Large-scale chemical-genetics of the human gut bacterium *Bacteroides thetaiotaomicron*

Hualan Liu[1], Morgan N. Price[1], Hans K. Carlson[1], Yan Chen[2], Jayashree Ray[1], Anthony L. Shiver[3], Christopher J. Petzold[2], Kerwyn Casey Huang[3,4,5], Adam P. Arkin[1,6], Adam M. Deutschbauer[1,7,*]

[1]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[2]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[3]Department of Bioengineering, Stanford University, Stanford, CA, USA

[4]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA

[5]Chan Zuckerberg Biohub, San Francisco, CA, USA

[6]Department of Bioengineering, University of California, Berkeley, CA, USA

[7]Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA

[*]To whom correspondence should be addressed: AMDeutschbauer@lbl.gov

Abstract

21

22 The genomic catalogue of the human microbiota has expanded dramatically in recent years,
23 and insights derived from human microbiota genomics has vast potential to generate treatments
24 for human diseases. However, predictably harnessing the microbiota for beneficial outcomes is
25 currently limited by our lack of understanding of the physiology of the constituent bacteria. For
26 instance, the functions of most of their genes are not known. Here, we systematically measure
27 mutant phenotypes for genes from the gut commensal *Bacteroides thetaiotaomicron*. Using a
28 barcoded transposon mutant library, we measured the fitness of 4,055 *B. thetaiotaomicron*
29 genes across 492 experiments, including growth on 45 carbon substrates and in the presence
30 of 57 stress-inducing compounds. Our data is in strong agreement with previous studies, and
31 more importantly also uncovers the biological roles of poorly annotated genes. We identified
32 497 genes with a specific phenotype in only one or a handful of conditions, thus enabling
33 informed predictions of gene function for a subset of these genes. For example, we identified a
34 glycoside hydrolase important for growth on type I rhamnogalacturonan, a DUF4861 protein for
35 glycosaminoglycan utilization, a DUF1080 protein for disaccharide utilization, and a tripartite
36 multidrug resistance system specifically important for bile salt tolerance. Our approach can be
37 applied to other members of the human microbiota to experimentally characterize their genes.

Introduction

Since the United States National Institutes of Health launched the Human Microbiome Project in 2008 [1], our understanding of the human microbiota has advanced significantly. It is now established that the composition and diversity of the gut microbiota play important roles in human health and disease [2,3]. With these advances, it is now anticipated that manipulation of the human microbiota can be used to treat or manage a number of diseases [4]. However, these efforts will likely be limited until we have a greater understanding of the activity, physiology, and interactions of the individual bacteria that comprise the human microbiota. One primary bottleneck in our understanding of the human microbiota is that approximately half (40-70%) of the protein-coding genes in the human microbiota do not yet have a predicted molecular function [5]. Moreover, of the remaining genes, very few have experimentally determined functions; most predictions of gene function are derived by homology to genes characterized in other species from outside the human microbiota, which are often too distant to be accurate [6,7].

One of the most well studied constituents of the human microbiota is the Gram-negative gut anaerobe *Bacteroides thetaiotaomicron*, which is prevalent in a substantial fraction of the world's population [8]. Pioneering work on the human microbiota revealed the broad range of complex polysaccharides and simple carbohydrates that *B. thetaiotaomicron* can degrade and ferment in the gut [9-11]. Subsequently, analyses enabled by the full genome sequence of *B. thetaiotaomicron* provided new insights into gene regulation and carbon utilization [12-14]. It was found that many carbon degradation systems are encoded within Polysaccharide Utilization Loci (PULs), which typically contain a cluster of genes involved in the regulation, transport, and catabolism of various polysaccharides [15]. The most well studied PUL in *B. thetaiotaomicron* is the starch utilization system (Sus), encoded by *susRABCDEFG* [16]. SusR is the sensor/regulator, SusCDEF are membrane proteins responsible for starch binding and transport, and SusABG are hydrolytic enzymes. With the minimal requirement of a *susCD* gene pair, genome-wide computational analysis has predicted more than 80 PULs in the *B. thetaiotaomicron* genome [17], but the majority of these are experimentally uncharacterized.

Although traditional single-gene genetics and classical enzymology are useful for elucidating protein function, they can be time consuming and labor intensive. As such, it is challenging to apply a "one gene at a time" approach to characterize hundreds to thousands of bacterial genes. Thus, large-scale functional genomic methods are attractive for the systematic characterization of genes without a currently known function [18]. Methods based on transposon site sequencing (TnSeq) are useful for assaying the importance of thousands of genes in parallel under multiple conditions [19], including for human commensals *in vivo*. For example, the transposon insertion sequencing (INSeq) approach was developed in *B. thetaiotaomicron* and revealed genes that were required for colonizing and surviving in mice [20]. Subsequent application of INSeq to multiple *Bacteroides* species identified species-specific genes important for *in vivo* activity [21]. The development of large, multi-condition gene phenotype datasets has proven utility for inferring gene function in multiple organisms [7,22-24]. However, genome-wide fitness assays have yet to be performed systematically across many conditions to identify condition-specific phenotypes in *B. thetaiotaomicron*.

81

82   In this work, we describe the use of a barcoded variant of TnSeq (RB-TnSeq) to generate a

83   large genetic dataset for *B. thetaiotaomicron.* In RB-TnSeq, the incorporation of random DNA

84   barcodes simplifies the assessment of strain and gene fitness in a pooled, competitive assay [25],

85   and enables the systematic identification of mutant phenotypes across many conditions [7], thus

86   massively decreasing the time and effort required for large-scale genetic characterization. Using

87   RB-TnSeq, we performed hundreds of genome-wide fitness assays in *B. thetaiotaomicron,*

88   including in the presence of 113 different compounds such as simple sugars, complex

89   polysaccharides, and antibiotics. This rich dataset provides experimental evidence for many

90   computationally predicted gene functions in *B. thetaiotaomicron*, and also links genes that

91   previously lacked informative annotations to specific processes including polysaccharide

92   degradation, disaccharide catabolism, and bile salt tolerance.

93

Results

95

A gene-phenotype map of *B. thetaiotaomicron*

97   To identify an effective transposon delivery vector for *B. thetaiotaomicron* VPI-5482, we first

98   tested the mutagenesis efficiency of hundreds of different vectors in parallel. This approach,

99   which we term "magic pools" [26], enabled us to design a *mariner* transposon vector that was

100   suitable for constructing a genome-wide library with minimal insertion biases (Methods). The

101   final mutant library contains 316,025 uniquely barcoded strains with insertions that we could

102   confidently map to a single location in the *B. thetaiotaomicron* genome. These transposon

103   insertions are evenly distributed across the chromosome (Figure 1A), and are equally likely to

104   be on the coding and non-coding strands of the mutated genes (of insertions within protein-

105   coding genes, 49.4% had the drug marker on the coding strand). To accelerate follow-up

106   studies, we also generated an archived collection of individual transposon insertion mutants in

107   96-well microplate format and mapped the identity of the majority of these strains (A.L.S, A.M.D,

108   K.C.H; in review).

109

110   To validate the utility of the *B. thetaiotaomicron* mutant library for genome-wide fitness assays,

111   we grew the pooled library in a defined growth medium with either glucose or xylose as the sole

112   carbon source. For each experiment, we measured the abundance of the DNA barcodes by

113   sequencing (BarSeq) before and after growth selection, and we defined the fitness for each

114   gene as the $\log_2$ change in abundance of mutants of the gene (Figure 1B). Negative and

115   positive values mean that the mutants in that gene were less and more fit, respectively

116   compared to the average strain. Many genes that are predicted to be involved in amino acid

117   biosynthesis were found to be important for fitness with either carbon source, because the

118   defined media lacked most of the amino acids. Genes from the predicted xylose utilization

119   pathway, BT0792 (xylulose kinase), BT0793 (xylose isomerase) and BT0794 (xylose

120   transporter), were important for growth on xylose only (Figure 1B). This simple example

121   illustrates the biological consistency of fitness assays using our mutant library.

122

123   The primary advantage of a barcoded transposon mutant library is that each genome-wide
124   fitness assay only requires quantification of the DNA barcodes with BarSeq, which is extremely
125   scalable [25]. This feature of our method allowed us to perform hundreds of genome-wide fitness
126   assays to create a large gene-phenotype compendium for *B. thetaiotaomicron*. Many of these
127   experiments were assays in defined media (Varel-Bryant (VB)) with a single carbon source for
128   the bacteria to consume. In addition, we performed hundreds of competitive growth experiments
129   with an inhibitory but sublethal concentration of a stressor such as an antibiotic, biocide, or bile
130   salt. Overall, we performed 492 genome-wide fitness assays that passed our metrics for internal
131   consistency [25] (Methods). All analyses in this manuscript are derived from these 492
132   experiments, which represent fitness data in the presence of 113 unique compounds. All data is
133   available online for interactive analyses at the Fitness Browser (http://fit.genomics.lbl.gov).

135   We successfully measured fitness for 4,055 protein-coding genes in each of these experiments,
136   with the fitness for the median gene calculated from 24 independent transposon insertions.
137   Among the genes with no fitness data, we estimate that 378 of them are likely essential for
138   viability in the Brain Heart Infusion Supplemented (BHIS) media we used to select the mutants
139   (Supplementary Table 1). Some non-essential genes may not have fitness data because they
140   are too short, repetitive, or their mutants are at a low abundance because they are nearly
141   essential for viability.

143   Across the entire dataset, we identified 497 different genes with a specific phenotype
144   (Supplementary Table 2), defined as |fitness| > 1 and statistically significant in an experiment
145   but lacking a phenotype in most other experiments (Methods). A specific phenotype often links a
146   gene to a condition that is directly related to its function [7]. To illustrate this principle, we present
147   examples of specific phenotypes for BT3788 and BT4507. BT3788 is a SusC homolog located
148   in the PUL for mannan utilization [27]. Across all 45 different carbon sources we profiled, BT3788
149   was only important for fitness on mannan (Figure 1C). BT4507 encodes a class A serine β-
150   lactamase [28], and we found that mutants in this gene were only impaired in the presence of two
151   β-lactam antibiotics, ceftriaxone and cephalothin, but not in any of the other conditions that we
152   profiled, including dozens of other antibiotics (Figure 1D).

