1
2 **A Distinct Contractile Injection System Found in a Majority of Adult Human**
3 **Microbiomes**
4
5
6 Maria I. Rojas^{*1,2}, Giselle S. Cavalcanti^{*1,2}, Katelyn McNair^{1,3}, Sean Benler^{1,2}, Amanda
7 T. Alker^{1,2}, Ana G. Cobián-Güemes^{1,2}, Melissa Giluso^{1,3}, Kyle Levi^{1,3}, Forest Rohwer^{1,2,3},
8 Sinem Beyhan^{2,4}, Robert A. Edwards^{1,2,3}, Nicholas J. Shikuma^{#1,2,3,4}
9
10
11 1Viral Information Institute
12 San Diego State University
13 San Diego, California 92182 USA
14
15 2Department of Biology
16 San Diego State University
17 San Diego, California 92182 USA
18
19 3Computational Science Research Center
20 San Diego State University
21 San Diego, California 92182 USA
22
23 4Department of Infectious Diseases
24 J. Craig Venter Institute
25 La Jolla, California 92037 USA
26
27 *co-first authors
28 #corresponding author
29 nshikuma@sdsu.edu

30 **ABSTRACT**

31 An imbalance of normal bacterial groups such as Bacteroidales within the human gut is
 32 correlated with diseases like obesity. A current grand challenge in the microbiome field is
 33 to identify factors produced by normal microbiome bacteria that cause these observed
 34 health and disease correlations. While identifying factors like a bacterial injection system
 35 could provide a missing explanation for why Bacteroidales correlates with host health, no
 36 such factor has been identified to date. The lack of knowledge about these factors is a
 37 significant barrier to improving therapies like fecal transplants that promote a healthy
 38 microbiome. Here we show that a previously ill-defined Contractile Injection System is
 39 carried in the gut microbiome of 99% of individuals from the United States and Europe.
 40 This type of Contractile Injection System, we name here Bacteroidales Injection System
 41 (BIS), is related to the contractile tails of bacteriophage (viruses of bacteria) and have
 42 been described to mediate interactions between bacteria and diverse eukaryotes like
 43 amoeba, insects and tubeworms. Our findings that BIS are ubiquitous within adult human
 44 microbiomes suggest that they shape host health by mediating interactions between
 45 Bacteroidales bacteria and the human host or its microbiome.

46

47 INTRODUCTION

48 The human gut harbors a dynamic and densely populated microbial community (Gill et
 49 al., 2006). The gut microbiome of healthy individuals is characterized by a microbial
 50 composition where members of the Bacteroidetes phylum (*Bacteroides* and
 51 *Parabacteroides*) constitute 20-80% of the total (Consortium et al., 2012). Several studies
 52 have demonstrated that dysbiosis in the human gut is correlated with microbiome
 53 immaturity, and diseases like obesity and inflammatory bowel disease (Carroll et al.,
 54 2011; Hooper and Gordon, 2001; Kau et al., 2011; Ley et al., 2006; Subramanian et al.,
 55 2014). However, we currently do not know the identity of bacterial factors that explain
 56 these correlations between Bacteroidales abundances and healthy human phenotypes.
 57 The lack of knowledge about these factors is a significant barrier to improving therapies
 58 like fecal transplants that manipulate microbial communities to treat microbiome-related
 59 illness.

60

61 Many bacteria such as Bacteroidales produce syringe-like secretion systems called
 62 Contractile Injection Systems (CIS) that are related to the contractile tails of
 63 bacteriophage (bacterial viruses) (Cianfanelli et al., 2016; Salmond and Fineran, 2015).
 64 Most CIS characterized to date, termed Type 6 Secretion Systems (T6SS), are produced
 65 and act from within an intact bacterial cell. In contrast, extracellular CIS (eCIS) are
 66 released by bacterial cell lysis, paralleling the mechanism used by tailed phages to
 67 escape their host cell (**Figure 1A**). To date, Bacteroidales from the human gut have only
 68 been shown to produce one type of CIS, a Subtype-3 T6SS that mediates bacteria-
 69 bacteria interactions and can govern the microbial composition of the gut microbiome

70 (Chatzidaki-Livanis et al., 2016; Coyne et al., 2016; Russell et al., 2014; Verster et al.,
71 2017). However, CIS from other bacterial groups such as Gammaproteobacteria are
72 known to help bacteria interact with diverse eukaryotic organisms such as amoeba,
73 insects, tubeworms and humans.

74

75 Distinct from Bacteroidales Subtype-3 T6SS is a different class of CIS that may have
76 evolved independently (Böck et al., 2017). Intriguingly, all previously described examples
77 of these distinct CIS mediate bacteria-eukaryote interactions. Three are classified as
78 eCIS and include: (1) MACs (Metamorphosis Associated Contractile Structures) that
79 stimulate the metamorphosis of tubeworms (Shikuma et al., 2016, 2014), (2) PVCs
80 (*Photorhabdus* Virulence Cassettes) that mediate virulence in grass grubs (Yang et al.,
81 2006), and (3) Afp (Anti-Feeding Prophage) that cause cessation of feeding and death of
82 grass grub larvae (Heymann et al., 2013; Hurst et al., 2018, 2007, 2004). A fourth CIS
83 from *Amoebophilus asiaticus* (Böck et al., 2017) promotes intracellular survival in amoeba
84 and defines the Subtype-4 T6SS group. Although this class of CIS forms a superfamily of
85 structures found in diverse microbes (Chen et al., 2019), until the present work, examples
86 of this distinct class of CIS have only been identified in a few isolated Bacteroidetes
87 genomes (Böck et al., 2017; Penz et al., 2012; Sarris et al., 2014).

88

89 In this study, we show that diverse Bacteroidales species from the human gut encode a
90 distinct CIS within their genome that is related to eCIS and T6SS that mediate tubeworm,
91 insect and amoeba interactions (MACs, PVCs, Afp, Subtype-4 T6SS). Strikingly, we show
92 that this distinct CIS from Bacteroidales are present within the microbiomes of over 99%

93 of human adult individuals from Western countries (Europe and the United States) and
94 are expressed *in vivo*. Our results suggest that a previously enigmatic class of contractile
95 injection system is present and expressed within the microbiomes of nearly all adults from
96 Western countries.

97

98

99 RESULTS

100 **Bacteroidales bacteria from the human gut possess genes encoding a distinct**
 101 **Contractile Injection System.** Using PSI-BLAST to compare previously identified eCIS
 102 and Subtype-4 T6SS proteins to the non-redundant (nr) protein sequence database, we
 103 identified CIS structural proteins (baseplate, sheath and tube) that matched with proteins
 104 from various human Bacteroidales isolates, including a bacterial isolate from the human
 105 gut (*Bacteroides cellulosilyticus* WH2, **Table S1**) (McNulty et al., 2013). To determine the
 106 relatedness of these distinct CIS with all known CIS subtypes, we performed phylogenetic
 107 analyses of CIS proteins that are key structural components of CIS—the CIS sheath and
 108 tube. Multiple methods of phylogenetic analyses (maximum likelihood, neighbor joining,
 109 maximum parsimony, UPGMA, and minimum evolution) showed that Bacteroidales
 110 sheath and tube proteins consistently formed a monophyletic group with other eCIS and
 111 Subtype-4 T6SS sheath and tube proteins (**Figures 1B, S1, Table S2**). Moreover, the
 112 BIS sheath and tube were clearly distinct when compared to previously characterized
 113 *Bacteroides* Subtype-3 T6SS (**Figures 1B, S1**) (Chatzidaki-Livanis et al., 2016; Russell
 114 et al., 2014). Based on these data and results below, we name this distinct CIS as
 115 Bacteroidales Injection Systems (BIS).

