechanisms governing predation. We compared protein-coding gene
622 content between NC01 and HD100 and showed that ~30% of genes found in one of the strains
623 are missing in the other. These differentially present genes encode proteins from a wide range of
624 functional and structural families. To further investigate the possible impact of differentially
625 present genes on predation by NC01, we assembled a set of 343 HD100 protein-coding genes
626 identified as important for predation by previous studies using gene expression analysis or
627 mutagenesis (50, 51). About a quarter of these genes are missing in NC01. Most are annotated as
628 hypothetical proteins, but ten were annotated with functions that may be important for lytic
629 activity at different stages of the predatory lifecycle. In particular, four HD100 protein-coding
630 genes that are missing in NC01 are predicted to play a role in peptidoglycan metabolism.

631 Invasion of a prey cell by *Bdellovibrio* involves opening a hole in the prey cell peptidoglycan,
632 then resealing and reshaping peptidoglycan to form a rounded bdelloplast (11, 15). If NC01 is
633 missing enzymes that play a role in these processes, it may impact the ability of this strain to
634 invade prey cells and establish itself in the prey periplasm.

635 In addition to genome comparisons, we also compared predatory phenotypes between
636 *Bdellovibrio* sp. NC01 and *B. bacteriovorus* type strain HD100. Based on plaque formation on
637 lawns of eight different Gram-negative bacteria, NC01 has a more restricted prey range than
638 HD100. In particular, HD100 was able to attack and kill both tested strains of *E. coli*, whereas
639 NC01 was only able to attack and kill *E. coli* ML35. Other researchers have also observed that a
640 particular *Bdellovibrio* strain may vary in its ability to kill different members of the same species
641 (17, 63); however, the mechanisms governing these outcomes are not known. By demonstrating
642 that strain-level variation within a Gram-negative species determines susceptibility to predation,
643 these data have potentially important implications for developing predatory bacteria as
644 biocontrol agents in the treatment of bacterial infections. If a *Bdellovibrio* isolate attacks and
645 kills one representative of a Gram-negative species, it does not guarantee that the isolate will
646 attack and kill other strains within the species. Further work to define the molecular mechanisms
647 underlying variation in prey range may enable prediction of whether a Gram-negative strain will
648 be susceptible to predation by a particular *Bdellovibrio* strain.

649 Although both NC01 and HD100 can attack and kill *E. coli* ML35, predation by HD100
650 causes a greater reduction in the ML35 population after 24 hours *in vitro* compared to predation
651 by NC01. HD100 then maintains a low ML35 population density, although it does not
652 completely eliminate viable ML35. By contrast, in co-cultures of NC01 and ML35, the ML35
653 population recovers after 24 hours, whereas the NC01 population decreases. This is likely not the

654 result of general predator-prey dynamics, such as an increase in search time due to reduced
655 ML35 numbers, because we did not observe ML35 population recovery in co-cultures with
656 HD100. In contrast to NC01, the HD100 population increases after co-culture with *E. coli* ML35.
657 When incubated under similar conditions but without prey, both NC01 and HD100 populations
658 decrease. Overall, data on *Bdellovibrio* population density suggest that NC01 is experiencing
659 starvation conditions during co-culture with *E. coli* ML35 despite the presence of viable prey.
660 Differences in predation efficiency between NC01 and HD100 on *E. coli* ML35 may also
661 be related to variation in plaque phenotype. Plaques formed by NC01 on lawns of ML35 are
662 turbid, whereas plaques formed by HD100 are completely clear. Turbid plaques have been
663 reported as a characteristic of prey-independent mutants of *Bdellovibrio*; however, these plaques
664 are typically smaller than those formed by the wild-type strain and sometimes have a small
665 colony of host-independent *Bdellovibrio* at their center (64, 65). Neither of these features were
666 observed when comparing NC01 and HD100 plaques, suggesting that the NC01 plaque
667 phenotype is not the result of prey-independent growth.
668 The predatory phenotypes observed during NC01 co-culture with *E. coli* ML35 may be
669 the result of a mechanism specific to the prey, the predator, or a combination of the two. Other
670 researchers have observed prey population recovery (66, 67), and Shemesh and Jurkevitch
671 demonstrated that recovery of *Erwinia carotovora* in co-cultures with *Bdellovibrio* isolated from
672 soil was due to development of phenotypic resistance to predation. It is possible that *E. coli*
673 ML35 uses a similar mechanism to develop resistance to predation by *Bdellovibrio* sp. NC01,
674 thereby causing the NC01 population to decrease as a result of starvation due to lack of
675 susceptible prey. This mechanism is either not expressed or not successful in co-cultures with *B.*
676 *bacteriovorus* HD100.

677 It is also possible that *E. coli* ML35 remains susceptible to predation throughout co-
678 culture, but NC01 is deficient in some aspect of the predatory lifecycle. Our comparison of
679 NC01 and HD100 protein-coding gene content indicated that NC01 is missing some lytic
680 enzymes that may be important for predation. If these functions are not fulfilled by other protein-
681 coding genes, multiple stages of the predatory lifecycle may be affected, resulting in reduced
682 predation efficiency compared to HD100. For example, the turbid plaques formed by NC01 on
683 lawns of *E. coli* ML35 may arise from incomplete lysis of ML35 cells. If prey cells are
684 successfully invaded but not lysed, *Bdellovibrio* progeny will be trapped, which would explain
685 the incomplete clearance of the prey lawn and the decrease in the NC01 population observed
686 during predation efficiency assays. Explanations of NC01 predatory phenotypes that are specific
687 to prey, such as plastic phenotype resistance, and those specific to NC01 itself, such as deficient
688 or missing lytic enzymes, may not be mutually exclusive. Further work will test these
689 explanations.