154   A second approach for inferring gene function from large-scale genetics data is cofitness, where
155   a pair of genes from the same organism have very similar fitness profiles across all conditions [7].
156   To illustrate the utility of cofitness to infer gene function in *B. thetaiotaomicron*, we investigated
157   genes important for arginine biosynthesis. The predicted operon BT3758-3762 encodes several
158   homologs to genes in the *Escherichia coli* arginine biosynthesis pathway [29], except that BT3761
159   is not closely related to an *E. coli* enzyme (by BLAST), but rather shares homology to
160   Cabys_1732 from *Caldithrix abyssi*. These two proteins share 58% identity with 88% sequence
161   coverage. Cabys_1732 is annotated as a newly characterized isoform of N-acetylglutamate
162   synthase [30]. From our data, we found that BT3761 has very high cofitness (>0.92) with the other
163   four genes in the same putative operon, and it has the highest cofitness (0.97) with BT3733
164   (Figure 1E), which is a homolog of *E. coli* argininosuccinate lyase. Based on these data,
165   BT3761 likely encodes an N-acetylglutamate synthase required for arginine biosynthesis. A full
166   list of genes with high cofitness is available in Supplementary Table 3.

167

168   Overall, we identified a functional association (either a specific phenotype or cofitness) for 612
169   of the 4,055 protein coding genes for which we collected data. Genes with a functional
170   association were more likely to have a specific COG functional category assigned compared to
171   genes with no functional association (58% vs. 38%; $P < 10^{-15}$, Fisher's exact test) [31].
172   Nonetheless, we were able to identify a functional association for 256 genes without a specific
173   COG function category. According to the PaperBLAST database [32], 193 of these 256 genes
174   have not been previously discussed in the literature, nor have close homologs (at >70% amino
175   acid identity) from other *Bacteroides* species.

176

177

Linking PULs to specific substrates with genetics

178
179   To associate the PULs of *B. thetaiotaomicron* to their substrates, we analyzed the fitness data
180   collected by growing the mutant library in defined media with various compounds added as the
181   sole source of carbon. The 45 carbon sources we successfully assayed represent a range of
182   substrate complexity and include 15 monosaccharides and their derivatives, 10 disaccharides, 1
183   trisaccharide, 2 tetrasaccharides, 3 oligosaccharides (>4 monomers), and 14 complex
184   polysaccharides. The complex polysaccharides we assayed are typically derived from natural
185   sources (for example, starch from potato) and are likely more heterogeneous than the simpler
186   sugars; we note the source of each compound in Supplementary Table 4. Lastly, because we
187   performed all of our genome-wide fitness assays in pools, we cannot rule out the possibility of
188   cross-feeding between strains, with the potential implication that some genes will have no
189   phenotype in our data despite their importance for growth on a particular substrate.

190
191   Fitness profiling across a range of substrate complexity provides a broad view of the importance
192   of each PUL component to the catabolism of a complex polysaccharide. As an illustration, we
193   re-examined the canonical Sus system for starch degradation by comparing the phenotypes of
194   these genes during growth on starch, the component polysaccharides of starch (amylopectin
195   and amylose), the oligosaccharide α-cyclodextrin, and the simpler starch break-down products
196   glucose, maltose, maltotetraose, and maltohexaose (Figure 2). The SusCD transporter proteins
197   were important for fitness on all saccharides of length 4 or more, but largely dispensable for
198   growth on maltose and glucose, which agrees with previous findings that the importance of
199   SusD varies depending on the size of the glucose polymers [33].

200
201   The three glycoside hydrolases (GHs), SusA, SusB, and SusG, were variably important across
202   these substrates. SusA was previously characterized biochemically as a neopullulanase that
203   acts to cleave alpha-1,4 linkages in pullulan, a polymer of maltotriose subunits, and was shown
204   to be mildly important for growth on starch [34]. In our data, SusA is only important for growth on
205   α-cyclodextrin, a cyclic oligosaccharide of 6 glucose monomers linked by α(1-4) glycosidic
206   bonds (Figure 2). SusA homologs in other species are often annotated as cyclomaltodextrinase
207   (an enzyme that linearizes circular oligosaccharides), and homologs of this protein are required
208   for growth on α-cyclodextrin in diverse bacteria including *Caulobacter crescentus* and
209   *Echinocola vietnamensis* KMM 6221 [7]. These data suggest that SusA also has

210   cyclomaltodextrinase activity, as has previously been proposed [35]. SusB is a GH97 family
211   member that hydrolyzes multiple linkages, including α(1-4) bonds [36]. In contrast to a previous
212   study that found only a minimal growth defect of a *susB* mutant on starch [34], our results showed
213   that SusB was required for optimal growth on all of these saccharides except glucose,
214   suggesting a key role in hydrolyzing diverse starch derivatives, including the simple maltose
215   disaccharide. SusG is an alpha-amylase localized to the cell surface that breaks down starch
216   into simpler oligosaccharides [37], which is supported by our fitness data: SusG was very
217   important for growth on starch and amylopectin, and to a lesser extent on amylose and α-
218   cyclodextrin (Figure 2). Lastly, while the outer membrane proteins SusE and SusF have been
219   demonstrated to bind starch [38], their precise function remains unclear. We found that these
220   proteins were only important for growth on amylopectin, suggesting that SusEF preferentially
221   bind this component of starch or are important for stabilization of the SusG-starch complex.
222   Taken together, our re-evaluation of the starch utilization system based on genetics data across
223   multiple substrates provides new insight into this well-studied system.
224
225   To globally associate PULs to different carbon substrates, we identified instances where at least
226   one gene in a given PUL had a specific and important phenotype (gene met criteria for specific
227   phenotype and fitness < -1) on a small number of the 45 carbon sources. Based on these
228   criteria, we were able to associate 20 of the 93 PULs (from [39]) to 32 substrates (Table 1, Figure
229   3). Some PULs were only important for growth on a single compound, for example PUL85 and
230   heparin, while others were important on multiple substrates (such as the aforementioned Sus
231   locus). Conversely, some compounds could be linked to a single PUL, for example dextran with
232   PUL48, while others like arabinogalactan appear to be catabolized by multiple PULs (Figure 3).
233   Overall, our experimental data is in strong agreement with previous findings for several PULs.
234   We identified specific and important phenotypes for at least 3 genes from multiple PULs on their
235   predicted or known polysaccharides (Table 1, Figure 3): PUL5 on arabinogalactan [40]; PUL22 on
236   fructans [41]; PUL48 on dextran [42]; PUL57 on chondroitin sulfate and hyaluronic acid [43]; PUL68 on
237   mannan [27]; PUL75 on pectic glycans [44]; PUL85 on heparin [45]; and PUL86 on type I
238   rhamnogalacturonan [44].
239
240   In addition to confirming the results of these previous studies, we provide the first genetic
241   evidence for the prediction that, besides PUL5, PUL47 and PUL80 are also important for
242   arabinogalactan utilization [46]. The arabinogalactan phenotypes for genes in PUL47 and PUL80
243   were mild (fitness values ranged from -0.7 to -1.4 for statistically significant phenotypes), but
244   these mild phenotypes were observed across at least 3 genes in each PUL, and the genes in
245   these clusters do not have clear phenotypes in any of the other conditions that we profiled
246   (Figure 4A,B). These data demonstrate that the efficient breakdown of certain complex
247   polysaccharides like arabinogalactan requires multiple PULs. It was also proposed that PUL65
248   (BT3674-3687) is involved in arabinogalactan breakdown [40], but no genes in this cluster had a
249   strong phenotype in our dataset, including on arabinogalactan (all genes fitness > -0.5).
250
251   Finally, our analysis was able to link some previously uncharacterized PULs to specific
252   monosaccharides and disaccharides (Figure 4C-E). For example, the SusCD homologs of
253   PUL15 (BT1118-1119) were important for growth on galacturonic acid and 1,4-β-D-galactobiose

254 (Figure 4C). In addition, we found that components of PUL37 were only important for growth on
255 D-ribose (Figure 4D), in particular a hypothetical protein (BT2802), two predicted ribokinases
256 (BT2803-2804), and a sugar transporter (BT2809). The SusCD homologs in this cluster were
257 also mildly important for growth on D-ribose (fitness of BT2805 = -1.3 and fitness of BT2806 = -
258 1.4). Lastly, we found that a putative glucosidase (BT3567), SusD-like protein (BT3568), and
259 SusC-like protein (BT3569) from PUL61 were required for optimal growth on the disaccharide
260 laminaribiose (Figure 4E). While the PULs of *B. thetaiotaomicron* have typically been studied in
261 the context of complex polysaccharides, our data suggests that some of these systems may
262 have evolved to specialize in the catabolism of simpler substrates. In support of this view,
263 PUL37 and PUL61 contain the core SusCD proteins, but do not have homologs of SusEF or
264 hydrolases expected to act on complex substrates. Also, in each instance, the SusCD proteins
265 are important for using simple sugars, suggesting that they bind these substrates on the outer
266 surface of the cell.

267

268

269 # Identification of new polysaccharide degradation genes

270 Although the degradation pathways for a number of polysaccharides are well described in *B.*
271 *thetaiotaomicron*, it is likely that additional genes outside of the predicted PULs are also
272 involved in these processes. Indeed, prior work demonstrated that genes important in glycan
273 utilization can be located outside the corresponding PUL(s) [43,47]. By examining genes with
274 specific and important (fitness < -1) phenotypes, we identified three new genes important for the
275 utilization of these polysaccharides.

276

277 The first example is in the metabolism of type I rhamnogalacturonan (RG-I), a polysaccharide
278 containing repeats of a disaccharide of galacturonic acid and rhamnose. A recent study
279 proposed the involvement of five GH28-family glycosyl hydrolases (BT4123, BT4146, BT4149,
280 BT4153, and BT4155) in breaking down RG-I [44]. In our fitness data, only BT4149 was important
281 for utilizing RG-I, although this phenotype was mild (fitness = -1.3). Rather, we found that
282 another GH28-family member, BT4187, was more important for growth on RG-I (fitness = -2.2)
283 (Figure 5). In further support of the importance of BT4187 in RG-I degradation, the operon
284 containing BT4187 is predicted to be in the same regulon as PUL75 [13], which is both important
285 for fitness (Figure 3) and induced during growth on RG-I [46]. While it is possible that the other
286 hydrolases are functionally redundant, our results demonstrate that the loss of BT4187 alone
287 results in a strong fitness defect during growth on RG-I, suggesting that BT4187 has a unique
288 hydrolase activity.