Figure 1. Bacteroidales possess a distinct Contractile Injection System. (A) Contractile Injection Systems are related to the contractile tails of bacteriophage. There are two main types of CIS; Type 6 Secretion Systems (T6SS) act from within a bacterial cell, while extracellular CIS (eCIS) are released by bacterial cell lysis and bind to target cells. **(B)** Unrooted phylogeny of CIS sheath protein sequences. BIS group with known Subtype-4 T6SS and eCIS (orange) and are distinct from known Subtype-1, Subtype-2 and Subtype-3 T6SS. Bacteria with BIS identified in this study are highlighted in red.

Genes encoding BIS are found in a conserved cluster of genes and form three genetic arrangements. To identify Bacteroidales species that possess a bonafide BIS gene cluster, we performed a comprehensive search of 759 sequenced *Bacteroides* and *Parabacteroides* genomes from the Refseq database. Our sequence-profile search revealed 66 genomes from *Bacteroides* and *Parabacteroides* species that harbor complete BIS gene clusters (**Table S3**) in three conserved gene arrangements (**Figure 2A**). The first architecture is exemplified by *B. cellulosilyticus* WH2, which harbors two sheath proteins, two tube proteins, and a protein with unknown function intervening

138 between putative genes encoding the baseplate (gp25, gp27 and gp6). The second
 139 architecture is exemplified by *B. fragilis* BE1. This architecture has a single sheath protein
 140 and lacks the hypothetical proteins observed in architecture one between gp25 and gp27,
 141 and between Tube2 and LysM. Finally, the third architecture defined by *P. distasonis* D25
 142 is the most compact, lacks four hypothetical proteins found in architectures 1 and 2.
 143 Additionally, gp27 and gp6 proteins are shorter, and the genes FtsH/ATPase and
 144 DUF4157 are inverted. Importantly, all three genetic architectures have genes with
 145 significant sequence similarity to MAC, Afp and PVC genes shown previously to produce
 146 a functional CIS, including baseplate proteins (gp25, gp27, and gp6), sheath, tube, and
 147 FtsH/ATPase (**Figure 2B**).

148

Figure 2. Bacteroidales produce three architectures of BIS. (A) Synteny plot of BIS gene clusters in *Bacteroides* and *Parabacteroides* species compared to **(B)** those of *P. luteoviolacea* MACs, *S. entomophila* Afp, *Photorhabdus* PVCs, and *A. asiaticus* Subtype-4 T6SS. Representative CIS gene cluster architectures are shown, with genes color coded according to function. Genes with no significant sequence similarity at the amino acid level to any characterized proteins are colored light grey. Sequence coordinates of all gene clusters are provided in **Table S3**.

BIS genes are present in human gut, mouth, and nose microbiomes. To determine the prevalence and distribution of BIS genes in human microbiomes, we searched shotgun DNA sequencing data from 11,219 microbiomes from the Human Microbiome Project database, taken from several locations on the human body (Consortium et al., 2012; Levi et al., 2018). We sampled these metagenomes for the presence of 18 predicted BIS proteins (**Table S4**). Across all HMP metagenomes, 8,320 (74%) showed hits to at least one of the 18 BIS proteins. Hits were distributed across metagenomes from various mucosal tissues, and were more abundant in the gut and in the mouth, where Bacteroidales are found in high abundance (Consortium et al., 2012). The dataset included stool (1,851, 99.6% of total stool metagenomes), oral (4,739, 79.2% of total oral metagenomes), nasal (630, 41.8% of total nasal metagenomes), and vaginal (232, 27.7% of total vaginal metagenomes) samples (**Figure 3A**).

175

To determine how often any of the 18 genes co-occurred within the same metagenome sample, we constructed a co-occurrence network. Ten genes appeared together at high frequencies including Sheath1, Sheath2, FtsH/ATPase, baseplate (gp25, gp27, and gp6), LysM, Spike, and two hypothetical proteins (**Figure 3B**). The gene with the highest hit abundance encodes for an ATPase homologous to the *Escherichia coli* FtsH, known to be involved in cleavage of the lambda prophage repressor, followed by a hypothetical protein, and Sheath1. The other genes, including Tube1, Tube2, Tip, DUF4255 domain-containing protein, DUF4157 domain-containing protein, and three hypothetical proteins, were detected together less often within the microbiome samples.

185

186 (A)

187

188

189 (B)

190

191

192

193

194

195

196

197

198

Figure 3. BIS genes are abundant in human gut and mouth microbiomes. (A) Coverage plot of BIS genes (Log_{10} of $1,000,000 \times \text{hits}/\text{reads}$) in 8,320 microbiomes associated with mucosal tissue: gut, mouth, nose, and other (includes vaginal and skin tissues) from 300 healthy humans. **(B)** Ten BIS genes are found more often together in human metagenomes (co-occurrence network). Node size represents the number of hits for each protein across all runs. Line weight represents the number of times any two proteins occurred together within a dataset.

199 **BIS genes are expressed *in vivo* in the gut of humanized mice, and *in vitro* when**
200 **cultured with various polysaccharides.** To determine whether BIS genes are
201 transcribed under laboratory growth conditions or within the human gut, we searched for
202 the 18 major BIS proteins in publicly available RNA sequencing data from *in vivo*
203 metatranscriptomes of humanized mice (Ridaura et al., 2013) (gnotobiotic mice colonized
204 with human microbiome bacteria) and *in vitro* *B. cellulosilyticus* WH2 pure cultures
205 (McNulty et al., 2013). We inspected 59 metatranscriptomes from a previously published
206 *in vivo* study (Ridaura et al., 2013) for the presence of the 18 major BIS proteins, where
207 gnotobiotic mice were inoculated with human gut microbiome cultures. In 48 out of 59
208 metatranscriptomes (81.4%), we found expression of at least 15 BIS proteins (**Figure 4**).
209 Similarly, when *B. cellulosilyticus* WH2 was cultured in minimum media (MM)
210 supplemented with 31 different simple and complex sugars (McNulty et al., 2013), all 18
211 genes were expressed at least once in at least two of the three replicate cultures (**Figure**
212 **S2**). The highest expression was seen under growth in N-Acetyl-D-Galactosamine
213 (GalNAc) and N-Acetyl-Glucosamine (GlcNAc), amino sugars that are common
214 components of the bacterial peptidoglycan, in high abundance in the human colon, and
215 implicated in many metabolic diseases (Baudoin and Issad, 2014; Myslicki et al., 2014).
216 Our analyses of metatranscriptomes show that BIS genes are transcribed by
217 Bacteroidales bacteria under laboratory growth conditions and within humanized mouse
218 microbiomes *in vivo*.

219

Figure 4. BIS genes are expressed *in vivo* in humanized mice. Coverage plot of BIS genes (normalized by number of reads and protein nt size) from stool metatranscriptomes of humanized mouse microbiomes.

BIS genes are present in the microbiomes of nearly all adult individuals. To determine the prevalence of BIS within the microbiomes of human populations, we analyzed 2,125 fecal metagenomes from 339 individuals; 124 Individuals from Europe (Qin et al., 2010) and 215 individuals from North America (Consortium et al., 2012). We found that all individuals possessed at least 1 of the 18 BIS genes within their gut microbiome (**Figure 5**). Most individuals carried at least 9 BIS proteins (83.0% in HMP and 90.3% in Qin dataset). A lower number possessed all 18 BIS proteins (8.96% in HMP and 6.45% in Qin dataset).

Figure 5. BIS genes are present in the microbiomes of a majority (99%) of adult individuals from the United States and Europe. Frequencies of 18 BIS proteins from fecal samples of 339 individuals. HMP n=215, from healthy North American individuals (Consortium et al., 2012); Qin n=124, from a study of European individuals (Qin et al., 2010). Protein hits are normalized by individual donor.