690 Overall, comparisons between *Bdellovibrio* sp. NC01 and *B. bacteriovorus* HD100
691 identified variation in predation phenotypes that may have important implications for the
692 ecological impact of each of these strains and their effectiveness in clinical applications. The
693 molecular mechanisms governing phenotypes such as prey range and predation efficiency are not
694 known. Comparative genomics can generate hypotheses linking gene content differences to
695 observed phenotype differences, but the lack of functional annotation for many protein-coding
696 genes complicates this effort. It is also likely that the outcomes of interactions between predatory
697 bacteria and prey, such as whether or not two different *Bdellovibrio* can kill the same Gram-
698 negative strain, are dependent upon mechanisms specific to the *Bdellovibrio* strain and

699 mechanisms specific to the prey strain. Defining these mechanisms is key for understanding
700 variation in predatory phenotypes among divergent *Bdellovibrio*.

701 **Author statements**

702 **Conflicts of interest**

703 The authors declare that there are no conflicts of interest.

704 **Funding information**

705 This research was supported by an Institutional Development Award (IDeA) from the
706 National Institute of General Medical Sciences of the National Institutes of Health under grant
707 no. P20GM103430 and by funding to LEW from Providence College. This material is based
708 upon work conducted at a Rhode Island NSF EPSCoR research facility, the Genomics and
709 Sequencing Center, supported in part by the National Science Foundation EPSCoR Cooperative
710 Agreement #EPS-1004057. This material is based upon work supported in part by the National
711 Science Foundation under EPSCoR Cooperative Agreement #OIA-1655221. The funders had no
712 role in study design, data collection and interpretation, or the decision to submit the work for
713 publication.

714 **Acknowledgements**

715 We thank Lisa Sadzewicz and Luke Tallon at the Institute for Genome Sciences at the
716 University of Maryland Baltimore for PacBio sequencing services and Janet Atoyan at the Rhode
717 Island Genomics and Sequencing Center for Illumina MiSeq sequencing services. We thank
718 Sean O'Donnell for isolating freshwater prey bacteria and Mark Martin and Brett Pellock for
719 donating bacterial strains. We thank Joey Cerra for contributing some of the scripts used for
720 comparative genomics.

721

722 **References**

723 1. Jurkevitch E, Davidov Y. 2007. Phylogenetic diversity and evolution of predatory
724 prokaryotes, p. 11–56. *In* Jurkevitch, E (ed.), Predatory Prokaryotes: Biology, Ecology and
725 Evolution. Springer Berlin Heidelberg, Berlin, Heidelberg.

726 2. Chen H, Laws EA, Martin JL, Berhane T-K, Gulig PA, Williams HN. 2018. Relative
727 contributions of *Halobacteriovorax* and bacteriophage to bacterial cell death under various
728 environmental conditions. MBio 9.

729 3. Welsh RM, Zaneveld JR, Rosales SM, Payet JP, Burkepile DE, Thurber RV. 2016.
730 Bacterial predation in a marine host-associated microbiome. ISME J 10:1540–1544.

731 4. Feng S, Tan CH, Constancias F, Kohli GS, Cohen Y, Rice SA. 2017. Predation by
732 *Bdellovibrio bacteriovorus* significantly reduces viability and alters the microbial
733 community composition of activated sludge flocs and granules. FEMS Microbiol Ecol 93.

734 5. Dwidar M, Monnappa AK, Mitchell RJ. 2012. The dual probiotic and antibiotic nature of
735 *Bdellovibrio bacteriovorus*. BMB Rep 45:71–78.

736 6. Kadouri DE, To K, Shanks RMQ, Doi Y. 2013. Predatory bacteria: a potential ally against
737 multidrug-resistant Gram-negative pathogens. PLoS ONE 8:e63397.

738 7. Tyson J, Sockett RE. 2017. Nature knows best: employing whole microbial strategies to
739 tackle antibiotic resistant pathogens. Environmental Microbiology Reports 9:47–49.

740 8. Martin MO. 2002. Predatory prokaryotes: an emerging research opportunity. J Mol
741 Microbiol Biotechnol 4:467–477.

742 9. Koval SF, Hynes SH, Flannagan RS, Pasternak Z, Davidov Y, Jurkevitch E. 2013.
743 *Bdellovibrio exovorus* sp. nov., a novel predator of *Caulobacter crescentus*. Int J Syst Evol
744 Microbiol 63:146–151.

745 10. Sockett RE. 2009. Predatory lifestyle of *Bdellovibrio bacteriovorus*. Annu Rev Microbiol
746 63:523–539.

747 11. Kuru E, Lambert C, Rittichier J, Till R, Ducret A, Derouaux A, Gray J, Biboy J, Vollmer
748 W, VanNieuwenhze M, Brun YV, Sockett RE. 2017. Fluorescent D-amino-acids reveal bi-
749 cellular cell wall modifications important for *Bdellovibrio bacteriovorus* predation. Nat
750 Microbiol 2:1648–1657.

751 12. Avidan O, Petrenko M, Becker R, Beck S, Linscheid M, Pietrokovski S, Jurkevitch E.
752 2017. Identification and characterization of differentially-regulated Type IVb pilin genes
753 necessary for predation in obligate bacterial predators. Sci Rep 7:1013.

754 13. Evans KJ, Lambert C, Sockett RE. 2007. Predation by *Bdellovibrio bacteriovorus* HD100
755 requires Type IV pili. Journal of Bacteriology 189:4850–4859.