289

290 The next two examples are previously unknown genes for utilizing host-derived
291 glycosaminoglycans (GAGs), which are structural components of animal tissues and are
292 composed of repeating disaccharide units of a hexuronic acid linked to an amino sugar. *B.*
293 *thetaiotaomicron* is capable of using multiple GAGs as nutrients, including chondroitin sulfate
294 (CS), hyaluronic acid (HA), and heparin [45,48,49]. CS and HA differ by their repeating disaccharide:
295 CS contains sulfated N-acetylgalactosamine and glucuronic acid, and HA contains N-
296 acetylglucosamine and glucuronic acid. Previous studies have proposed a model for CS and HA
297 catabolism by *B. thetaiotaomicron* involving PUL57 and BT4410, a CS lyase [17,43]. In our

298 experiments, many components of PUL57 were indeed required for optimal growth on CS and
299 HA, with the SusCD homologs (BT3331-3332), glucuronyl hydrolase (BT3348), and hybrid two-
300 component regulator (BT3334) being particularly important on both substrates (fitness ≤ -1.8 in
301 all CS and HA experiments) (Figure 6A). Other components of PUL57 were preferentially
302 important on one substrate only. BT3349 (a sulfatase) was only important on CS, as expected
303 given that CS contains sulfated sugars, which are not present in HA. The recently described
304 polysaccharide lyase family PL29 enzyme encoded by BT3328 was only important for fitness on
305 HA, despite the activity of this enzyme against both HA and CS *in vitro* [49].
306
307 The first new GAG utilization gene we identified was BT4411, which was important for growth
308 on HA but not on CS (Figure 6A). BT4411 is predicted to be in the same operon as the CS lyase
309 BT4410, which was also more important for fitness on HA. BT4411 contains a domain of
310 unknown function, DUF4627, which to our knowledge has not been previously characterized.
311 DUF4627 has significant similarity to CBM_4_9 (carbohydrate binding domain PF02018) [50], the
312 BT4410-4411 operon is conserved in several other *Bacteroides* species, and both proteins have
313 putative signal peptides [51]. Based on these data, we speculate that BT4410 and BT4411 form a
314 complex and that BT4411 contributes to substrate binding.
315
316 Finally, we found that BT3233 was important for growth on 3 GAGs: CS (fitness = -1.8 and -1.4
317 in two experiments), HA (fitness = -0.7 in two experiments), and heparin (fitness = -2.3 and -2.0
318 in two experiments) (Figure 6A). BT3233 is a member of DUF4861, an uncharacterized protein
319 family with predicted carbohydrate binding activity [50]. Homologs of BT3233 are mostly confined
320 to *Bacteroides* species, including a number of human commensal species. Using an individual
321 transposon mutant from our archived collection, we confirmed that BT3233 is important for
322 growth on CS (Figure 6B,C). Although its precise function is unknown, our data suggests that
323 BT3233 plays an important role in the catabolism of diverse GAGs. Overall, our investigation of
324 novel polysaccharide utilization genes identified three new genes with roles in GAG catabolism,
325 including two genes with domains of unknown function.
326
327

Disaccharide catabolism through a putative 3-keto intermediate

329 We observed that components of a gene cluster (BT2157, BT2159, and BT2160) were
330 important for catabolizing the disaccharides trehalose (glucose-α-1,1-glucose), leucrose
331 (glucose-α-1,5-fructose), palatinose (glucose-α-1,6-fructose), and the trisaccharide raffinose
332 (galactose-α-1,6-glucose-β-1,2-fructose) (Figure 7A). BT2160 encodes a SusR-family
333 transcription factor that likely regulates the cluster. In support of this view, a previous study used
334 comparative genomics to identify a binding motif for BT2160, and predicted that this
335 transcription factor only regulates target genes within its own predicted operon (BT2156-2160)
336 [13], although a biological role for this system was not identified. BT2156 did not have a significant
337 phenotype under any of our assayed conditions. This lack of phenotypes was potentially due to
338 genetic redundancy, as *B. thetaiotaomicron* contains a second gene cluster (BT4446-4449) that
339 is highly similar to BT2156-2159 (Supplementary Figure 1), except that BT4446-4449 does not
340 contain a SusR-family regulator nor does it have a significant phenotype in our dataset. BT2158

341   is very similar to BT4448 (the nucleotide sequences are 98% identical), and we were not able to

342   unambiguously map transposon insertions within these genes using our standard mapping

343   parameters. To determine if BT2158 was important for fitness, we re-ran our transposon

344   insertion mapping with less stringency by only requiring one nucleotide difference to assign

345   insertions to BT2158 and BT4448, which successfully generated fitness data for both of these

346   genes. While mutants of BT4448 did not have strong phenotypes, BT2158 was important for

347   growth on trehalose, leucrose, palatinose, and raffinose (fitness ≤ -1.7 in all experiments on

348   these substrates). Furthermore, BT2158 had the highest cofitness with BT2160 ($r = 0.62$).

349   These results demonstrate that BT2157-2160 are all important for fitness on similar substrates,

350   and thus are very likely to act in the same pathway.

351

352   To further explore this system, we examined the growth of an individual BT2160 transposon

353   mutant on different substrates. The mutant completely lost the ability to use trehalose as a

354   carbon source, and had a substantially longer lag phase than the wild-type strain on raffinose

355   (Figure 7B-D). One explanation for the mutant's growth on raffinose is that the fructose

356   monomer can be removed and consumed by the fructans PUL22, independent of the activity of

357   BT2156-2160. Indeed, components of the fructans PUL are mildly important for consuming all 3

358   fructose-containing sugars (leucrose, palatinose, and raffinose; Figure 3 shows data for

359   palatinose and raffinose), demonstrating that two different systems in *B. thetaiotaomicron* are

360   required for optimal growth on certain small fructose-containing sugars. Next, we examined the

361   induction of this operon on different substrates. A targeted proteomic analysis of stationary

362   phase cultures revealed a significant increase in abundance for three of the target proteins

363   (BT2157-2159) when grown on trehalose (Figure 7E). However, despite the importance of these

364   genes for growth on raffinose, we did not observe increased abundance of these proteins in

365   stationary phase of cultures grown on raffinose (Figure 7E).

366

367   Sequence analysis of the *B. thetaiotaomicron* BT2156-2159 proteins revealed similarities to

368   some proteins putatively involved in the catabolism of disaccharides through a 3-ketoglycoside

369   intermediate. Although we currently lack a complete understanding of the 3-ketoglycoside

370   catabolic pathway, biochemical evidence for the oxidation of disaccharides by *Agrobacterium*

371   *tumefaciens* has long been known [52]. Furthermore, a cloned genomic fragment of *A.*

372   *tumefaciens* conferred sucrose-degrading capabilities to *E. coli*, and a number of biochemical

373   activities were identified and used to propose a catabolic pathway: sucrose is first oxidized to 3-

374   ketodisaccharide and hydrolyzed in the periplasm, and then the 3-ketoglucose is transported

375   into the cytoplasm and reduced (Supplementary Figure 2) [53]. Although this genomic fragment

376   was not sequenced, it did reveal that these biochemical activities are encoded within a single

377   gene cluster in some bacteria. Similar evidence for disaccharide catabolic pathways via a 3-

378   ketoglycoside intermediate have been observed for trehalose utilization in *Sinorhizobium meliloti*

379   [54] and lactose utilization in *Caulobacter crescentus* [55], with the latter example expected to

380   involve a 3-ketogalactose intermediate. Only the first step of this proposed pathway has been

381   definitively linked to genes. This disaccharide 3-dehydrogenase is encoded by two unrelated

382   gene clusters, *thuAB* from *S. meliloti* and *lacABC* from *C. crescentus*. The 3-ketoglycoside

383   hydrolase and the 3-ketosugar reductase activities have not, to our knowledge, been associated

384   to genes in any bacterium.

385

386   Here, we propose that *B. thetaiotaomicron* catabolizes disaccharides through a 3-ketoglycoside

387   intermediate, and that BT2156-2160 are involved in this process. Our evidence relies on two

388   observations. First, by mining fitness data from a previous large survey [7], we identified shared

389   genes that are chromosomally clustered with *lacABC* and are important for catabolizing diverse

390   glycosides in *C. crescentus*, *Pedobacter* sp. GW460-11-11-14-LB5, and *Echinicola*

391   *vietnamensis* KMM 6221. In addition to the known LacABC disaccharide dehydrogenase, we

392   found that one or more IolE-like sugar epimerases, MviM-like dehydrogenases, MFS

393   transporters, and DUF1080-containing proteins were important for fitness on different

394   disaccharides, raffinose, and/or salicin (a glucoside of glucose and salicyl alcohol) in these three

395   species (Supplementary Figure 3). In addition, in each of these three species, at least one copy

396   of all of these proteins is encoded within a single gene cluster, although in some instances a

397   paralogous protein outside the cluster has a stronger or variable phenotype (Supplementary

398   Figure 1).

399

400   Our second line of evidence is that proteins in BT2156-2160 are similar to proteins from the *C.*

401   *crescentus*, *Pedobacter* sp. GW460-11-11-14-LB5, and *E. vietnamensis lacABC* gene clusters,

402   even though *B. thetaiotaomicron* does not contain the full gene cluster (Supplementary Figure

403   1). BT2156 encodes a predicted IolE-like epimerase, BT2157 a DUF1080-containing protein,

404   and BT2158 and BT2159 encode MviM-like dehydrogenases. BT2158 is likely localized to the

405   periplasm [51], while BT2159 is expected to be in the cytoplasm. According to Pfam [50], DUF1080

406   is a hydrolase with structural similarity to endo-1,3-1,4-β-glucanase. Given the importance of

407   various members of this protein family for consuming different glucosides in diverse bacteria,

408   and the genomic proximity of this protein family with other components of the 3-ketoglycoside

409   pathway, we propose that DUF1080-containing proteins are more specifically 3-ketoglycoside

410   hydrolases. In addition, DUF1080-containing proteins are likely localized to the periplasm, as

411   they often contain a putative signal peptide [51]. While the exact details remain to be elucidated,

412   we propose that in *C. crescentus*, *Pedobacter* sp. GW460-11-11-14-LB5, and *E. vietnamensis*,

413   various glucosides are oxidized by LacABC to a 3-ketoglucoside in the periplasm, hydrolyzed to

414   a 3-ketosugar and a second monomer by a DUF1080 hydrolase in the periplasm, and the 3-

415   ketosugar is taken up via the MFS transporter and further metabolized by the IolE-like

416   epimerase and MviM-like dehydrogenase. In *B. thetaiotaomicron*, we have no genetic evidence

417   for the 3-glucoside dehydrogenase nor the MFS transporter, so it remains unclear how the 3-

418   ketoglucoside is formed and where this metabolism occurs.