DISCUSSION

Here, we show that a previously ill-described Contractile Injection System is present in the gut microbiomes of nearly all adult individuals from Western countries. We found that BIS genes are present in human microbiomes throughout mucosal tissues (oral, nasal, vaginal, ocular) and enriched in metagenomes from gut samples. CIS have gained recent recognition as secretion systems promoting disease in several prominent human pathogens like *Pseudomonas aeruginosa* and *Vibrio cholerae*. However, our discovery of a distinct class of CIS produced by Bacteroidales from human microbiomes stems from studies of symbiotic interactions between environmental bacteria and diverse eukaryotic

252 hosts like tubeworms, insects and amoeba (Böck et al., 2017; Chen et al., 2019;
 253 Pechenik, 1999; Shikuma et al., 2014, 2016; Vlisidou et al., 2019). The close relatedness
 254 of BIS with other structures promoting microbe-eukaryotic interactions (MACs, Afp, PVCs
 255 and Subtype-4 T6SS) suggests that BIS mediate interactions between bacterial species
 256 within the human microbiome or Bacteroidales bacteria and their human host.

257

258 If BIS do interact with human cells, they may promote either symbiotic or pathogenic
 259 interactions. Injection systems closely related to BIS are described to mediate both
 260 beneficial and infectious microbe-host relationships. For example, MACs mediate
 261 metamorphosis of a marine tube worms (Ericson et al., 2019; Shikuma et al., 2014) and
 262 a Subtype-4 T6SS (Böck et al., 2017) mediates membrane interaction between *A.*
 263 *asiaticus* and its amoeboid host. In contrast, Afp and PVCs inject toxic effectors into
 264 insects (Hurst et al., 2004; Yang et al., 2006). While a Subtype-3 T6SS has been
 265 described in *Bacteroides* bacteria, we speculate that BIS could be the first Bacteroidales
 266 CIS to mediate microbe-animal interactions.

267

268 Structures that are closely related to BIS have been described to form T6SS that act from
 269 within a bacterial cell (e.g. Subtype-4 T6SS from *A. asiaticus*; Böck et al., 2017) or eCIS
 270 that are released by cell lysis (e.g. MACs and PVCs; (Shikuma et al., 2014; Vlisidou et
 271 al., 2019). While we currently do not know whether BIS generate a T6SS and/or an eCIS
 272 structure, we do know that intestinal bacteria like *Bacteroides* are physically separated
 273 from the intestinal epithelium by a layer of mucus (Marcobal et al., 2011; Tropini et al.,
 274 2017; Van Der Waaij et al., 2005). In contrast to T6SS that are bound to and act from

275 within bacterial cells, an eCIS from Bacteroidales could cross this mucus barrier to inject
276 effectors into human epithelial cells.

277

278 We show here that BIS are produced by Bacteroidales *in vivo* during colonization of
279 humanized mice (Ridaura et al., 2013) and under laboratory conditions with various
280 carbon sources (McNulty et al., 2013). However, we do not yet know the conditions that
281 promote BIS production within normal or diseased human individuals. BIS may not have
282 been extensively described before this work because they likely evolved independently
283 from previously described CIS like Subtype-3 T6SS (Böck et al., 2017; Coyne et al., 2016)
284 and possess significantly divergent sequence homologies (**Figure 1B**). Like other
285 described CIS, BIS gene clusters possess genes encoding the syringe-like structural
286 components and could encode for effectors that elicit specific cellular responses from
287 target host cells. For example, the closely related injection system called MACs
288 possesses two different effectors; one effector protein promotes the metamorphic
289 development of a tubeworm (Ericson et al., 2019) and a second toxic effector kills insect
290 and mammalian cell lines (Rocchi et al., 2019).

291

292 Many correlations between Bacteroidales abundances in the human gut and host health
293 are currently unexplained. Future research into the conditions that promote the production
294 BIS and its protein effectors will yield new insight into how Bacteroidales abundances
295 correlate with host health, potentially by promoting the direct interaction between the
296 microbiome and human host. In addition to their effect on host health, BIS provide the
297 tantalizing potential as biotechnology platforms because they may be manipulated to

298 inject engineered proteins of interest into other microbiome bacteria or directly into human

299 cells.

300

301 MATERIALS AND METHODS

302

303 **Phylogenetic analyses of CIS structural proteins.** Whole genomes and assembled
304 contigs of representative phage-like clusters (**Table S2**) were downloaded from NCBI to
305 construct a database using BLAST+(2.6.0). The *B. cellulosilyticus* WH2 Sheath1 (WP
306 029427210.1) and Tube1 (WP 118435218) protein sequences were downloaded from
307 NCBI and a tBLASTn was performed against the genome database. The recovered
308 nucleotide sequences were then translated using EMBOSS Transeq (EMBL-EBI,
309 https://www.ebi.ac.uk/Tools/st/emboss_transeq/) to generate a list of protein
310 sequences. To capture highly divergent protein homologs, we used T6SS-Hcp, VipA/B;
311 Phage- gp18, gp19 as reference proteins. The amino acid sequences were aligned
312 using the online version of MAFFT (v7) with the iterative refinement alignment method
313 E-INS-i. The aligned fasta file was converted into a phylip file using Seaview (Gouy et
314 al., 2010). PhyML was performed through the ATCG Bioinformatics web server and
315 utilized the Smart Model Selection (SMS) feature and the Maximum-Likelihood method
316 (Guindon et al., 2010; Lefort et al., 2017). The best models for the Sheath1 and Tube1
317 phylogenies were rtREV+G+F and LG+G+I+F, respectively. Bootstrap values (1000
318 resamples) were calculated to ensure tree robustness. The Maximum likelihood tree
319 topology was confirmed using other methods, including Neighbor joining, Maximum-
320 Parsimony, Minimum-Evolution, and UPGMA. Trees were manipulated and viewed in
321 iTOL (Letunic and Bork, 2016).

322

323 **BIS gene cluster synteny analyses.** To identify CIS gene clusters in Bacteroidetes we
324 used a modified protocol used to identify T6SS (Coyne et al., 2016). Briefly, the
325 assemblies for 759 *Bacteroides* and *Parabacteroides* genomes included in the Refseq
326 database (release 92, 26553804) were downloaded. Proteins from assembly were
327 searched with HHMER v3.2.1 (<http://hmmer.org/>) for a match above the gathering
328 threshold of Pfam HMM profile 'phage_sheath_1' (PF04984) (El-Gebali et al., 2019).
329 For each match, up to 20 proteins were extracted from either side. All proteins from the
330 resulting set (phage sheath $\pm$ 20 proteins) were sorted by length and clustered at 50%
331 amino acid identity using UClust v1.2.22q (Edgar, 2010). Clusters containing $\geq$ 4
332 members were analyzed further. Cluster representatives were annotated using protein-
333 profile searches against three databases: the Pfam-A database using HMMER3 (El-
334 Gebali et al., 2019), the NCBI Conserved Domain Database using RPS-BLAST
335 (Marchler-Bauer et al., 2017, 2011; Marchler-Bauer and Bryant, 2004), and the
336 Uniprot30 database (accessed February 2019, available from
337 http://wwwuser.gwdg.de/~compbiol/data/hhsuite/databases/hhsuite_dbs/) using HHblits
338 (McGarvey et al., 2019; Remmert et al., 2012). Multiple sequence alignments were
339 automatically generated from three iterations of the HHblits search and used for profile-
340 profile comparisons against the PDB70 database (accessed February 2019, available
341 from http://wwwuser.gwdg.de/~compbiol/data/hhsuite/databases/hhsuite_dbs/).
342 Significant hits to cluster representatives were used to assign an annotation to all
343 proteins contained within the parent cluster. Manual inspection of *Bacteroides* and
344 *Parabacteroides* loci enabled consistent trimming of each genetic architecture;
345 specifically, the genes intervening DUF4255 and FtsH/ATPase were retained.