756 14. Mahmoud KK, Koval SF. 2010. Characterization of type IV pili in the life cycle of the
757 predator bacterium *Bdellovibrio*. Microbiology (Reading, Engl) 156:1040–1051.

758 15. Lerner TR, Lovering AL, Bui NK, Uchida K, Aizawa S-I, Vollmer W, Sockett RE. 2012.
759 Specialized peptidoglycan hydrolases sculpt the intra-bacterial niche of predatory
760 *Bdellovibrio* and increase population fitness. PLOS Pathogens 8:e1002524.

761 16. Dashiff A, Junka RA, Libera M, Kadouri DE. 2011. Predation of human pathogens by the
762 predatory bacteria *Micavibrio aeruginosavorus* and *Bdellovibrio bacteriovorus*. Journal of
763 Applied Microbiology 110:431–444.

764 17. Jurkevitch E, Minz D, Ramati B, Barel G. 2000. Prey range characterization, ribotyping,
765 and diversity of soil and rhizosphere *Bdellovibrio* spp. isolated on phytopathogenic
766 bacteria. Appl Environ Microbiol 66:2365–2371.

767 18. Chanyi RM, Ward C, Pechey A, Koval SF. 2013. To invade or not to invade: two
768 approaches to a prokaryotic predatory life cycle. Can J Microbiol 59:273–279.

769 19. Rogosky AM, Moak PL, Emmert EAB. 2006. Differential predation by *Bdellovibrio*
770 *bacteriovorus* 109J. Curr Microbiol 52:81–85.

771 20. Rendulic S, Jagtap P, Rosinus A, Eppinger M, Baar C, Lanz C, Keller H, Lambert C, Evans
772 KJ, Goesmann A, Meyer F, Sockett RE, Schuster SC. 2004. A predator unmasked: life
773 cycle of *Bdellovibrio bacteriovorus* from a genomic perspective. Science 303:689–692.

774 21. Negus D, Moore C, Baker M, Raghunathan D, Tyson J, Sockett RE. 2017. Predator versus
775 pathogen: How does predatory *Bdellovibrio bacteriovorus* interface with the challenges of
776 killing Gram-negative pathogens in a host setting? Annu Rev Microbiol 71:441–457.

777 22. Hoeniger JFM, Ladwig R, Moor H. 1972. The fine structure of "resting bodies" of
778 *Bdellovibrio* sp. strain W developed in *Rhodospirillum rubrum*. Can J Microbiol 18:87–92.

779 23. Oyedara OO, Segura-Cabrera A, Guo X, Elufisan TO, Cantú González RA, Rodríguez
780 Pérez MA. 2018. Whole-genome sequencing and comparative genome analysis provided

781 insight into the predatory features and genetic diversity of two *Bdellovibrio* species isolated
782 from soil. Int J Genomics 2018.

783 24. Enos BG, Anthony MK, DeGiorgis JA, Williams LE. 2018. Prey range and genome
784 evolution of *Halobacteriovorax marinus* predatory bacteria from an estuary. mSphere
785 3:e00508-17.

786 25. Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ, Dymock D, Wade WG.
787 1998. Design and evaluation of useful bacterium-specific PCR primers that amplify genes
788 coding for bacterial 16S rRNA. Appl Environ Microbiol 64:795–799.

789 26. Heuer H, Krsek M, Baker P, Smalla K, Wellington EM. 1997. Analysis of actinomycete
790 communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic
791 separation in denaturing gradients. Appl Environ Microbiol 63:3233–3241.

792 27. Ewing B, Green P. 1998. Base-calling of automated sequencer traces using Phred. II. Error
793 probabilities. Genome Res 8:186–194.

794 28. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces
795 using Phred. I. Accuracy assessment. Genome Res 8:175–185.

796 29. Gordon D, Abajian C, Green P. 1998. Consed: A graphical tool for sequence finishing.
797 Genome Res 8:195–202.

798 30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search
799 tool. Journal of Molecular Biology 215:403–410.

800 31. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: Accurate high-throughput multiple
801 sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823–1829.

802 32. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naïve Bayesian classifier for rapid
803 assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol
804 73:5261–5267.

805 33. Jurkevitch E. 2012. Isolation and classification of *Bdellovibrio* and Like Organisms.
806 Current Protocols in Microbiology 7B.1.1-7B.1.20.

807 34. Koval SF. 2007. The search for hunters: culture-dependent and -independent methods for
808 analysis of *Bdellovibrio* and Like Organisms, p. 191–211. *In* Jurkevitch, E (ed.), Predatory
809 Prokaryotes: Biology, Ecology and Evolution. Springer Berlin Heidelberg, Berlin,
810 Heidelberg.

811 35. Cerra J, Donohue H, Kral A, Oser M, Rostkowski L, Zappia L, Williams LE. 2018.
812 Complete genome sequence of *Pseudomonas* sp. strain NC02, isolated from soil. Genome
813 Announc 6.

814 36. Kozlov A, Darriba D, Flouri T, Morel B, Stamatakis A. 2018. RAxML-NG: A fast,
815 scalable, and user-friendly tool for maximum likelihood phylogenetic inference. bioRxiv
816 447110.

817 37. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for
818 specialized analyses of microbial genomes and metagenomes. PeerJ Preprints 4:e1900v1.

819 38. Casale A, Clark S, Grasso M, Kryschuk M, Ritzer L, Trudeau M, Williams LE. 2018.
820 Complete genome sequence of *Escherichia coli* ML35. Genome Announc 6.