419

420

421   An efflux system for bile salts

422   During growth in the human gut, *B. thetaiotaomicron* is exposed to complex polysaccharides

423   and also to bile salts, which can inhibit bacterial growth. Bile acids are acidic steroids that are

424   synthesized by the liver from cholesterol. Primary bile acids are conjugated with taurine or

425   glycine before secretion, and the conjugation step greatly increases their water solubility under

426   physiological pH, thus the conjugated form of bile acids are usually called bile salts. Once they

427   reach the colon, bile salts can be deconjugated and oxidized by bacteria into so-called

428   secondary bile acids. Bile salts play an important role as an innate immune defense through

429   their potent antibacterial activity, primarily by disrupting bacterial membrane integrity [56,57].

430   Consequently, commensal bacteria have evolved a number of strategies to cope with bile salts

431   including removal by efflux pumps, deconjugation by bile salt hydrolases (BSH) and oxidation

432   and epimerization of hydroxyl groups by hydroxysteroid dehydrogenases (HSDH) [57]. BSH and

433   HSDH homologs have been identified in *B. thetaiotaomicron* [58], and BT2086 has recently been

434   demonstrated to have BSH activity *in vitro* and *in vivo* [59].

435

436   To identify *B. thetaiotaomicron* genes involved in bile salt tolerance, we examined fitness data

437   from growing the mutant library in the presence of inhibitory concentrations of either the primary

438   bile acid chenodeoxycholic acid, the secondary bile acids lithocholic acid or deoxycholic acid, or

439   cholic acid-derived bile salts (a mixture of sodium cholate and sodium deoxycholate, which we

440   refer to here as "bile salts"). Furthermore, we performed these experiments in two different

441   growth media: BHIS and VB. From these data, we identified a predicted four-gene operon

442   BT2792-2795 that was important for fitness under deoxycholic acid, chenodeoxycholic acid, and

443   bile salt stress (Figure 8A). These genes were important for fitness in both growth media and at

444   multiple stressor concentrations; the phenotypes were more pronounced at higher

445   concentrations. For example, at 0.5 mg/mL bile salts, the fitness values for all four genes

446   ranged from -2.1 to -3.3 (irrespective of growth media). These genes were not important when

447   stressed with lithocholic acid, at least under the concentrations we assayed. Based on

448   homology [60], these proteins form a tripartite multidrug efflux system, with BT2793 as the inner

449   membrane component, BT2794 as the membrane fusion component, and BT2795 as the outer

450   membrane component. BT2792 is a transcriptional regulator that likely regulates the operon.

451   Interestingly, these genes did not have a significant phenotype in any of the other conditions

452   that we profiled (Figure 8A), including dozens of diverse antibiotics, suggesting that this efflux

453   system is specific to bile acids and bile salts. Closely related efflux systems are present in other

454   *Bacteroides* species [61], although the importance of those genes for bile salt tolerance remains to

455   be determined.

456

457   To confirm the contribution of this operon to bile salt tolerance in *B. thetaiotaomicron*, we

458   assayed the growth of individual transposon mutants from our arrayed collection. These data

459   revealed that mutants of BT2793, BT2794, and BT2795 all had longer lag phases and reduced

460   growth rates relative to control strains when stressed with 0.5 mg/mL bile salts (Figure 8B,C).

461   To determine if this efflux system is induced by the stressor, we performed targeted proteomic

462   analysis using wild-type cells grown in the presence of various concentrations of bile salts.

463   BT2792, BT2794, and BT2795 were all induced in the presence of bile salts, and the amount of

464   induction increased with higher bile salt concentrations (Figure 8D). Bile salt-mediated

465   upregulation of efflux pumps has also been observed in other bacteria from the human

466   microbiota [62-65]. Taken together, our data shows that the efflux system encoded by BT2792-

467   2795 is regulated by and important for resistance to bile salts, suggesting that this efflux system

468   in *B. thetaiotaomicron* has evolved to specifically mitigate bile salt stress in the human intestinal

469   tract.

470

471

Transport of antibiotics and biocides

473 Given the extensive use of antimicrobial compounds to treat infection, it is expected that
474 commensal members of the human microbiota are exposed to these inhibitors. Indeed,
475 antibiotic treatment has been shown in some instances to decrease the relative abundance of
476 *Bacteroides* species [66]. We examined the fitness data generated from assaying the mutant
477 library during growth in the presence of different classes of inhibitory compounds including
478 antibiotics and biocides. Given the prevalence of compound efflux as a mechanism of antibiotic
479 resistance in bacteria [67], we focused our analysis on transport-related proteins with strong
480 phenotypes on different inhibitors. For example, BexA (BT0588) belongs to the multidrug and
481 toxic compound extrusion (MATE) family and has previously been shown to be responsible for
482 the efflux of the fluoroquinolone antibiotics norfloxacin and ciprofloxacin in *B. thetaiotaomicron*
483 [68]. In our dataset, we found a mild fitness defect for BT0588 mutants in the presence of
484 ciprofloxacin (minimum fitness on ciprofloxacin = -1.6, although it was not statistically
485 significant). The loss of BT0588 resulted in a more pronounced phenotype with another
486 fluoroquinolone, lomefloxacin, or with paraquat (minimum fitness in a lomefloxacin experiment =
487 -3.1; minimum fitness in a paraquat experiment = -3.6) (Figure 9A). These results illustrate the
488 utility of fitness profiling across many inhibitory compounds to associate new antibiotics to a
489 previously studied efflux protein.
490
491 We also identified phenotypes for previously unstudied transport systems. For example, we
492 found that the RND transporter system encoded by BT3337-3339 was important for fusidic acid
493 tolerance (Figure 9B), and to a lesser extent for tolerating other inhibitors, including cefoxitin
494 and the antipsychotic drug chlorpromazine. These results suggest that the substrate range of
495 BT3337-3339 is broad and includes antimicrobials of diverse chemical classes.
496
497 Lastly, we identified a predicted two-gene operon BT1439-1440 with high positive fitness values
498 when grown in the presence of vancomycin (Figure 9C), a glycopeptide antibiotic that inhibits
499 the biosynthesis of peptidoglycan in the periplasmic space. Vancomycin is commonly used for
500 treating *Clostridium difficile* infections (CDI), and often taken orally. One study showed that fecal
501 levels of vancomycin in CDI patients can range from 0.05 mg/mL to 2 mg/mL during the
502 antibiotic treatment [69], which is above the inhibitory concentration we used in our fitness assays
503 (0.009-0.0186 mg/mL). These numbers suggest that vancomycin use in humans can impact *B.*
504 *thetaiotaomicron* in the gut. BT1439 is a SusD homolog and BT1440 is a SusC homolog, and
505 these genes have high cofitness ($r = 0.76$), indicating that these proteins likely act together.
506 Although SusCD complexes are best known to be involved in translocating carbohydrate
507 substrates from outside the cell to the periplasmic space, no growth deficiency was found in any
508 of the carbohydrate conditions that we assayed. To validate our genome-wide data, we assayed
509 the growth of a BT1439 deletion mutant and found that it was substantially more resistant to
510 vancomycin relative to the parental control (Figure 9DE). The concentration of vancomycin that
511 inhibited growth by half (IC_{50}) was over seven times higher in the BT1439 mutant compared to
512 the non-mutated control (0.0047 versus 0.00066 mg/mL) (Supplementary Figure 4).
513 Interestingly, the vancomycin phenotype of BT1439-1440 was only observed in BHIS media,
514 and not in VB media (Figure 9C). We propose that this media-specific difference in phenotype

515 could be due to altered regulation or the impact of vancomycin binding to another compound in
516 the growth media. Regardless, our data reveal surprising links between genes annotated as
517 carbohydrate transporters and antibiotic sensitivity.
518
519

520 Discussion

521 A small fraction of the gene content of the human microbiome has been experimentally
522 characterized, and as such, there are large knowledge gaps in our molecular understanding of
523 these bacteria. Here, we address this knowledge gap using high-throughput genetics in *B.*
524 *thetaiotaomicron*, the most commonly studied gut commensal bacterium. We performed
525 hundreds of chemical-genetic experiments with a randomly barcoded transposon mutant library
526 and generated nearly two million gene-phenotype measurements. Using these data, we
527 identified novel phenotypes for PULs, linked proteins with domains of unknown function to
528 polysaccharide and disaccharide degradation, and identified an efflux system specific for bile
529 salts tolerance. This large dataset contains phenotypes for hundreds of additional *B.*
530 *thetaiotaomicron* genes and should be an invaluable resource for gene function inference and
531 hypothesis generation for the scientific community. In particular, the 612 genes with either a
532 specific phenotype or high cofitness to another gene are attractive candidates for follow-up
533 investigations. To facilitate these future studies, the *B. thetaiotaomicron* fitness data is available
534 for comparative analyses at the Fitness Brower (http://fit.genomics.lbl.gov), a web portal with
535 data from over 5,000 genome-wide fitness experiments from 35 bacteria.
536
537 Much of the research to date on *B. thetaiotaomicron* has focused on its ability to degrade
538 complex polysaccharides using PULs. While we were able to link 20 PULs to carbon sources
539 via specific phenotypes, this number represents less than 25% of the total number of predicted
540 PULs in the genome. There are a number of potential reasons for our inability to detect a carbon
541 source phenotype for many PULs. First, *B. thetaiotaomicron* may contain two or more PULs that
542 act on the same substrate, and this genetic redundancy will mask the impact of any single-gene
543 loss-of-function mutations. To address this redundancy, future efforts could investigate genetic
544 interactions via double mutants, or by interrogating carbon source utilization in other
545 *Bacteroides* strains under the expectation that not all of these strains will have redundant
546 systems. Indeed, it is known that different gut *Bacteroides* species have different
547 polysaccharide-degrading capabilities [46]. Second, the PUL (or individual genes within the PUL)
548 may perform its activity extracellularly, such that other mutants in the pooled library can
549 complement any putative growth deficiency *in trans*. In addition to assaying mutants individually
550 (including those in our archived mutant collection), newer approaches for performing genome-
551 wide fitness assays within droplets are becoming available, which hold the potential for
552 eliminating complementation by other mutant strains within the pooled library [70]. Third, some of
553 the PULs may not be expressed in our simple growth conditions or are involved in the
554 breakdown of polysaccharides that we did not profile. Fourth, some of these PULs may not have
555 evolved to eat complex polysaccharides, but rather simpler sugars and oligosaccharides. For
556 example, we identified a PUL (BT3567-3569) that was only important for fitness on the
557 disaccharide laminaribiose, suggesting that chemical-genetic profiling with additional simple