346

347 **Metagenomic mining analyses.**

348 To find the prevalence of the BIS genes within the Human Microbiome Project, we
 349 downloaded 11219 metagenomes using NCBI's fastq-dump API. The metagenomes
 350 were parsed where: left-right tags were clipped, technical reads (adapters, primers,
 351 barcodes, etc) were dropped, low quality reads were dropped, and paired reads were
 352 treated as two distinct reads. A subject database was created from the amino-acid
 353 sequences of the 18 BIS genes. Then the fastq files were piped through seqtk (Shen et
 354 al., 2016), to convert them to fasta format, which was then piped to DIAMOND via stdin.
 355 Then DIAMOND aligned the six-frame translation of the input reads against the subject
 356 database, with all default parameters. For each metagenome the number of non-
 357 mutually exclusive hits to each CIS gene were then summed providing a hit 'count
 358 score'. From the hit counts, a heatmap was created by taking the number of hits of each
 359 gene per metagenome and dividing by the total number of reads and multiplying the
 360 result by one million, which was then $\log_{10}(x+1)$ transformed. To estimate the co-
 361 occurrence between pairs of genes, the previous the hit count scores from the previous
 362 calculation were taken, and for each pair combination, the hit count of the lower of the
 363 two genes was added to a running total. The co-occurrence was then visualized on a
 364 network graph, where each edge corresponds to the number of times the pair of genes
 365 co-occurred in all the metagenomes (Csardi and Nepusz, 2006; Hadley, 2016) (R core
 366 Team 2017 - <https://www.R-project.org/>; ggraph - [https://CRAN.R-](https://CRAN.R-project.org/package=ggraph)
 367 [project.org/package=ggraph](https://CRAN.R-project.org/package=ggraph))

368

369 **Metatranscriptome mining analyses.**

370 Fastq files from transcriptomes were downloaded from the Sequence Read Archive
 371 using the sra toolkit (<https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/>). Low quality
 372 reads were removed using prinseq++ (<https://peerj.com/preprints/27553/>). Reads were
 373 compared to the amino acid sequences of the *Bacteroides cellulosilyticus* WH2 BIS
 374 protein cluster using blastX with an E-value cutoff of 0.001. The best hit for each read
 375 was kept. Hits to each protein were normalized by the number of reads of each
 376 transcriptome and the length of each protein using the program Fragment Recruitment
 377 Assembly Purification (<https://github.com/yinacobian/frap>).

378

379

380 ACKNOWLEDGEMENTS

381 We thank Dr. Martin Pilhofer for providing constructive comments on the manuscript. This
382 work was supported by the Office of Naval Research (N00014-17-1-2677, N.J.S. and
383 S.B.), Office of Naval Research (N00014-16-1-2135, N.J.S), Alfred P. Sloan Foundation,
384 Sloan Research Fellowship (N.J.S.), San Diego State University (N.J.S.), National
385 Science Foundation PIRE grant (OISE1243541, F.L.R.).

386

387 COMPETING INTERESTS

388 N.J.S and S.B have two provisional patents pending related to MACs in the United States,
389 Application Number: 62/768,240 and 62/844,988. The other authors declare that no
390 competing interests exist.

391

392 REFERENCES

- 393 Baudoin L, Issad T. 2014. O-GlcNAcylation and inflammation: A vast territory to explore.
394 *Front Endocrinol (Lausanne)* **5**:1–8. doi:10.3389/fendo.2014.00235
- 395 Böck D, Medeiros JM, Tsao H, Penz T, Weiss GL, Aistleitner K, Horn M, Pilhofer M.
396 2017. In situ architecture, function, and evolution of a contractile injection system
397 injection system. *Scien* **357**:713–717.
- 398 Carroll IM, Ringel-Kulka T, Keku TO, Chang YH, Packey CD, Balfour Sartor R, Ringel
399 Y. 2011. Molecular analysis of the luminal- and mucosal-associated intestinal
400 microbiota in diarrhea-predominant irritable bowel syndrome. *Am J Physiol -*
401 *Gastrointest Liver Physiol* **301**:799–807. doi:10.1152/ajpgi.00154.2011
- 402 Chatzidaki-Livanis M, Geva-Zatorsky N, Comstock LE. 2016. *Bacteroides fragilis* type
403 VI secretion systems use novel effector and immunity proteins to antagonize
404 human gut Bacteroidales species. *Proc Natl Acad Sci U S A* **113**:3627–32.
405 doi:10.1073/pnas.1522510113
- 406 Chen L, Song N, Liu B, Zhang N, Ali Khan NF, Zhou Z, Zhou Y, Zhou S, Zheng D, Chen
407 M, Hapeshi A, Healey J, Waterfield NR, Yang J, Yang G. 2019. Genome-wide
408 Identification and Characterization of a Superfamily of Bacterial Extracellular
409 Contractile Injection Systems. *Cell Rep* **29**:511-521.e2.
410 doi:10.1016/j.celrep.2019.08.096
- 411 Cianfanelli FR, Monlezun L, Coulthurst SJ. 2016. Aim, Load, Fire: The Type VI
412 Secretion System, a Bacterial Nanoweapon. *Trends Microbiol* **24**:51–62.
413 doi:10.1016/j.tim.2015.10.005
- 414 Consortium THMP, Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH,
415 Chinwalla AT, Creasy HH, Earl AM, FitzGerald MG, Fulton RS, Giglio MG,
416 Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny
417 DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK,
418 Zeng Q, Aagaard KM, Abolude OO, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen
419 GL, Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T,
420 Baker CC, Begg L, Belachew T, Bhonagiri V, Bihan M, Blaser MJ, Bloom T,
421 Bonazzi V, Paul Brooks J, Buck GA, Buhay CJ, Busam DA, Campbell JL, Canon
422 SR, Cantarel BL, Chain PSG, Chen I-MA, Chen L, Chhibba S, Chu K, Ciulla DM,
423 Clemente JC, Clifton SW, Conlan S, Crabtree J, Cutting MA, Davidovics NJ, Davis
424 CC, DeSantis TZ, Deal C, Delehaunty KD, Dewhirst FE, Deych E, Ding Y, Dooling
425 DJ, Dugan SP, Michael Dunne W, Scott Durkin A, Edgar RC, Erlich RL, Farmer
426 CN, Farrell RM, Faust K, Feldgarden M, Felix VM, Fisher S, Fodor AA, Forney LJ,
427 Foster L, Di Francesco V, Friedman J, Friedrich DC, Fronick CC, Fulton LL, Gao H,
428 Garcia N, Giannoukos G, Giblin C, Giovanni MY, Goldberg JM, Goll J, Gonzalez A,
429 Griggs A, Gujja S, Kinder Haake S, Haas BJ, Hamilton HA, Harris EL, Hepburn TA,
430 Herter B, Hoffmann DE, Holder ME, Howarth C, Huang KH, Huse SM, IZard J,
431 Jansson JK, Jiang H, Jordan C, Joshi V, Katancik JA, Keitel WA, Kelley ST, Kells
432 C, King NB, Knights D, Kong HH, Koren O, Koren S, Kota KC, Kovar CL, Kyrpides
433 NC, La Rosa PS, Lee SL, Lemon KP, Lennon N, Lewis CM, Lewis L, Ley RE, Li K,
434 Liolios K, Liu B, Liu Y, Lo C-C, Lozupone CA, Dwayne Lunsford R, Madden T,
435 Mahurkar AA, Mannon PJ, Mardis ER, Markowitz VM, Mavromatis K, McCorrison
436 JM, McDonald D, McEwen J, McGuire AL, McInnes P, Mehta T, Mihindukulasuriya
437 KA, Miller JR, Minx PJ, Newsham I, Nusbaum C, O’Laughlin M, Orvis J, Pagani I,