821 39. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A,
822 Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial
823 genome assemblies from long-read SMRT sequencing data. Nature Methods 10:563–569.

824 40. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler
825 transform. Bioinformatics 25:1754–1760.

826 41. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin
827 R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–
828 2079.

829 42. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q,
830 Wortman J, Young SK, Earl AM. 2014. Pilon: An integrated tool for comprehensive
831 microbial variant detection and genome assembly improvement. PLOS ONE 9:e112963.

832 43. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D.
833 2013. Using Tablet for visual exploration of second-generation sequencing data. Brief
834 Bioinform 14:193–202.

835 44. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S,
836 Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil
837 LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O,
838 Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using
839 subsystems technology. BMC Genomics 9:75.

840 45. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S,
841 Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the
842 Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic
843 Acids Res 42:D206-214.

844 46. Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches.
845 Bioinformatics 29:2933–2935.

846 47. Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for
847 analysis of transfer RNA genes. Nucleic Acids Res 44:W54-57.

848 48. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer
849 RNA genes in genomic sequence. Nucleic Acids Res 25:955–964.

850 49. Zhang Y, Sievert SM. 2014. Pan-genome analyses identify lineage- and niche-specific
851 markers of evolution and adaptation in *Epsilonproteobacteria*. Front Microbiol 5.

852 50. Lambert C, Chang C-Y, Capeness MJ, Sockett RE. 2010. The first bite--profiling the
853 predatosome in the bacterial pathogen *Bdellovibrio*. PLoS ONE 5:e8599.

854 51. Duncan MC, Gillette RK, Maglasang MA, Corn EA, Tai AK, Lazinski DW, Shanks RMQ,
855 Kadouri DE, Camilli A. 2019. High-throughput analysis of gene function in the bacterial
856 predator *Bdellovibrio bacteriovorus*. mBio 10:e01040-19.

857 52. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator
858 (BRIG): simple prokaryote genome comparisons. BMC Genomics 12:402.

859 53. Lambert C, Sockett RE. 2008. Laboratory maintenance of *Bdellovibrio*. Current Protocols
860 in Microbiology 9:7B.2.1-7B.2.13.

861 54. Stackebrandt E, Goebel BM. 1994. Taxonomic note: A place for DNA-DNA reassociation
862 and 16S rRNA sequence analysis in the present species definition in bacteriology.
863 International Journal of Systematic and Evolutionary Microbiology 44:846–849.

864 55. Stackebrandt E, Ebers J. 2006. Taxonomic parameters revisited: tarnished gold standards.
865 Microbiology Today 33:152–155.

866 56. Ravenhall M, Škunca N, Lassalle F, Dessimoz C. 2015. Inferring horizontal gene transfer.
867 PLoS Comput Biol 11.

868 57. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005.
869 InterProScan: protein domains identifier. Nucleic Acids Res 33:W116–W120.

870 58. Ma D, Cook DN, Alberti M, Pon NG, Nikaido H, Hearst JE. 1995. Genes *acrA* and *acrB*
871 encode a stress-induced efflux system of *Escherichia coli*. Mol Microbiol 16:45–55.

872 59. Aoki SK, Malinverni JC, Jacoby K, Thomas B, Pamma R, Trinh BN, Remers S, Webb J,
873 Braaten BA, Silhavy TJ, Low DA. 2008. Contact-dependent growth inhibition requires the
874 essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane
875 transport protein AcrB. Molecular Microbiology 70:323–340.

876 60. Gophna U, Charlebois RL, Doolittle WF. 2006. Ancient lateral gene transfer in the
877 evolution of *Bdellovibrio bacteriovorus*. Trends Microbiol 14:64–69.

878 61. Pan A, Chanda I, Chakrabarti J. 2011. Analysis of the genome and proteome composition
879 of *Bdellovibrio bacteriovorus*: indication for recent prey-derived horizontal gene transfer.
880 Genomics 98:213–222.

881 62. Hobley L, Lerner TR, Williams LE, Lambert C, Till R, Milner DS, Basford SM, Capeness
882 MJ, Fenton AK, Atterbury RJ, Harris MATS, Sockett RE. 2012. Genome analysis of a
883 simultaneously predatory and prey-independent, novel *Bdellovibrio bacteriovorus* from the
884 River Tiber, supports in silico predictions of both ancient and recent lateral gene transfer
885 from diverse bacteria. BMC Genomics 13:670.

886 63. Fratamico PM, Whiting RC. 1995. Ability of *Bdellovibrio bacteriovorus* 109J to lyse
887 Gram-negative food-borne pathogenic and spoilage bacteria. Journal of Food Protection
888 58:160–164.

889 64. Cotter TW, Thomashow MF. 1992. Identification of a *Bdellovibrio bacteriovorus* genetic
890 locus, *hit*, associated with the host-independent phenotype. J Bacteriol 174:6018–6024.