558   carbon sources may reveal new functions for some of these PULs. Lastly, the definition of a
559   PUL is very minimal, and only requires a SusCD gene pair. Some of these minimal "PULs" may
560   have other cellular roles. For instance, we showed that a two-gene PUL (BT1439-1440) confers
561   susceptibility to vancomycin in BHIS, while we were unable to detect a carbon source
562   phenotype for this system.
563
564   Our genetic approach can be applied generally to discover new gene functions in the human
565   microbiome. In particular, the barcoding of mutant libraries enables the rapid assessment of
566   gene importance across many conditions at low cost [7]. Nevertheless, challenges still need to be
567   overcome, including the accelerated development of genetic tools for non-model species of the
568   human microbiota [71,72], and the development of scalable phenotypic assays that more
569   accurately reflect the natural ecology of these bacteria, including interactions with other
570   bacteria, phage, and the human host. In addition, the conservation of a phenotype in multiple
571   bacteria is a powerful indicator of gene function [7], and the generation of similar gene-phenotype
572   maps in related bacteria will expand the scope of analysis. In particular, extending our approach
573   to additional *Bacteroides* species from the human gut will provide higher confidence inferences
574   of gene function for conserved genes, as well as provide insights into the functions of accessory
575   genes that are present in a subset of the genomes. A greatly improved understanding of the
576   genes in the human microbiota will accelerate the development of interventions and
577   therapeutics to improve human health.
578
579

580   Materials and Methods

581   **Strains, plasmids, and compounds.** The strains, plasmids, and oligonucleotides used in this
582   study are listed in Supplementary Tables 5-7. All oligonucleotides and gBlocks were ordered
583   from Integrated DNA Technologies. All PCR reactions were carried out with Q5 hot start DNA
584   polymerase from New England Biolabs (NEB). PCR products were cleaned with the DNA Clean
585   & Concentrator kit (Zymo Research). Gibson assembly reactions were performed with the
586   Gibson Assembly Master Mix from NEB. Golden Gate assembly enzymes Esp3I (BsmBI) and
587   BpiI (BbsI) were purchased from ThermoFisher Scientific. T4 DNA ligase and buffer were
588   purchased from NEB. Plasmid isolation was accomplished using the QIAprep Spin Miniprep Kit
589   (Qiagen). In this study, we used many of the compounds that we previously reported in an
590   investigation of bacterial mutant fitness [7]. We also purchased and assayed additional
591   compounds that are specifically relevant to *B. thetaiotaomicron*, such as polysaccharides and
592   bile salts. Given that some of the polysaccharides contain heterogeneous compounds, we
593   provide detailed vendor information in Supplementary Table 4.
594
595   **Constructing a barcoded transposon delivery vector for *B. thetaiotaomicron*.** We
596   previously described a *mariner* transposon delivery vector library pHLL255 (magic pool)
597   containing hundreds of combinations of promoters and erythromycin drug markers [26]. The
598   plasmids in the magic pool contain DNA barcodes that enable the parallel tracking of
599   mutagenesis efficiency for each plasmid design in a single assay. To identify a suitable vector
600   for mutagenizing *B. thetaiotaomicron*, we first constructed a small mutant library by conjugating

601   with an *E. coli* strain carrying pHLL255. We performed TnSeq on this library to map transposon
602   insertion locations and their associated DNA barcodes, enabling us to infer which vector design
603   in the magic pool resulted in the mutation [26]. Analysis of this library revealed that the optimal
604   drug marker for mutagenizing *B. thetaiotaomicron* was the erythromycin resistance gene *ermG*
605   [20,73], the optimal promoter for expressing the *mariner* transposase was the native *rpoD* (BT1311)
606   promoter [20], and the optimal promoter for expressing *ermG* was the native *dnaE* (BT2230)
607   promoter.
608
609   Although this vector design had the highest insertion efficiency relative to the other designs in
610   the magic pool, it still resulted in a significant strand bias whereby the insertion orientation of the
611   transposon was more often in the same transcriptional orientation as the mutated gene. Based
612   on our previous work, we proposed that stronger expression of the selection marker on the
613   transposon might eliminate this strand bias [26]. So, in addition to constructing the best but
614   potentially suboptimal vector design from the magic pool results (pTGG45), we also constructed
615   another vector (pTGG46) by replacing the *dnaE* promoter in pTGG45 with a *cepA* hybrid
616   promoter from *Bacteroides fragilis* RBF-103. This hybrid promoter contains a putative insertion
617   sequence element (IS1224), and has been shown to have increased promoter activity relative to
618   the native *cepA* promoter [74]. To determine which vector was the best combination of high
619   efficiency and low strand bias, we constructed two small mutant libraries with either pTGG45 or
620   pTGG46. TnSeq analysis of these libraries revealed that both had high insertion efficiency, but
621   while the strand bias was again significant for pTGG45 (60% of insertions were in the same
622   orientation as transcription of the mutated gene), there was essentially no strand bias for
623   pTGG46 (49.4%). We then barcoded pTGG46 (pTGG46_NN1) with millions of random 20-
624   nucleotide barcodes using Golden Gate assembly, as previously described [26].
625
626   **Transposon mutant library construction.** We carried out transposon mutagenesis by
627   conjugating *B. thetaiotaomicron* recipient cells with *E. coli* donor cells carrying the barcoded
628   transposon vector library pTGG46_NN1 at 1:1 ratio. Specifically, we grew 10 mL of wild-type *B.*
629   *thetaiotaomicron* in BHIS overnight at 37 °C. The next morning, we recovered a 2 mL freezer
630   stock of AMD776 (*E. coli* WM3064 with pTGG46_NN1) in 50 mL LB supplemented with 50
631   µg/mL carbenicillin and 300 µM diaminopimelic acid (DAP) at 37 °C. When the OD_{600} of the *E.*
632   *coli* donor strain reached ~1.0, we harvested 15 OD_{600} units of the culture and washed the cells
633   3 times with fresh BHIS supplemented with DAP. Then, 15 OD_{600} units of *B. thetaiotaomicron*
634   wild-type cells were harvested and mixed with the washed donor strain cells, and resuspended
635   in a final volume of 60 µL with BHIS supplemented with DAP. The resuspension was spotted
636   onto 0.45-µm gravimetric analysis membrane filters (Millipore), and incubated aerobically
637   overnight on BHIS agar plates supplemented with DAP at 37 °C. The next day, the conjugation
638   mixture was scraped from the membrane and resuspended into 6 mL fresh BHIS medium
639   supplemented with 10 µg/mL erythromycin, and plated at different dilutions on BHIS plates
640   supplemented with 10 µg/mL erythromycin. The plates were incubated at 37 °C for 48 hours to
641   let visible colonies develop. We then pooled ~250,000 colonies, and grew the pooled library in
642   liquid BHIS supplemented with 10 µg/mL erythromycin and 200 µg/mL gentamicin (to further
643   select against the *E. coli* conjugation donor strain). When the cells reached saturation, we made
644   multiple, single-use glycerol stocks of the final library and extracted genomic DNA for TnSeq

645   analysis. To map the genomic location of the transposon insertions and to link these insertions
646   to their associated DNA barcode, we used the same TnSeq protocol that we described
647   previously [25]. The final mutant library was named Btheta_ML6. Lastly, using single-cell sorting,
648   we generated and mapped an archived collection of individual transposon insertion mutants
649   from this library (A.L.S, A.M.D, and K.C.H, in review).

650

651   **Individual mutant strain construction and assays.** We constructed an unmarked, in-frame
652   deletion mutant of BT1439 using a previously established approach [75]. The individual
653   transposon insertion mutants we assayed were from the arrayed collection. We performed
654   growth curves of individual mutants in a 96 well plate in a Tecan Infinite F200 instrument
655   housed in a Coy anaerobic chamber. To calculate IC_{50} values for the BT1439 deletion mutant
656   and parent strain on vancomycin, we used the drc package in R [76]. Specifically, we used drm()
657   with the LL.4() model and defined the response as the average OD_{600} up until the time that
658   uninhibited cells (no vancomycin) reached their maximum OD_{600} value.

659

660   **Growth conditions.** Growth media was purchased from BD and most compounds were
661   purchased from Sigma-Aldrich. Some carbohydrates were purchased from other vendors
662   (Supplementary Table 4**)**. We typically grew *B. thetaiotaomicron* anaerobically in a Coy chamber
663   at 37 °C with one of two different growth media. First, we used Brain Heart Infusion
664   Supplemented (BHIS) medium using Bacto Brain Heart Infusion (BHI) as the base. Second, we
665   used a defined medium developed by Varel and Bryant (VB) [77]. The recipes for both media are
666   provided in Supplementary Table 8. For culturing individual *B. thetaiotaomicron* transposon
667   insertion mutants, we supplemented the media with 10 μg/ml erythromycin. For growth on solid
668   media, we used 1.5% (w/v) agar.