438 Palaniappan K, Patel SM, Pearson M, Peterson J, Podar M, Pohl C, Pollard KS,
439 Pop M, Priest ME, Proctor LM, Qin X, Raes J, Ravel J, Reid JG, Rho M, Rhodes R,
440 Riehle KP, Rivera MC, Rodriguez-Mueller B, Rogers Y-H, Ross MC, Russ C,
441 Sanka RK, Sankar P, Fah Sathirapongsasuti J, Schloss JA, Schloss PD, Schmidt
442 TM, Scholz M, Schriml L, Schubert AM, Segata N, Segre JA, Shannon WD, Sharp
443 RR, Sharpton TJ, Shenoy N, Sheth NU, Simone GA, Singh I, Smillie CS, Sobel JD,
444 Sommer DD, Spicer P, Sutton GG, Sykes SM, Tabbaa DG, Thiagarajan M,
445 Tomlinson CM, Torralba M, Treangen TJ, Truty RM, Vishnivetskaya TA, Walker J,
446 Wang L, Wang Z, Ward D V, Warren W, Watson MA, Wellington C, Wetterstrand
447 KA, White JR, Wilczek-Boney K, Wu Y, Wylie KM, Wylie T, Yandava C, Ye L, Ye Y,
448 Yooseph S, Youmans BP, Zhang L, Zhou Y, Zhu Y, Zoloth L, Zucker JD, Birren
449 BW, Gibbs RA, Highlander SK, Methé BA, Nelson KE, Petrosino JF, Weinstock
450 GM, Wilson RK, White O. 2012. Structure, function and diversity of the healthy
451 human microbiome. *Nature* **486**:207.

452 Coyne MJ, Roelofs KG, Comstock LE. 2016. Type VI secretion systems of human gut
453 Bacteroidales segregate into three genetic architectures, two of which are
454 contained on mobile genetic elements. *BMC Genomics* **17**:1–21.
455 doi:10.1186/s12864-016-2377-z

456 Csardi G, Nepusz T. 2006. The igraph software package for complex network research.
457 *InterJournal*.

458 Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST.
459 *Bioinformatics* **26**:2460–2461. doi:10.1093/bioinformatics/btq461

460 El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M,
461 Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L,
462 Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in
463 2019. *Nucleic Acids Res* **47**:D427–D432. doi:10.1093/nar/gky995

464 Ericson C, Eisenstein F, Medeiros J, Malter K, Cavalcanti G, Zeller R, Newman D,
465 Pilhofer M, Shikuma N. 2019. A contractile injection system stimulates tubeworm
466 metamorphosis by translocating a proteinaceous effector. *Elife* 1–19.
467 doi:https://doi.org/10.7554/eLife.46845

468 Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman
469 DA, Fraser-liggett CM, Nelson KE. 2006. Metagenomic Analysis of the Human
470 Distal Gut Microbiome **312**.

471 Gouy M, Guindon S, Gascuel O. 2010. Sea view version 4: A multiplatform graphical
472 user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*
473 **27**:221–224. doi:10.1093/molbev/msp259

474 Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New
475 algorithms and methods to estimate maximum-likelihood phylogenies: Assessing
476 the performance of PhyML 3.0. *Syst Biol* **59**:307–321. doi:10.1093/sysbio/syq010

477 Hadley W. 2016. ggplot2: Elegant Graphics for Data Analysis. New York: Springer.
478 doi:10.1002/wics.147

479 Heymann JB, Bartho JD, Rybakova D, Venugopal HP, Winkler DC, Sen A, Hurst MRH,
480 Mitra AK. 2013. Three-dimensional structure of the toxin-delivery particle
481 antifeeding prophage of *Serratia entomophila*. *J Biol Chem* **288**:25276–25284.
482 doi:10.1074/jbc.M113.456145

483 Hooper L V., Gordon JI. 2001. Commensal host-bacterial relationships in the gut.

484 *Science* (80-) **292**:1115–1118. doi:10.1126/science.1058709

485 Hurst MRH, Beard SS, Jackson TA, Jones SM. 2007. Isolation and characterization of
486 the *Serratia entomophila* antifeeding prophage. *FEMS Microbiol Lett* **270**:42–48.
487 doi:10.1111/j.1574-6968.2007.00645.x

488 Hurst MRH, Beattie A, Jones SA, Laugraud A, van Koten C, Harper L. 2018. *Serratia*
489 *proteamaculans* Strain AGR96X Encodes an Antifeeding Prophage (Tailocin) with
490 Activity against Grass Grub (*Costelytra giveni*) and Manuka Beetle (*Pyronota*
491 *Species*) Larvae. *Appl Environ Microbiol* **84**:e02739-17.

492 Hurst MRH, Glare TR, Jackson TA. 2004. Cloning *Serratia entomophila* antifeeding
493 genes - A putative defective prophage active against the grass grub *Costelytra*
494 *zealandica*. *J Bacteriol* **186**:7023–7024. doi:10.1128/JB.186.20.7023-7024.2004

495 Kau AL, Ahern PP, Griffin NW, Goodman AL, Gordon JI. 2011. Human nutrition, the gut
496 microbiome and the immune system. *Nature* **474**:327.

497 Lefort V, Longueville JE, Gascuel O. 2017. SMS: Smart Model Selection in PhyML. *Mol*
498 *Biol Evol* **34**:2422–2424. doi:10.1093/molbev/msx149

499 Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display
500 and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**:W242–W245.
501 doi:10.1093/nar/gkw290

502 Levi K, Abeysinghe E, Rynge M, Edwards RA. 2018. Searching the sequence read
503 archive using Jetstream and Wrangler. *ACM Int Conf Proceeding Ser* 1–7.
504 doi:10.1145/3219104.3229278

505 Ley RE, Turnbaugh PJ, Klein S, Gordon JI. 2006. Human gut microbes associated with
506 obesity. *Nature* **444**:1022–1023. doi:10.1038/4441022a

507 Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK,
508 Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS,
509 Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017.
510 CDD/SPARCLE: Functional classification of proteins via subfamily domain
511 architectures. *Nucleic Acids Res* **45**:D200–D203. doi:10.1093/nar/gkw1129

512 Marchler-Bauer A, Bryant SH. 2004. CD-Search: Protein domain annotations on the fly.
513 *Nucleic Acids Res* **32**:327–331. doi:10.1093/nar/gkh454

514 Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C,
515 Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke
516 Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko M V., Robertson
517 CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH.
518 2011. CDD: A Conserved Domain Database for the functional annotation of
519 proteins. *Nucleic Acids Res* **39**:225–229. doi:10.1093/nar/gkq1189

520 Marcobal A, Barboza M, Sonnenburg ED, Pudlo N, Martens EC, Desai P, Lebrilla CB,
521 Weimer BC, Mills DA, German JB, Sonnenburg JL. 2011. *Bacteroides* in the infant
522 gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host*
523 *Microbe* **10**:507–514. doi:10.1016/j.chom.2011.10.007

524 McGarvey PB, Nightingale A, Luo J, Huang H, Martin MJ, Wu C, Consortium UP. 2019.
525 UniProt genomic mapping for deciphering functional effects of missense variants.
526 *Hum Mutat* **40**:694–705. doi:10.1002/humu.23738

527 McNulty NP, Wu M, Erickson AR, Pan C, Erickson BK, Martens EC, Pudlo NA, Muegge
528 BD, Henrissat B, Hettich RL, Gordon JI. 2013. Effects of Diet on Resource
529 Utilization by a Model Human Gut Microbiota Containing *Bacteroides*