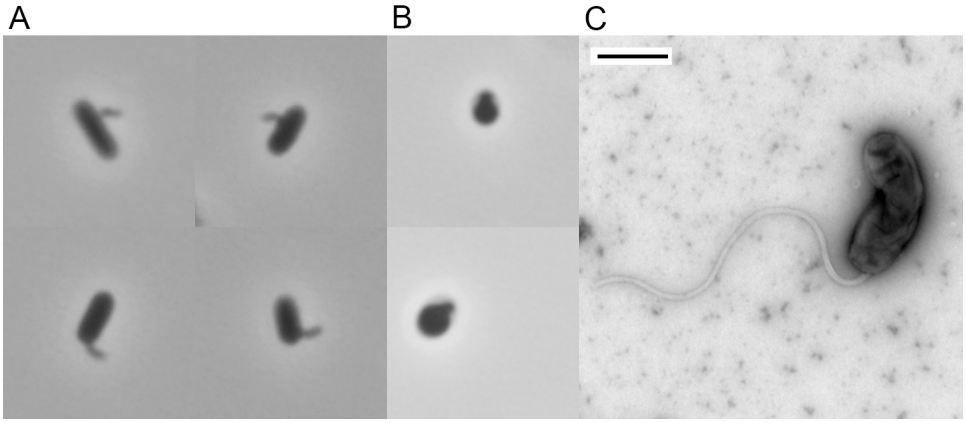
891 65. Roschanski N, Klages S, Reinhardt R, Linscheid M, Strauch E. 2011. Identification of
892 genes essential for prey-independent growth of *Bdellovibrio bacteriovorus* HD100. Journal
893 of Bacteriology 193:1745–1756.

894 66. Baker M, Negus D, Raghunathan D, Radford P, Moore C, Clark G, Diggle M, Tyson J,
895 Twycross J, Sockett RE. 2017. Measuring and modelling the response of *Klebsiella*
896 *pneumoniae* KPC prey to *Bdellovibrio bacteriovorus* predation, in human serum and
897 defined buffer. Sci Rep 7:8329.

898 67. Shemesh Y, Jurkevitch E. 2004. Plastic phenotypic resistance to predation by *Bdellovibrio*
899 and like organisms in bacterial prey. Environ Microbiol 6:12–18.

900

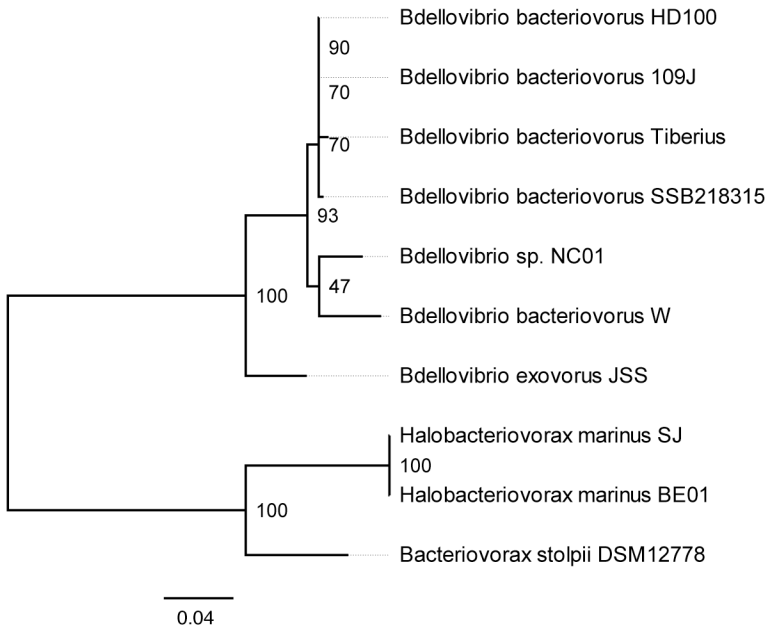
901 **Figures and Figure Legends**



902

903 **Figure 1.** Microscopy of *Bdellovibrio* sp. NC01 isolated from bioswale soil. (A) Compilation of
904 four images taken using 1000x phase-contrast microscopy to show small, comma-shaped NC01
905 attack phase cells attached to larger *E. coli* ML35 cells after 48 hours of co-culture. (B)
906 Compilation of two images taken using 1000x phase-contrast microscopy to show rounded
907 structures characteristic of bdelloplasts after 48 hours of co-culture. (C) Negative stain electron
908 microscopy of a NC01 attack phase cell. The scale bar is 500 nm.

909



910

911 **Figure 2.** 16S rRNA gene phylogenetic tree. We aligned 16S rRNA gene sequences using SINA

912 1.2.11, then constructed a phylogenetic tree using RAxML BlackBox with 100 bootstrap

913 replicates. The phylogenetic tree is rooted on the branch joining *Bdellovibrio* with

914 *Bacteriovorax*/*Halobacteriovorax*, and bootstrap support for each node is shown.