669

670   **Genome-wide mutant fitness assays.** We performed genome-wide mutant fitness assays as
671   described previously [7]. Briefly, we thawed an aliquot of the full transposon mutant library,
672   inoculated the entire aliquot into 50 mL of BHIS supplemented with 10 μg/ml erythromycin, and
673   grew the library anaerobically at 37°C until the cells reached mid-log phase. We then collected 6
674   cell pellets of ~1.0 OD_{600} unit each (the "Time0" samples; we collect multiple pellets for the
675   Time0 sample because we perform BarSeq sequencing on each of these). We then used the
676   remaining cells to inoculate competitive growth assays. Nearly all fitness assays were
677   performed in 1.2 mL of growth medium in either a 24-well transparent microplate (Greiner) or in
678   a 96-deepwell plate (Costar). For fitness assays in BHIS or VB, we diluted the recovered mutant
679   library to a starting OD_{600} of 0.02 for each experiment. For carbon source assays in VB, we
680   washed the cells in VB lacking a carbon source two times. For nitrogen source experiments, we
681   washed the cells in VB lacking a nitrogen source two times. For all growth-based fitness assays,
682   we grew the cells until the cultures reached stationary phase, and then collected cell pellets (the
683   "Condition" sample). We also performed a few "survival" experiments in which we incubated the
684   mutant library aerobically for different amounts of time. After this incubation, we inoculated cells
685   into fresh BHIS and grew anaerobically to select for cells that survived the stress. We extracted
686   genomic DNA from the Time0 and Condition samples using either the DNeasy Blood and
687   Tissue kit (Qiagen) or in a 96-well microplate format with the QIAamp 96 DNA Qiacube HT kit
688   (Qiagen).

689

690   We performed barcode sequencing (BarSeq) as previously described [7,78]. For most
691   experiments, we performed BarSeq with indexed P2 oligos and a mix of non-indexed P1 oligos
692   with variable lengths of N's (2-5) to stagger the reads. For some experiments, we used BarSeq
693   oligos with both P1 and P2 indexed to minimize the impact of incorrectly assigned indexes in
694   Illumina HiSeq4000 runs [79]. Strain and gene fitness scores were calculated as previously
695   described [25]. Briefly, strain fitness was calculated as the $\log_2$ ratio of barcode abundance after
696   growth selection (Condition) versus before growth selection (Time0). Gene fitness was
697   calculated as the weighted average of individual strain fitness values and was normalized to
698   remove the effect of chromosomal position (due to variation in copy number along the
699   chromosome). Finally, gene fitness was normalized so that the mode of the distribution was 0.
700

701   **Analysis of fitness data.** Overall, we attempted 772 genome-wide fitness assays in this study,
702   of which 492 passed our metrics for internal consistency [25]. From these 492 experiments, we
703   identified genes with a specific phenotype in only one or a handful of conditions using previously
704   described criteria [7]: |fitness| > 1 and |t| > 5 and 95% of |fitness| < 1 and |fitness| > 95th
705   percentile(|fitness|) + 0.5, where t is fitness/estimated standard error and |t| > 5 ensures that the
706   phenotype is statistically significant. For the analysis of the total number of genes with a specific
707   phenotype, we excluded one gene that had specific phenotypes in control experiments with
708   plain BHIS medium. A specific important phenotype has the same criteria as above except that
709   the fitness score has to be negative. We identified pairs of genes with highly correlated patterns
710   of phenotypes across all 492 successful experiments (cofitness) using the linear (Pearson)
711   correlation and a threshold of 0.8 [7].
712

713   COG assignments and functional codes were downloaded from MicrobesOnline [61].
714   Polysaccharide utilization loci assignments were derived from the CAZy database [39]. Most data
715   analyses to infer the putative functions of genes were performed with the Fitness Browser
716   (http://fit.genomics.lbl.gov).
717

718   **Targeted proteomics experiments.** For carbon source experiments, we grew wild-type *B.*
719   *thetaiotaomicron* in VB with 20 mM D-trehalose, D-raffinose, or D-glucose. After overnight
720   incubation, the OD_{600} values of the cultures were between 1.0 and 1.5, and the cultures were
721   used for proteomics analysis. For bile salt experiments, we grew wild-type *B. thetaiotaomicron*
722   in defined VB medium with 20 mM D-glucose and supplemented with bile salts at 4 different
723   concentrations: 0, 0.125, 0.25, and 0.5 mg/mL. These cultures were growth for 20 hours, after
724   which the OD_{600} values were between 0.5 and 0.9. For all experiments, we harvested 2 OD_{600}
725   units of cells and washed two times with 10 mM NH_4HCO_3 before extracting the total protein
726   using a methanol-chloroform precipitation method. Protein concentration was determined by
727   Bio-Rad DC protein assay (Lot# 5000112) following manufacturer's protocol, and the same
728   amount of proteins were subjected to standard reduction, alkylation, and trypsin digestion
729   procedures. A targeted selected reaction monitoring (SRM) method was developed to quantify
730   unique peptides of proteins with the assistance of an in-house constructed *B. thetaiotaomicron*
731   spectral library [80].
732

733   SRM targeted proteomic assays were performed on an Agilent 6460 QQQ mass spectrometer
734   system coupled with an Agilent 1290 UHPLC system (Agilent Technologies, Santa Clara, CA).
735   Twenty microgram peptides of each investigated sample were separated on an Ascentis
736   Express Peptide C18 column (2.7-mm particle size, 160-Å pore size, 5-cm length × 2.1-mm
737   inside diameter (ID), coupled to a 5-mm × 2.1-mm ID guard column with the same particle and
738   pore size, operating at 60 °C; Sigma-Aldrich) operating at a flow rate of 0.4 mL/min via the
739   following gradient: initial conditions were 98% solvent A (0.1% formic acid), 2% solvent B
740   (99.9% acetonitrile, 0.1% formic acid). Solvent B was increased to 40% over 5 min, and then
741   increased to 80% over 0.5 min, and held for 2 min at a flow rate of 0.6 mL/min, followed by a
742   ramp back down to 2% solvent B over 0.5 min where it was held for 1 min to re-equilibrate the
743   column to original conditions. The eluted peptides were ionized via an Agilent Jet Stream ESI
744   source operating in positive-ion mode with the following source parameters: gas temperature =
745   250 °C, gas flow = 13 L/min, nebulizer pressure = 35 psi, sheath gas temperature = 250 °C,
746   sheath gas flow = 11 L/min, capillary voltage = 3500 V, nozzle voltage = 0 V. Data were
747   acquired using Agilent MassHunter v. B.08.02. Acquired SRM data were analyzed by Skyline v.
748   3.70 (MacCoss Lab Software).
749
750   **Strain, data, and code availability.** The barcoded *B. thetaiotaomicron* transposon mutant
751   library (Btheta_ML6) is available upon request. Fitness data from the 492 successful
752   experiments is available for comparative analysis at the Fitness Browser
753   (http://fit.genomics.lbl.gov). In addition, the fitness data (and all associated metadata) is
754   available at: https://doi.org/10.6084/m9.figshare.7742855.v1. The software for RB-TnSeq is
755   available at https://bitbucket.org/berkeleylab/feba.
756

Acknowledgements

757

References

1. NIH HMP Working Group *et al.* The NIH Human Microbiome Project. *Genome Res* **19,** 2317–2323 (2009).

2. Zitvogel, L., Daillère, R., Roberti, M. P., Routy, B. & Kroemer, G. Anticancer effects of the microbiome and its products. *Nat Rev Microbiol* **15,** 465–478 (2017).

3. Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* **13,** 260–270 (2012).

4. Guinane, C. M. & Cotter, P. D. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Therap Adv Gastroenterol* **6,** 295–308 (2013).

5. Heintz-Buschart, A. & Wilmes, P. Human Gut Microbiome: Function Matters. *Trends Microbiol* **26,** 563–574 (2018).

6. Schnoes, A. M., Brown, S. D., Dodevski, I. & Babbitt, P. C. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput Biol* **5,** e1000605 (2009).

7. Price, M. N. *et al.* Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* **557,** 503–509 (2018).

8. Wexler, A. G. & Goodman, A. L. An insider's perspective: Bacteroides as a window into the microbiome. *Nat Microbiol* **2,** 17026 (2017).

9. Salyers, A. A., Vercellotti, J. R., West, S. E. & Wilkins, T. D. Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. *Applied and Environmental Microbiology* **33,** 319–322 (1977).

10. Salyers, A. A. & O'Brien, M. Cellular location of enzymes involved in chondroitin sulfate breakdown by Bacteroides thetaiotaomicron. *J. Bacteriol.* **143,** 772–780 (1980).

11. Kotarski, S. F. & Salyers, A. A. Isolation and characterization of outer membranes of Bacteroides thetaiotaomicron grown on different carbohydrates. *J. Bacteriol.* **158,** 102–109 (1984).

12. Heinken, A., Sahoo, S., Fleming, R. M. T. & Thiele, I. Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. *Gut Microbes* **4,** 28–40 (2013).

13. Ravcheev, D. A., Godzik, A., Osterman, A. L. & Rodionov, D. A. Polysaccharides utilization in human gut bacterium Bacteroides thetaiotaomicron: comparative genomics reconstruction of metabolic and regulatory networks. *BMC Genomics* **14,** 873 (2013).

14. Xu, J. *et al.* A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* **299,** 2074–2076 (2003).

15. Bjursell, M. K., Martens, E. C. & Gordon, J. I. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period. *J Biol Chem* **281,** 36269–36279 (2006).

16. Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284,** 24673–24677 (2009).

17. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4,**

810   447–457 (2008).

811   18. Gray, A. N. *et al.* High-throughput bacterial functional genomics in the sequencing era.
812   *Current Opinion in Microbiology* **27,** 86–95 (2015).

813   19. van Opijnen, T. & Camilli, A. Transposon insertion sequencing: a new tool for systems-
814   level analysis of microorganisms. *Nat Rev Microbiol* **11,** 435–442 (2013).

815   20. Goodman, A. L. *et al.* Identifying genetic determinants needed to establish a human gut
816   symbiont in its habitat. *Cell Host Microbe* **6,** 279–289 (2009).

817   21. Wu, M. *et al.* Genetic determinants of in vivo fitness and diet responsiveness in multiple
818   human gut Bacteroides. *Science* **350,** aac5992–aac5992 (2015).

819   22. Nichols, R. J. *et al.* Phenotypic landscape of a bacterial cell. *Cell* **144,** 143–156 (2011).