530 *cellulosilyticus* WH2, a Symbiont with an Extensive Glycobiome. *PLoS Biol* **11**.
531 doi:10.1371/journal.pbio.1001637
532 Myslicki JP, Belke DD, Shearer J. 2014. Role of O-GlcNAcylation in nutritional sensing,
533 insulin resistance and in mediating the benefits of exercise. *Appl Physiol Nutr*
534 *Metab* **39**:1205–1213. doi:10.1139/apnm-2014-0122
535 Pechenik JA. 1999. On the advantages and disadvantages of larval stages in benthic
536 marine invertebrate life cycles. *Mar Ecol Prog Ser* **177**:269–297.
537 doi:10.3354/meps177269
538 Penz T, Schmitz-Esser S, Kelly SE, Cass BN, Müller A, Woyke T, Malfatti SA, Hunter
539 MS, Horn M. 2012. Comparative Genomics Suggests an Independent Origin of
540 Cytoplasmic Incompatibility in *Cardinium hertigii*. *PLoS Genet* **8**.
541 doi:10.1371/journal.pgen.1003012
542 Qin J, Li R, Raes J, Arumugam M, Burgdorf SK, Manichanh C, Nielsen T, Pons N,
543 Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H,
544 Zheng H. 2010. A human gut microbial gene catalogue established by
545 metagenomic sequencing. *Nature* **464**:59–65. doi:10.1038/nature08821
546 Remmert M, Biegert A, Hauser A, Söding J. 2012. HHblits: Lightning-fast iterative
547 protein sequence searching by HMM-HMM alignment. *Nat Methods* **9**:173–175.
548 doi:10.1038/nmeth.1818
549 Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V,
550 Henrissat B, Bain JR, Muehlbauer MJ, Ilkayeva O, Semenkovich CF, Funai K,
551 Hayashi DK, Lyle BJ, Martini MC, Ursell LK, Clemente JC, Van Treuren W, Walters
552 W a, Knight R, Newgard CB, Heath AC, Gordon JL. 2013. Gut microbiota from twins
553 discordant for obesity modulate metabolism in mice. *Science* **341**:1241214.
554 doi:10.1126/science.1241214
555 Rocchi I, Ericson CF, Malter KE, Zargar S, Eisenstein F, Pilhofer M, Beyhan S, Shikuma
556 NJ. 2019. A Bacterial Phage Tail-like Structure Kills Eukaryotic Cells by Injecting a
557 Nuclease Effector. *Cell Rep* **28**:295-301.e4. doi:10.1016/j.celrep.2019.06.019
558 Russell AB, Wexler AG, Harding BN, Whitney JC, Bohn AJ, Goo YA, Tran BQ, Barry
559 NA, Zheng H, Peterson SB, Chou S, Gonen T, Goodlett DR, Goodman AL,
560 Mougous JD. 2014. A type VI secretion-related pathway in *Bacteroidetes* mediates
561 interbacterial antagonism. *Cell Host Microbe* **16**:227–236.
562 doi:10.1016/j.chom.2014.07.007
563 Salmond GPC, Fineran PC. 2015. A century of the phage: past, present and future. *Nat*
564 *Rev Microbiol* **13**:777–86. doi:10.1038/nrmicro3564
565 Sarris PF, Ladoukakis ED, Panopoulos NJ, Scoulica E V. 2014. A phage tail-derived
566 element with wide distribution among both prokaryotic domains: A comparative
567 genomic and phylogenetic study. *Genome Biol Evol* **6**:1739–1747.
568 doi:10.1093/gbe/evu136
569 Shen W, Le S, Li Y, Hu F. 2016. SeqKit: A cross-platform and ultrafast toolkit for
570 FASTA/Q file manipulation. *PLoS One* **11**:1–10. doi:10.1371/journal.pone.0163962
571 Shikuma NJ, Antoshechkin I, Medeiros JM, Pilhofer M, Newman DK. 2016. Stepwise
572 metamorphosis of the tubeworm *Hydroides elegans* is mediated by a bacterial
573 inducer and MAPK signaling. *Proc Natl Acad Sci* **113**:10097–10102.
574 doi:10.1073/pnas.1603142113
575 Shikuma NJ, Pilhofer M, Weiss GL, Hadfield MG, Jensen GJ, Newman DK. 2014.

576 Marine Tubeworm Metamorphosis Induced by Arrays of Bacterial Phage Tail-Like
577 Structures. *Science* (80-) **343**:529–533. doi:10.1126/science.1246794
578 Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, Benezra A,
579 Destefano J, Meier MF, Muegge BD, Barratt MJ, VanArendonk LG, Zhang Q,
580 Province MA, Petri WA, Ahmed T, Gordon JI. 2014. Persistent gut microbiota
581 immaturity in malnourished Bangladeshi children. *Nature* **510**:417–421.
582 doi:10.1038/nature13421
583 Tropini C, Earle KA, Huang KC, Sonnenburg JL. 2017. The Gut Microbiome:
584 Connecting Spatial Organization to Function. *Cell Host Microbe* **21**:433–442.
585 doi:10.1016/j.chom.2017.03.010
586 Van Der Waaij LA, Harmsen HJM, Madjipour M, Kroese FGM, Zwieters M, Van Dullemen
587 HM, De Boer NK, Welling GW, Jansen PLM. 2005. Bacterial population analysis of
588 human colon and terminal ileum biopsies with 16S rRNA-based fluorescent probes:
589 Commensal bacteria live in suspension and have no direct contact with epithelial
590 cells. *Inflamm Bowel Dis* **11**:865–871. doi:10.1097/01.mib.0000179212.80778.d3
591 Verster AJ, Ross BD, Radey MC, Bao Y, Goodman AL, Mougous JD, Borenstein E.
592 2017. The Landscape of Type VI Secretion across Human Gut Microbiomes
593 Reveals Its Role in Community Composition. *Cell Host Microbe* **22**:411–419.e4.
594 doi:10.1016/j.chom.2017.08.010
595 Vlisidou I, Hapeshi A, Healey JRJ, Smart K, Yang G, Waterfield NR. 2019. The
596 photorhabdus asymbiotica virulence cassettes deliver protein effectors directly into
597 target eukaryotic cells. *Elife* **8**:1–24. doi:10.7554/eLife.46259
598 Yang G, Dowling AJ, Gerike U, Ffrench-Constant RH, Waterfield NR. 2006.
599 *Photorhabdus* Virulence Cassettes Confer Injectable Insecticidal Activity against
600 the Wax Moth. *J Bacteriol* **188**:2254–2261. doi:10.1128/JB.188.6.2254-2261.2006
601
602
603