915

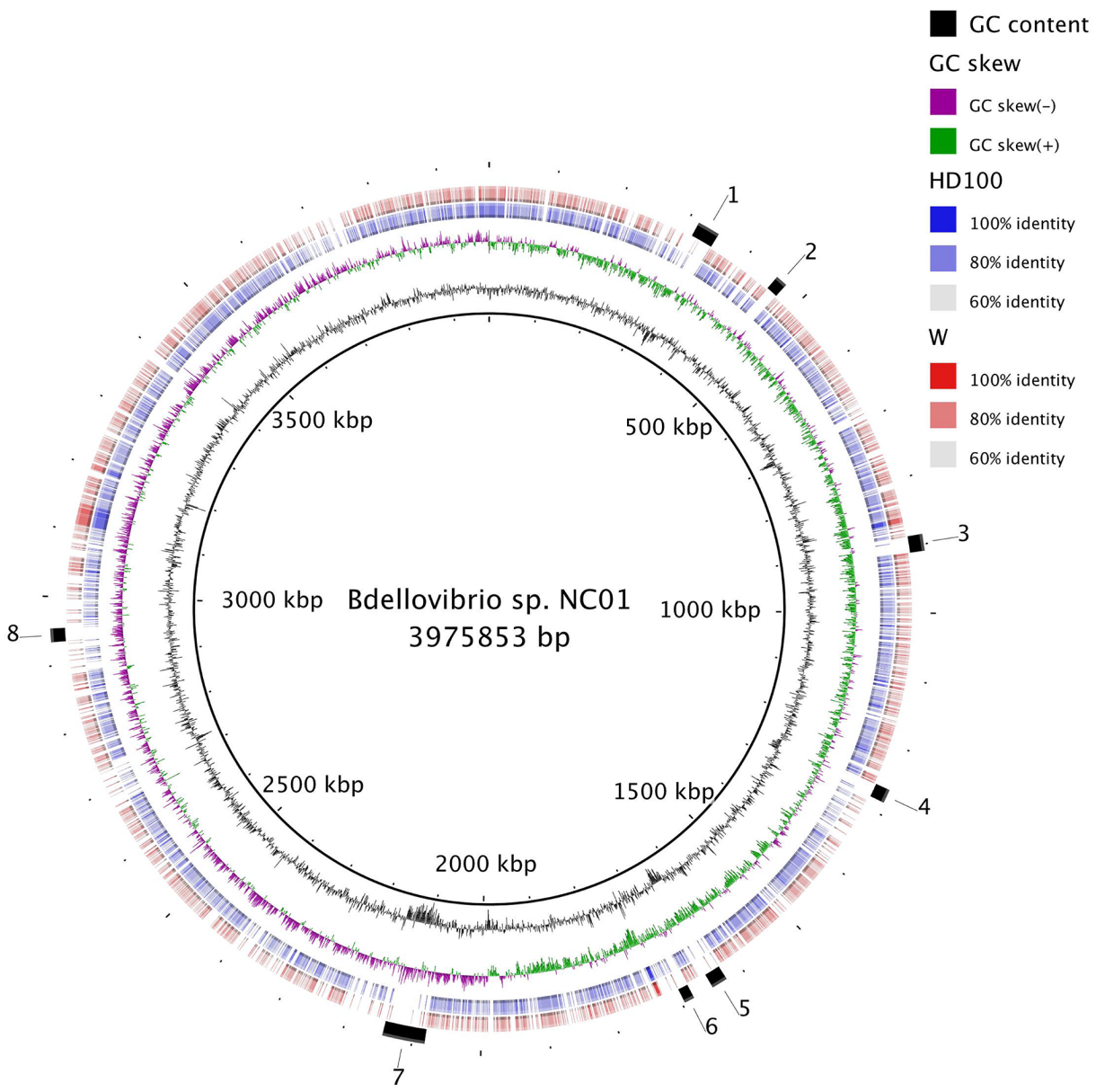
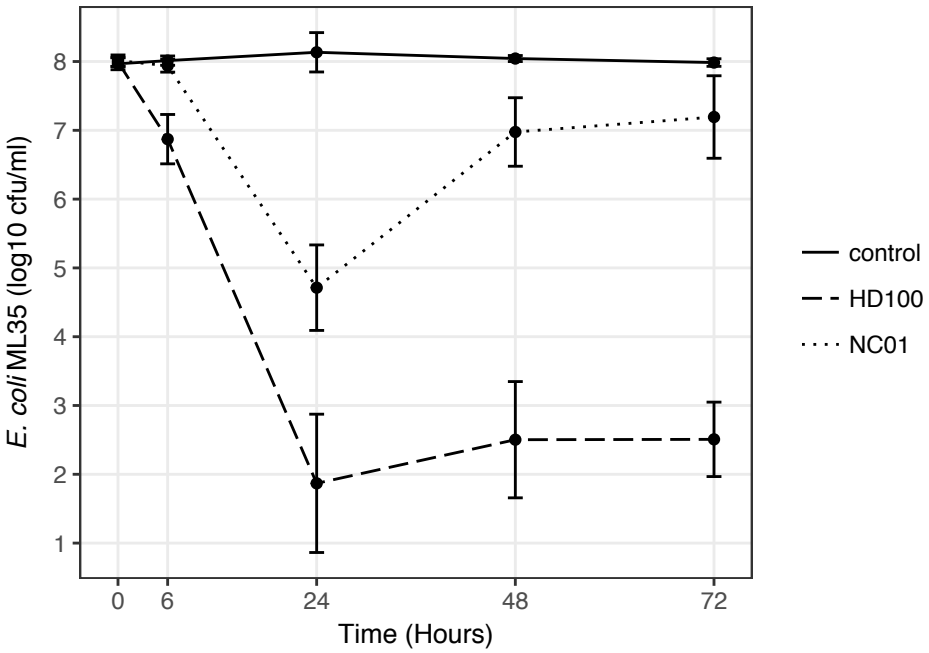


Figure 3. Genome-wide pairwise comparisons of *Bdellovibrio* nucleotide sequences. The genome of *Bdellovibrio* sp. NC01 is shown as the innermost ring. The next ring is a plot of GC content and the purple and green ring is a plot of GC skew, both of which were generated by the BLAST Ring Image Generator (BRIG) using sliding windows to determine deviation from the average value for the genome. The blue ring and red ring show pairwise nucleotide sequence alignments of NC01 to HD100 and W, respectively. Each individual alignment is shown as a block along the ring, and color gradation of the blocks reflects the percent similarity of

924 alignments. The outermost ring denotes eight regions lacking alignments, which may indicate

925 unique NC01 gene content.