820   23. Hillenmeyer, M. E. *et al.* Systematic analysis of genome-wide fitness data in yeast
821   reveals novel gene function and drug action. *Genome Biol.* **11,** R30 (2010).

822   24. Deutschbauer, A. *et al.* Evidence-based annotation of gene function in Shewanella
823   oneidensis MR-1 using genome-wide fitness profiling across 121 conditions. *PLoS Genet.*
824   **7,** e1002385 (2011).

825   25. Wetmore, K. M. *et al.* Rapid quantification of mutant fitness in diverse bacteria by
826   sequencing randomly bar-coded transposons. *MBio* **6,** e00306–15 (2015).

827   26. Liu, H. *et al.* Magic Pools: Parallel Assessment of Transposon Delivery Vectors in
828   Bacteria. *mSystems* **3,** e00143–17 (2018).

829   27. Cuskin, F. *et al.* Human gut Bacteroidetes can utilize yeast mannan through a selfish
830   mechanism. *Nature* **517,** 165–169 (2015).

831   28. Stentz, R. *et al.* Cephalosporinases associated with outer membrane vesicles released
832   by Bacteroides spp. protect gut pathogens and commensals against β-lactam antibiotics.
833   *J. Antimicrob. Chemother.* **70,** 701–709 (2015).

834   29. Caldara, M. *et al.* Arginine biosynthesis in Escherichia coli: experimental perturbation and
835   mathematical modeling. *J Biol Chem* **283,** 6347–6358 (2008).

836   30. Kublanov, I. V. *et al.* Genomic Analysis of Caldithrix abyssi, the Thermophilic Anaerobic
837   Bacterium of the Novel Bacterial Phylum Calditrichaeota. *Front Microbiol* **8,** 195 (2017).

838   31. Tatusov, R. L. *et al.* The COG database: an updated version includes eukaryotes. *BMC*
839   *Bioinformatics* **4,** 41 (2003).

840   32. Price, M. N. & Arkin, A. P. PaperBLAST: Text Mining Papers for Information about
841   Homologs. *mSystems* **2,** e00039–17 (2017).

842   33. Koropatkin, N. M., Martens, E. C., Gordon, J. I. & Smith, T. J. Starch catabolism by a
843   prominent human gut symbiont is directed by the recognition of amylose helices.
844   *Structure* **16,** 1105–1115 (2008).

845   34. D'Elia, J. N. & Salyers, A. A. Effect of regulatory protein levels on utilization of starch by
846   Bacteroides thetaiotaomicron. *J. Bacteriol.* **178,** 7180–7186 (1996).

847   35. Smith, K. A. & Salyers, A. A. Characterization of a neopullulanase and an alpha-
848   glucosidase from Bacteroides thetaiotaomicron 95-1. *J. Bacteriol.* **173,** 2962–2968
849   (1991).

850   36. Kitamura, M. *et al.* Structural and functional analysis of a glycoside hydrolase family 97
851   enzyme from Bacteroides thetaiotaomicron. *J Biol Chem* **283,** 36328–36337 (2008).

852   37. Koropatkin, N. M. & Smith, T. J. SusG: a unique cell-membrane-associated alpha-
853   amylase from a prominent human gut symbiont targets complex starch molecules.

854   *Structure* **18,** 200–215 (2010).

855 38. Cameron, E. A. *et al.* Multidomain Carbohydrate-binding Proteins Involved in Bacteroides
856   thetaiotaomicron Starch Metabolism. *J Biol Chem* **287,** 34614–34625 (2012).

857 39. Terrapon, N. *et al.* PULDB: the expanded database of Polysaccharide Utilization Loci.
858   *Nucleic Acids Res.* **46,** D677–D683 (2018).

859 40. Cartmell, A. *et al.* A surface endogalactanase in Bacteroides thetaiotaomicron confers
860   keystone status for arabinogalactan degradation. *Nat Microbiol* **3,** 1314–1326 (2018).

861 41. Sonnenburg, E. D. *et al.* Specificity of polysaccharide use in intestinal bacteroides
862   species determines diet-induced microbiota alterations. *Cell* **141,** 1241–1252 (2010).

863 42. Rogers, T. E. *et al.* Dynamic responses of Bacteroides thetaiotaomicron during growth on
864   glycan mixtures. *Mol Microbiol* **88,** 876–890 (2013).

865 43. Raghavan, V., Lowe, E. C., Townsend, G. E., Bolam, D. N. & Groisman, E. A. Tuning
866   transcription of nutrient utilization genes to catabolic rate promotes growth in a gut
867   bacterium. *Mol Microbiol* **93,** 1010–1025 (2014).

868 44. Luis, A. S. *et al.* Dietary pectic glycans are degraded by coordinated enzyme pathways in
869   human colonic Bacteroides. *Nat Microbiol* **3,** 210–219 (2018).

870 45. Cartmell, A. *et al.* How members of the human gut microbiota overcome the sulfation
871   problem posed by glycosaminoglycans. *Proc. Natl. Acad. Sci. U.S.A.* **114,** 7037–7042
872   (2017).

873 46. Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by
874   two human gut symbionts. *PLoS Biol* **9,** e1001221 (2011).

875 47. Benjdia, A., Martens, E. C., Gordon, J. I. & Berteau, O. Sulfatases and a radical S-
876   adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the
877   prominent human gut symbiont, Bacteroides thetaiotaomicron. *J Biol Chem* **286,** 25973–
878   25982 (2011).

879 48. Raghavan, V. & Groisman, E. A. Species-specific dynamic responses of gut bacteria to a
880   mammalian glycan. *J. Bacteriol.* **197,** 1538–1548 (2015).

881 49. Ndeh, D. *et al.* The human gut microbe Bacteroides thetaiotaomicron encodes the
882   founding member of a novel glycosaminoglycan-degrading polysaccharide lyase family
883   PL29. *J Biol Chem* **293,** 17906–17916 (2018).

884 50. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47,**
885   D427–D432 (2019).

886 51. Yu, N. Y. *et al.* PSORTb 3.0: improved protein subcellular localization prediction with
887   refined localization subcategories and predictive capabilities for all prokaryotes.
888   *Bioinformatics* **26,** 1608–1615 (2010).

889 52. Bernaerts, M. J. & De Ley, J. Microbiological Formation and Preparation of 3-
890   Ketoglycosides from Disaccharides. *Journal of General Microbiology* **22,** 129–136 (1960).

891 53. Schuerman, P. L., Liu, J. S., Mou, H. & Dandekar, A. M. 3-Ketoglycoside-mediated
892   metabolism of sucrose in E. coli as conferred by genes from Agrobacterium tumefaciens.
893   *Appl. Microbiol. Biotechnol.* **47,** 560–565 (1997).

894 54. Ampomah, O. Y. *et al.* The thuEFGKAB Operon of Rhizobia and Agrobacterium
895   tumefaciens Codes for Transport of Trehalose, Maltitol, and Isomers of Sucrose and
896   Their Assimilation through the Formation of Their 3-Keto Derivatives. *J. Bacteriol.* **195,**
897   3797–3807 (2013).

898   55.   Arellano, B. H., Ortiz, J. D., Manzano, J. & Chen, J. C. Identification of a Dehydrogenase
899   Required for Lactose Metabolism in Caulobacter crescentus. *Applied and Environmental*
900   *Microbiology* **76,** 3004–3014 (2010).
901   56.   Ridlon, J. M., Kang, D. J., Hylemon, P. B. & Bajaj, J. S. Bile acids and the gut
902   microbiome. *Curr. Opin. Gastroenterol.* **30,** 332–338 (2014).
903   57.   Begley, M., Gahan, C. G. M. & Hill, C. The interaction between bacteria and bile. *FEMS*
904   *Microbiol. Rev.* **29,** 625–651 (2005).
905   58.   Ridlon, J. M., Kang, D.-J. & Hylemon, P. B. Bile salt biotransformations by human
906   intestinal bacteria. *J. Lipid Res.* **47,** 241–259 (2006).
907   59.   Yao, L. *et al.* A selective gut bacterial bile salt hydrolase alters host metabolism. *Elife* **7,**
908   675 (2018).
909   60.   Overbeek, R. *et al.* The subsystems approach to genome annotation and its use in the
910   project to annotate 1000 genomes. *Nucleic Acids Res.* **33,** 5691–5702 (2005).
911   61.   Dehal, P. S. *et al.* MicrobesOnline: an integrated portal for comparative and functional
912   genomics. *Nucleic Acids Res.* **38,** D396–400 (2010).
913   62.   Prouty, A. M., Brodsky, I. E., Falkow, S. & Gunn, J. S. Bile-salt-mediated induction of
914   antimicrobial and bile resistance in Salmonella typhimurium. *Microbiology (Reading,*
915   *Engl.)* **150,** 775–783 (2004).
916   63.   Pumbwe, L. *et al.* Bile salts enhance bacterial co-aggregation, bacterial-intestinal
917   epithelial cell adhesion, biofilm formation and antimicrobial resistance of Bacteroides
918   fragilis. *Microb. Pathog.* **43,** 78–87 (2007).
919   64.   Lin, J. *et al.* Bile salts modulate expression of the CmeABC multidrug efflux pump in
920   Campylobacter jejuni. *J. Bacteriol.* **187,** 7417–7424 (2005).
921   65.   Ruiz, L., Margolles, A. & Sánchez, B. Bile resistance mechanisms in Lactobacillus and
922   Bifidobacterium. *Front Microbiol* **4,** 396 (2013).
923   66.   Jernberg, C., Löfmark, S., Edlund, C. & Jansson, J. K. Long-term ecological impacts of
924   antibiotic administration on the human intestinal microbiota. *ISME J* **1,** 56–66 (2007).
925   67.   Webber, M. A. & Piddock, L. J. V. The importance of efflux pumps in bacterial antibiotic
926   resistance. *J. Antimicrob. Chemother.* **51,** 9–11 (2003).
927   68.   Miyamae, S. *et al.* A MATE family multidrug efflux transporter pumps out fluoroquinolones
928   in Bacteroides thetaiotaomicron. *Antimicrob. Agents Chemother.* **45,** 3341–3346 (2001).
929   69.   Gonzales, M. *et al.* Faecal pharmacokinetics of orally administered vancomycin in
930   patients with suspected Clostridium difficile infection. *BMC Infect. Dis.* **10,** 363 (2010).
931   70.   Thibault, D., Wood, S., Jensen, P. & van Opijnen, T. droplet-Tn-Seq combines
932   microfluidics with Tn-Seq identifying complex single-cell phenotypes. *bioRxiv* (2018).
933   71.   Peters, J. M. *et al.* Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi.
934   *Nat Microbiol* **4,** 244–250 (2019).
935   72.   Liu, H. & Deutschbauer, A. M. Rapidly moving new bacteria to model-organism status.
936   *Curr. Opin. Biotechnol.* **51,** 116–122 (2018).
937   73.   Shoemaker, N. B., Vlamakis, H., Hayes, K. & Salyers, A. A. Evidence for extensive
938   resistance gene transfer among Bacteroides spp. and among Bacteroides and other
939   genera in the human colon. *Applied and Environmental Microbiology* **67,** 561–568 (2001).
940   74.   Bayley, D. P., Rocha, E. R. & Smith, C. J. Analysis of cepA and other Bacteroides fragilis
941   genes reveals a unique promoter structure. *FEMS Microbiol Lett* **193,** 149–154 (2000).