604 SUPPLEMENTARY MATERIAL

605

606 TABLE S1.

607

	Protein	Locus tag	<i>B. cell.</i> WH2 CIS protein	<i>B. cell.</i> WH2 Locus tag	Coverage	e-value	Identity
<i>P. luteo</i> MACs	MacB	WP_029427202.1	baseplate	WP_029427202.1	84%	5.00E-32	22%
<i>C. hertigii</i> T6SS	baseplate	WP_014934193.1			85%	2.00E-57	23%
<i>S. entomophila</i> Afp	baseplate	WP_010895813.1			69%	7.00E-18	25%
<i>P. asymbiotica</i> PVCs	baseplate	WP_015834374.1			68%	3.00E-16	25%
<i>A. asiaticus</i> T6SS ^{iv}	baseplate	WP_012472726.1			77%	2.00E-67	24%
<i>P. luteo</i> MACs	MacS	WP_039609824.1	Sheath1	WP_029427210.1	99%	6.00E-94	35%
<i>C. hertigii</i> T6SS	Sheath	WP_014934609.1			89%	8.00E-65	50%
<i>S. entomophila</i> Afp	Sheath1	WP_010895805.1			87%	9.00E-49	45%
<i>P. asymbiotica</i> PVCs	Sheath1	WP_015835470.1			77%	4.00E-54	46%
<i>A. asiaticus</i> T6SS ^{iv}	Sheath	WP_012473177.1			91%	5.00E-66	50%
<i>P. luteo</i> MACs	MacS	WP_039609824.1	Sheath2	WP_029427209.1	85%	1.00E-63	52%
<i>C. hertigii</i> T6SS	Sheath	WP_014934609.1			97%	7.00E-122	39%
<i>S. entomophila</i> Afp	Sheath2	WP_010895804.1			94%	1.00E-51	48%
<i>P. asymbiotica</i> PVCs	Sheath2	WP_015835471.1			97%	3.00E-54	47%
<i>A. asiaticus</i> T6SS ^{iv}	Sheath	WP_012473177.1			97%	3.00E-114	40%
<i>P. luteo</i> MACs	MacT1	WP_039609825.1	Tube1	WP_007212392.1	97%	5.00E-23	35%
<i>C. hertigii</i> T6SS	Tube	WP_014934612.1			73%	6.00E-05	21%
<i>S. entomophila</i> Afp	Tube	WP_010895803.1			90%	7.00E-32	34%
<i>P. asymbiotica</i> PVCs	Tube1	WP_015835472.1			95%	4.00E-31	33%
<i>A. asiaticus</i> T6SS ^{iv}	Tube	WP_012473180.1			84%	3.00E-08	22%
<i>P. luteo</i> MACs	MacT2	WP_039609826.1	Tube2	WP_007212393.1	92%	6.00E-01	25%
<i>C. hertigii</i> T6SS	Tube	WP_014934612.1			57%	1.00E-03	24%
<i>S. entomophila</i> Afp	Tube	WP_010895803.1			77%	9.00E-09	24%
<i>P. asymbiotica</i> PVCs	Tube2	WP_015834383.1			93%	2.00E-09	23%
<i>A. asiaticus</i> T6SS ^{iv}	Tube	WP_012473180.1			92%	3.00E-09	26%

608

609

610

611 **TABLE S2.**

Species	Strain	Genome Accession	Sheath Protein Accession	Tube Protein Accession
<i>Candidatus Amoebophilus asiaticus</i>	5a2	CP001102.1	WP_012473177.1	WP_012473180.1
<i>Bacteroides cellulosilyticus</i>	WH2	CP012801.1	WP_029427210.1	WP_007212392.1
<i>Bacteroides fragilis</i>	BFBE1.1 (BE1)	LN877293.1	WP_005803145.1, WP_053873779.1	WP_005803146.1
<i>Cardinium hertigii</i>	cEper1	NC_018605.1	WP_014934609.1	WP_014934612.1
<i>Enterobacteria phage P2</i>		NC_041848.1	YP_009591452.1	YP_009591453.1
<i>Enterobacteria phage T4</i>		NC_000866.4	NP_049780.1	WP_015969329.1
<i>Flavobacterium johnsoniae</i>	UW101	NC_009441.1	WP_012025137.1, WP_012025251.1	WP_012025138.1
<i>Francisella tularensis subsp. tularensis</i>	SCHU S4	AJ749949.2	WP_003023948.1	WP_003022149.1
<i>Parabacteroides sp.</i>	D25	NZ_JH976500.1	WP_009276514.1, WP_008669225.1	WP_005861441.1
<i>Pseudoalteromonas luteoviolacea</i>	HI1	KF724687.1	WP_039609824.1	WP_039609825.1
<i>Pseudomonas aeruginosa</i>	PAO1	NC_002516.2	WP_003113197.1, WP_003087596.1	WP_003083317.1, WP_003085175.1
<i>Salmonella enterica subsp. enterica serovar Typhi</i>	CT18	NC_003198.1	WP_000046142.1, WP_000013884.1	WP_000338756.1, WP_001207653.1
<i>Serratia entomophila (pADAP)</i>	A1MO2	NC_002523.4	WP_010895805.1	WP_010895803.1
<i>Vibrio cholerae O1 biovar El Tor</i>	N16961	NC_002506.1	WP_001882966.1	WP_001142947.1
<i>Photobacterium asymbiotica</i>	ATCC43949	NC_012962.1	WP_015835470.1	WP_015835472.1

612

613

614

615 **TABLE S3.**

616

Accession	Species	Strain	start	stop	strand	arch.
NZ_ATFI01000004.1	<i>Bacteroides cellulosilyticus</i> WH2	WH2	271076	294193	-	1
NZ_QRXS01000004.1	<i>Bacteroides cellulosilyticus</i>	AF17-25	276218	298574	-	1
NZ_QRVJ01000001.1	<i>Bacteroides cellulosilyticus</i>	AF22-3AC	604020	580904	+	1
NZ_QRSV01000007.1	<i>Bacteroides cellulosilyticus</i>	AF29-17	269200	292317	-	1
NZ_LRGD01000010.1	<i>Bacteroides cellulosilyticus</i>	CL09T06C25	44848	67991	-	1
NZ_JH724085.1	<i>Bacteroides cellulosilyticus</i> CL02T12C19	CL02T12C19	94259	117409	+	1
NZ_EQ973491.1	<i>Bacteroides cellulosilyticus</i> DSM 14838	DSM 14838	290576	313959	+	1
NZ_KQ968695.1	<i>Bacteroides intestinalis</i>	KLE1704	249385	272510	-	1
NZ_QSUL01000020.1	<i>Bacteroides oleiciplenus</i>	OM05-15BH	29099	52460	+	1
NZ_JH992940.1	<i>Bacteroides oleiciplenus</i> YIT 12058	YIT 12058	1533689	1555212	-	1
NZ_JAGH01000002.1	<i>Bacteroides</i> sp. 14(A)	14(A)	2478682	2501030	+	1
NZ_RAZN01000003.1	<i>Parabacteroides goldsteinii</i>	0.1X-D42-15	28300	49757	+	1
NZ_JH976474.1	<i>Parabacteroides goldsteinii</i> CL02T12C30	CL02T12C30	486316	507720	+	1
NZ_KE159513.1	<i>Parabacteroides goldsteinii</i> dnLKV18	dnLKV18	1159956	1181413	-	1
NZ_QSWF01000006.1	<i>Parabacteroides gordonii</i>	OM02-37	268051	289934	-	1
NZ_KE386763.1	<i>Parabacteroides gordonii</i> DSM 23371	DSM 23371	428020	449720	+	1
NZ_KQ033920.1	<i>Parabacteroides gordonii</i> MS-1	MS-1	770218	791918	-	1
NZ_QUHH01000016.1	<i>Parabacteroides</i> sp. AF14-59	AF14-59	19676	38492	+	1
NZ_QTMM01000005.1	<i>Parabacteroides</i> sp. AF17-3	AF17-3	296338	317796	-	1
NZ_QUGR01000001.1	<i>Parabacteroides</i> sp. AF18-52	AF18-52	476916	495726	-	1
NZ_QUDI01000023.1	<i>Parabacteroides</i> sp. AF48-14	AF48-14	70961	91946	+	1
NZ_KB822571.1	<i>Parabacteroides</i> sp. ASF519	ASF519	5930811	5952268	-	1
NZ_KQ033902.1	<i>Parabacteroides</i> sp. HGS0025	HGS0025	2808277	2829908	+	1
NZ_LT669941.1	<i>Parabacteroides timonensis</i>	Marseille-P3236	862055	883911	-	1
NZ_LN877293.1	<i>Bacteroides fragilis</i>	BE1	3202545	3220402	-	2
NZ_QRZO01000002.1	<i>Bacteroides fragilis</i>	AF14-14AC	123788	141668	-	2
NZ_QRZH01000002.1	<i>Bacteroides fragilis</i>	AF14-26	126212	144092	-	2
NZ_AKBY01000003.1	<i>Bacteroides fragilis</i> CL05T00C42	CL05T00C42	80558	98415	-	2
NZ_JH724200.1	<i>Bacteroides fragilis</i> CL05T12C13	CL05T12C13	80522	98379	-	2
NZ_JGDN01000068.1	<i>Bacteroides fragilis</i> str. 3397 N2	3397 N2	69183	87040	-	2
NZ_JGEG01000075.1	<i>Bacteroides fragilis</i> str. 3397 N3	3397 N3	69898	87755	-	2
NZ_JGDO01000041.1	<i>Bacteroides fragilis</i> str. 3397 T14	3397 T14	2229	20086	-	2
NZ_JH976506.1	<i>Parabacteroides</i> sp. D25	D25	344282	355992	+/-	3
NZ_GG705151.1	<i>Bacteroides</i> sp. 2_1_33B	2_1_33B	811723	828493	+/-	3
NZ_CYXP01000003.1	<i>Parabacteroides distasonis</i>	2789STDY5608872	797	17567	+/-	3
NZ_CZAR01000002.1	<i>Parabacteroides distasonis</i>	2789STDY5834901	1098	17868	+/-	3
NZ_CZBM01000003.1	<i>Parabacteroides distasonis</i>	2789STDY5834948	348507	365278	+/-	3
NZ_QRXX01000024.1	<i>Parabacteroides distasonis</i>	AF18-10	690	12400	+/-	3
NZ_QRPA01000023.1	<i>Parabacteroides distasonis</i>	AF36-3	57479	74255	+/-	3
NZ_NFJX01000005.1	<i>Parabacteroides distasonis</i>	An199	298514	315279	+/-	3