926



927

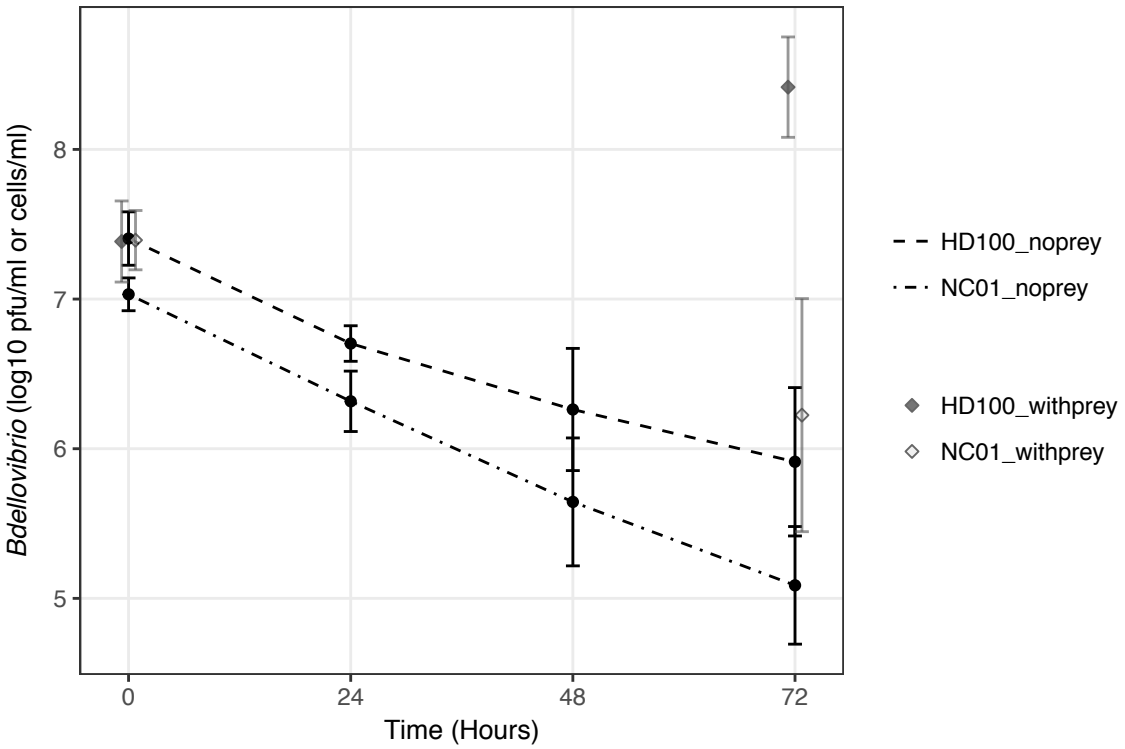
928 **Figure 4.** Efficiency of *Bdellovibrio* sp. NC01 and *B. bacteriovorus* HD100 predation on *E. coli*

929 ML35. We quantified viable *E. coli* ML35 over 72 hours on its own (control), combined with

930 *Bdellovibrio* sp. NC01 (NC01) and combined with *B. bacteriovorus* HD100 (HD100). The

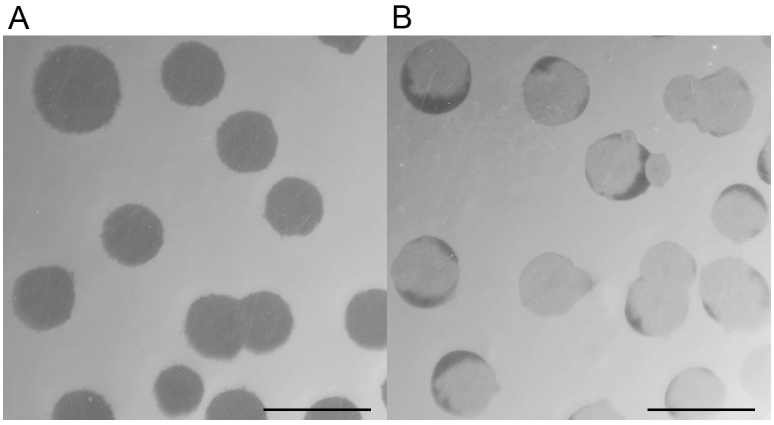
931 resulting cfu/ml data were log transformed and plotted using ggplot in RStudio. Each time point

932 shows the mean and standard deviation for at least five replicates.

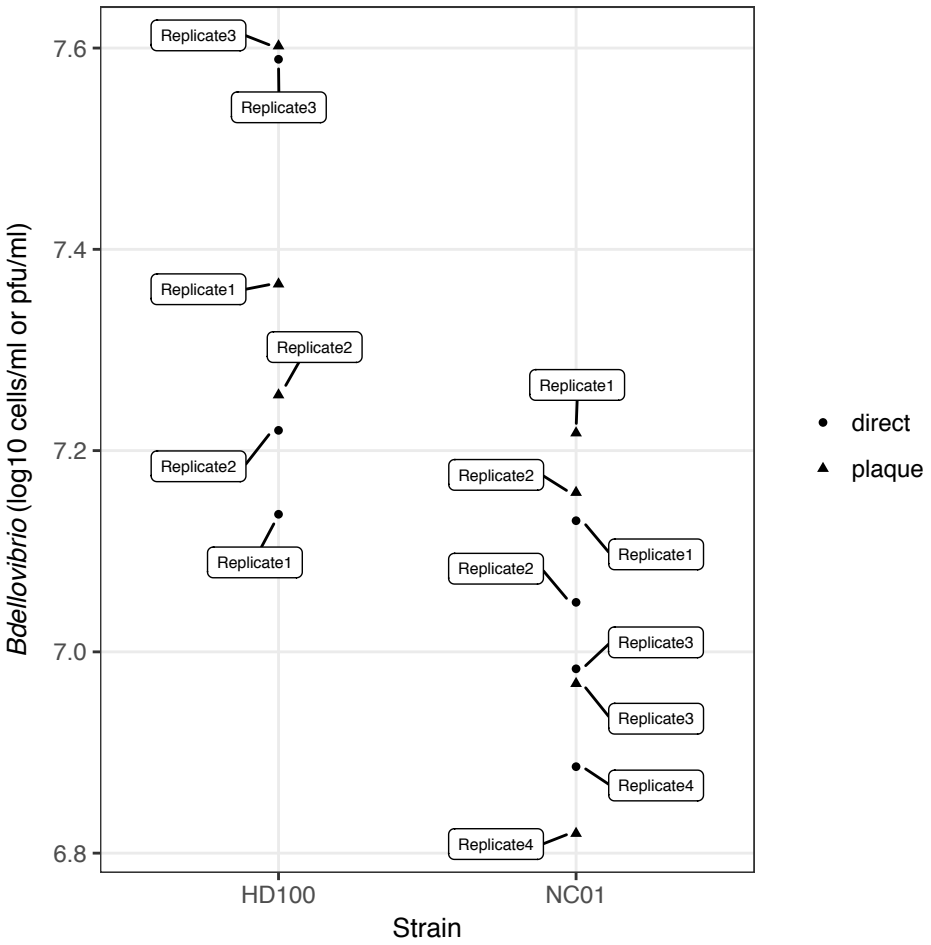


933

934 **Figure 5.** Population density of *Bdellovibrio* sp. NC01 and *B. bacteriovorus* HD100 in the
935 presence and absence of *E. coli* ML35. We quantified *Bdellovibrio* cells/ml at the start of the
936 predation efficiency assay and *Bdellovibrio* pfu/ml after 72 hours of co-culture with *E. coli*
937 ML35 (diamonds). For comparison, we quantified *Bdellovibrio* pfu/ml over 72 hours of
938 incubation in HM buffer alone (circles). The resulting cells/ml and pfu/ml data were log
939 transformed and plotted using ggplot in RStudio. Each time point shows the mean and standard
940 deviation for all replicates.

941 

942 **Figure 6.** Plaque phenotypes of *Bdellovibrio* sp. NC01 and *B. bacteriovorus* HD100. Using a
943 double agar overlay method, we assessed plaque formation on lawns of *E. coli* ML35. These
944 images were taken after 6 days of incubation and illustrate reproducible differences in the
945 appearance of plaques formed by HD100 (A) and NC01 (B). The scale bar is 10 mm.
946



947

948 **Supplementary Figure 1.** Quantification of *Bdellovibrio* using direct cell counts and plaque

949 counts. For each replicate of the survival assay without prey, initial concentrations of

950 *Bdellovibrio* HD100 and NC01 were estimated by both direct cell counts using a Petroff-Hausser

951 counting chamber (circles) and plaque counts using serial dilution and double agar overlay

952 (triangles). Replicates are labeled.

953

954 **Tables**

955 Table 1. Average amino acid identity among representative *Bdellovibrio* genomes

	NC01	W	HD100	109J
NC01		64.4	68.8	68.8
W	64.4		66.4	66.5
HD100	68.8	66.4		99.2
109J	68.8	66.5	99.2	

956

957 Table 2. Genome comparison of *Bdellovibrio* sp. NC01, *Bdellovibrio bacteriovorus* HD100 and
958 *Bdellovibrio bacteriovorus* W
959

	Bdellovibrio sp. NC01		*Bdellovibrio bacteriovorus* HD100		*Bdellovibrio bacteriovorus* W	
	PGAP	RAST and RFAM/ tRNAScan	PGAP	RAST and RFAM/ tRNAScan	PGAP	RAST and RFAM/ tRNAScan
Genome size (bp)	3,975,853		3,782,950		3,006,992	
GC content (%)	44.5		50.6		43.3	
Genes	3,840	3,828	3,597	3,651	2,834	2,901
Protein-coding	3,773	3,776	3,548	3,597	2,760	2,849
Hypothetical proteins	1,712	1,431	1,245	1,309	962	971
tRNA	34	34	36	36	34	34
rRNA	6	6	6	6	6	6
Other RNA	4	12	4	12	4	12

960

961 Table 3. Differentially present protein-coding genes between *Bdellovibrio* sp. NC01 and *B.*
962 *bacteriovorus* HD100
963

	PGAP	RAST
Present in NC01, absent in HD100	1175 (31.1%)	1213 (32.1%)
Present in HD100, absent in NC01	940 (26.5%)	1026 (28.5%)

964

965 Table 4. Prey range differences between *B. bacteriovorus* HD100 and *Bdellovibrio* sp. NC01

Genus	Strain ID	Environment	Plaque formation[a]	
			HD100	NC01
Acinetobacter	0036	Freshwater	Yes	No
Aeromonas	0031	Freshwater	Yes	Yes

Enterobacteriaceae	0032	Freshwater	Yes	Yes
Raoultella	0037	Freshwater	Yes	No
Pseudomonas	NC02 (0042)[b]	Soil	Yes	Yes
Serratia	0043	Soil	Yes	Yes
Escherichia	0057		Yes	No
Escherichia	ML35		Yes	Yes

966 [a]Plaque formation data based on three replicates

967 [b]*Pseudomonas* sp. NC02 is referred to as 0042 in (24)

968