NZ_NNCA01000001.1	<i>Parabacteroides distasonis</i>	CBA7138	2878268	2889978	+/-	3
NZ_JH976489.1	<i>Parabacteroides distasonis</i> CL09T03C24	CL09T03C24	423568	440338	+/-	3
NZ_JNHN01000089.1	<i>Parabacteroides distasonis</i> str. 3776 D15 i	3776 D15 i	3655	25273	+/-	3
NZ_JNHU01000057.1	<i>Parabacteroides distasonis</i> str. 3776 D15 iv	3776 D15 iv	236310	253075	+/-	3
NZ_JNHL01000054.1	<i>Parabacteroides distasonis</i> str. 3776 Po2 i	3776 Po2 i	82522	99287	+/-	3
NZ_KQ236096.1	<i>Parabacteroides</i> sp. 2_1_7	2_1_7	1621793	1633503	+/-	3
NZ_QSQY01000029.1	<i>Parabacteroides</i> sp. 20_3	TF09-4	28656	45426	+/-	3
NZ_QSQL01000017.1	<i>Parabacteroides</i> sp. 20_3	TF12-11	628	17398	+/-	3
NZ_QTMJ01000002.1	<i>Parabacteroides</i> sp. AF19-14	AF19-14	758	17529	+/-	3
NZ_QTMG01000015.1	<i>Parabacteroides</i> sp. AF21-43	AF21-43	616	17386	+/-	3
NZ_QTLZ01000003.1	<i>Parabacteroides</i> sp. AF27-14	AF27-14	356010	372781	+/-	3
NZ_QTLP01000005.1	<i>Parabacteroides</i> sp. AF39-10AC	AF39-10AC	654	17424	+/-	3
NZ_QTLI01000013.1	<i>Parabacteroides</i> sp. AM17-47	AM17-47	642	17412	+/-	3
NZ_QTLD01000037.1	<i>Parabacteroides</i> sp. AM25-14	AM25-14	1411	18181	+/-	3
NZ_RAYG01000046.1	<i>Parabacteroides</i> sp. CH2-D42-20	CH2-D42-20	801	17566	+/-	3
NZ_KQ236106.1	<i>Parabacteroides</i> sp. D26	D26	373154	389924	+/-	3
NZ_QTMX01000003.1	<i>Parabacteroides</i> sp. OF01-14	OF01-14	867	12577	+/-	3
NZ_LIDT01000035.1	<i>Bacteroides fragilis</i>	20793-3	14633	36151	-	Tn
NZ_PDCT01000010.1	<i>Bacteroides fragilis</i>	CM13	14711	36229	-	Tn
NZ_QTLE01000029.1	<i>Bacteroides</i> sp. AM23-18	AM23-18	24670	49228	+	Tn
NZ_QRZK01000006.1	<i>Bacteroides thetaiotaomicron</i>	AF14-20	31752	54562	+	Tn
NZ_QROV01000008.1	<i>Bacteroides thetaiotaomicron</i>	AF37-12	31743	54552	+	Tn
NZ_QRKS01000015.1	<i>Bacteroides thetaiotaomicron</i>	AM15-10	56434	79243	+	Tn
NZ_QSLC01000002.1	<i>Bacteroides thetaiotaomicron</i>	AM26-17LB	486821	509631	-	Tn
NZ_FOAL01000016.1	<i>Bacteroides thetaiotaomicron</i>	KPPR-3	50034	72843	-	Tn
NZ_JH976467.1	<i>Parabacteroides johnsonii</i> CL02T12C29	CL02T12C29	152432	172162	-	Tn

617

618

619 **Table S4.**

Locus Tag (IMG: GOLD Analysis Project ID)	KEGG ID	Protein ID	Gene annotation	T4 homolog	pADAP homolog
Ga0123724_112369	BcellWH2_03967	>WP_029427211.1	DUF4255	-	Afp16
Ga0123724_112370	BcellWH2_03968	>WP_029427210.1	Hypothetical protein	-	-
Ga0123724_112371	BcellWH2_03969	>WP_029427209.1	Sheath1	gp18	Afp2/3/4
Ga0123724_112372	BcellWH2_03970	>WP_007212392.1	Sheath2	gp18	Afp2/3/4
Ga0123724_112373	BcellWH2_03971	>WP_007212393.1	Tube1	gp19	Afp1/5
Ga0123724_112374	BcellWH2_03972	>WP_007212394.1	Tube2	gp19	Afp1/5
Ga0123724_112375	BcellWH2_03973	>WP_029427212.1	Hypothetical protein	-	Afp6
Ga0123724_112376	BcellWH2_03974	>WP_029427208.1	LysM	gp6	Afp7
Ga0123724_112377	BcellWH2_03975	>WP_029427207.1	Spike	gp5	Afp8
Ga0123724_112378	BcellWH2_03976	>WP_029427206.1	Tip	gp5.4	Afp10
Ga0123724_112379	BcellWH2_03977	>WP_029427205.1	Baseplate	gp25	Afp9
Ga0123724_112380	BcellWH2_03978	>WP_029427203.1	Hypothetical protein	-	-
Ga0123724_112381	BcellWH2_03979	>WP_029427202.1	Baseplate	gp27	Afp11
Ga0123724_112382	BcellWH2_03980	>WP_029427201.1	Hypothetical protein	-	Afp13
Ga0123724_112383	BcellWH2_03981	>WP_029427200.1	Baseplate	gp6	Afp12
Ga0123724_112384	BcellWH2_03982	>WP_029427199.1	Hypothetical protein	-	Afp14
Ga0123724_112385	BcellWH2_03983	>WP_029427198.1	DUF4157	-	-
Ga0123724_112386	BcellWH2_03984	>WP_007215181.1	FtsH/ATPase	-	Afp15

620

621

Figure S1. Unrooted phylogeny of CIS tube protein sequences. BIS group with known T6SS-iv and CIS (orange) and are distinct from known T6SSⁱ and T6SSⁱⁱ present in human pathogens, and known T6SSⁱⁱⁱ, characterized to mediate bacteria-bacteria interactions.

Figure S2. BIS genes are expressed during *in vitro* culture of *B. cellulosilyticus* WH2. Relative abundance of RNA hits to 18 major genes of the BIS in *B. cellulosilyticus* WH2 culture in MM supplemented with 31 different simple and complex sugars.