942   75.   Mishra, S. & Imlay, J. A. An anaerobic bacterium, Bacteroides thetaiotaomicron, uses a
943   consortium of enzymes to scavenge hydrogen peroxide. *Mol Microbiol* **90,** 1356–1371
944   (2013).
945   76.   Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-Response Analysis Using R. *PLoS*
946   *One* **10,** e0146021 (2015).
947   77.   Varel, V. H. & Bryant, M. P. Nutritional features of Bacteroides fragilis subsp. fragilis. *Appl*
948   *Microbiol* **28,** 251–257 (1974).
949   78.   Price, M. N. *et al.* Oxidative Pathways of Deoxyribose and Deoxyribonate Catabolism.
950   *mSystems* **4,** 2977 (2019).
951   79.   Sinha, R. *et al.* Index Switching Causes 'Spreading-Of-Signal' Among Multiplexed
952   Samples In Illumina HiSeq 4000 DNA Sequencing. *bioRxiv* (2017).
953   80.   Chen, Y. *et al.* A rapid methods development workflow for high-throughput quantitative
954   proteomic applications. *PLoS One* **14,** e0211582 (2019).
955   81.   Peterson, J. D., Umayam, L. A., Dickinson, T., Hickey, E. K. & White, O. The
956   Comprehensive Microbial Resource. *Nucleic Acids Res.* **29,** 123–125 (2001).

957

958 # Figures and Tables

959

PUL[a]	Genes[b]	Carbohydrates[c]
5	BT0262-0291	Arabinogalactan (BT0267, 0268, 0272, 0275, 0276, 0278, 0288)
7	BT0348-0356, 0358-0369	D-Arabinose (BT0351, 0355, 0356); L-Arabinose (BT0350, 0351, 0355, 0356)
15	BT1117-1119	1,4-β-D-Galactobiose (BT1118); Galacturonic acid (BT1118, 1119)
22	BT1754, 1757-1763, 1765	Fructose (BT1754, 1757, 1759); Raffinose (BT1754, 1757, 1759); Palatinose (BT1754); Fructooligosaccharides (BT1754, 1757, 1762); Stachyose (BT1754); Levan (BT1754, 1757);
24	BT1871-1878	Melibiose (BT1871)
36	BT2615-2633	Mannan from yeast (BT2629)
37	BT2802-2809	Ribose (BT2802, 2803, 2804, 2809)
47	BT3043-3049	Arabinogalactan (BT3044, 3045)
48	BT3086-3091	Dextran (BT3087, 3088, 3089, 3090, 3091); Red arabinan (BT3089)
57	BT3324, 3328-3334, 3348-335	Chondroitin sulfate A (BT3330, 3331, 3332, 3334, 3348, 3349); Hyaluronic acid (BT3328, 3329, 3330, 3331, 3332, 3334, 3348); Heparin (BT3348)
59	BT3461-3507	Fructooligosaccharides (BT3485); Hyaluronic acid (BT3487)
61	BT3566-3570	Laminaribiose (BT3567, 3568, 3569)
66	BT3698-3705	Maltose (BT3703, 3705); Maltotetraose (BT3701, 3702, 3703, 3705); Maltohexaose (BT3701, 3702, 3703, 3705); α-Cyclodextrin (BT3701, 3702, 3703, 3704, 3705); Amylose (BT3701, 3702, 3703, 3705); Amylopectin (BT3698, 3699, 3700, 3701, 3702, 3703, 3705); Starch (BT3698, 3701, 3702, 3703, 3705) Polygalacturonic acid (BT3698, 3701, 3702, 3705);
68	BT3773-3782, 3784, 3786-3792	Mannan from yeast (BT3784, 3787, 3788, 3789, 3790)
75	BT4108-4124	Pectin (BT4113); Polygalacturonic acid (BT4111, 4113, 4114, 4115); Rhamnogalacturonan I (BT4111, 4113, 4114)
77	BT4145-4183	Rhamnogalacturonan I (BT4169)
80	BT4294-4300	Galacturonic acid (BT4300); Arabinogalactan (BT4295, 4297)
84	BT4631-4636	Arabinogalactan (BT4636)
85	BT4652-4663, 4675	Heparin (BT4652, 4655, 4656, 4657, 4658, 4659, 4660, 4661, 4662, 4663)
86	BT4667-4673	1,4-β-D-Galactobiose (BT4667, 4669, 4670, 4671, 4673); Rhamnogalacturonan I (BT4667, 4670, 4671, 4673); Polygalacturonic acid (BT4670)

960

Table 1. Genes with specific and important phenotypes for 20 PULs on various carbon sources. A PUL was included in the table if at least one of its component genes was specifically important for fitness (fitness < -1) on one or a few of the 45 carbon sources we profiled.

[a]Literature-derived PULs from CAZy database (http://www.cazy.org/PULDB/) [39];[b]Proteins in the PUL; [c]Carbohydrate conditions in which gene(s) from the PUL have a specific and important phenotype.
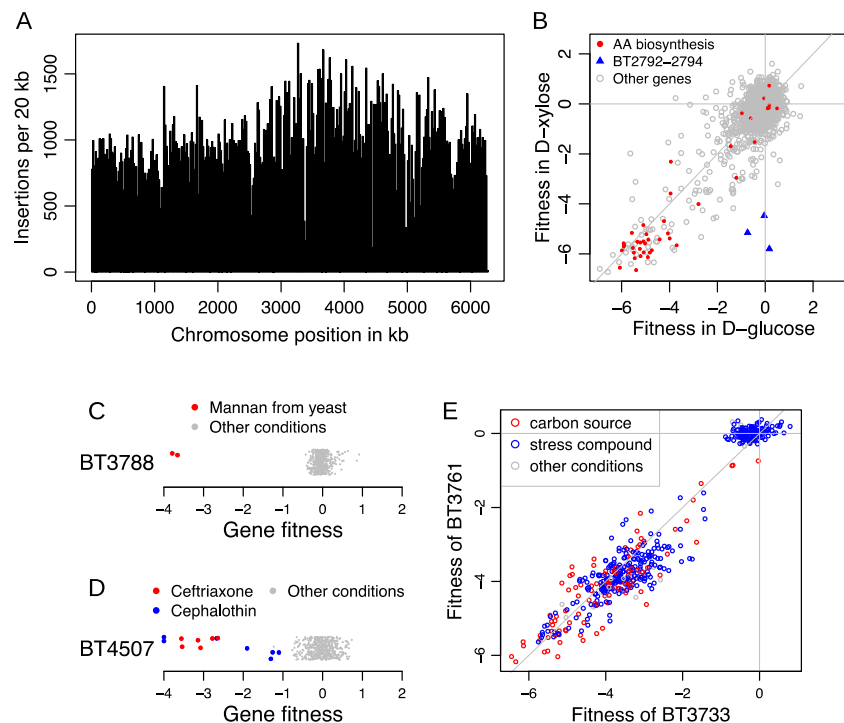
967

968
969

970
971

Figure 1. High-throughput genetics in *B. thetaiotaomicron*. (A) Transposon insertions were approximately evenly distributed across the *B. thetaiotaomicron* chromosome. (B) Gene fitness values during growth with glucose or xylose as the sole carbon source were consistent with previous knowledge. The amino acid biosynthesis genes are predicted by TIGR roles [81]. (C,D) Gene fitness values across conditions revealed that BT3788 is important for fitness on mannan, and mutants in BT4507 were impaired only under treatment with ceftriaxone and cephalothin. The *y*-axis is random and fitness values less than −4 are shown at −4. Negative values mean that the mutant is less fit than the typical mutant. The conditions in which these genes have a specific phenotype are marked. (E) Comparison of gene fitness values for BT3733 and BT3761 across all experiments. Because BT3733 encodes an enzyme in arginine biosynthesis, the high cofitness suggests that BT3761 is also involved in arginine biosynthesis.
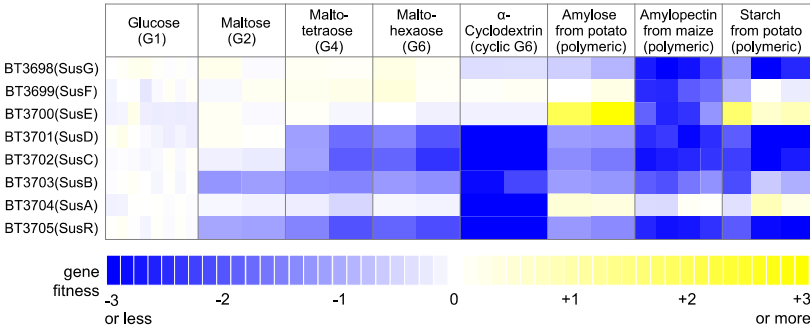
Figure 2. Fitness data for the Sus PUL shows a broad range of importance for substrate catabolism. A heatmap of gene fitness values for the starch utilization system genes during growth on different glucose-containing carbon sources is shown. The data from each replicate experiment is shown separately, for example 6 glucose experiments and 4 amylopectin experiments. The compounds are ordered from left to right by increasing complexity.
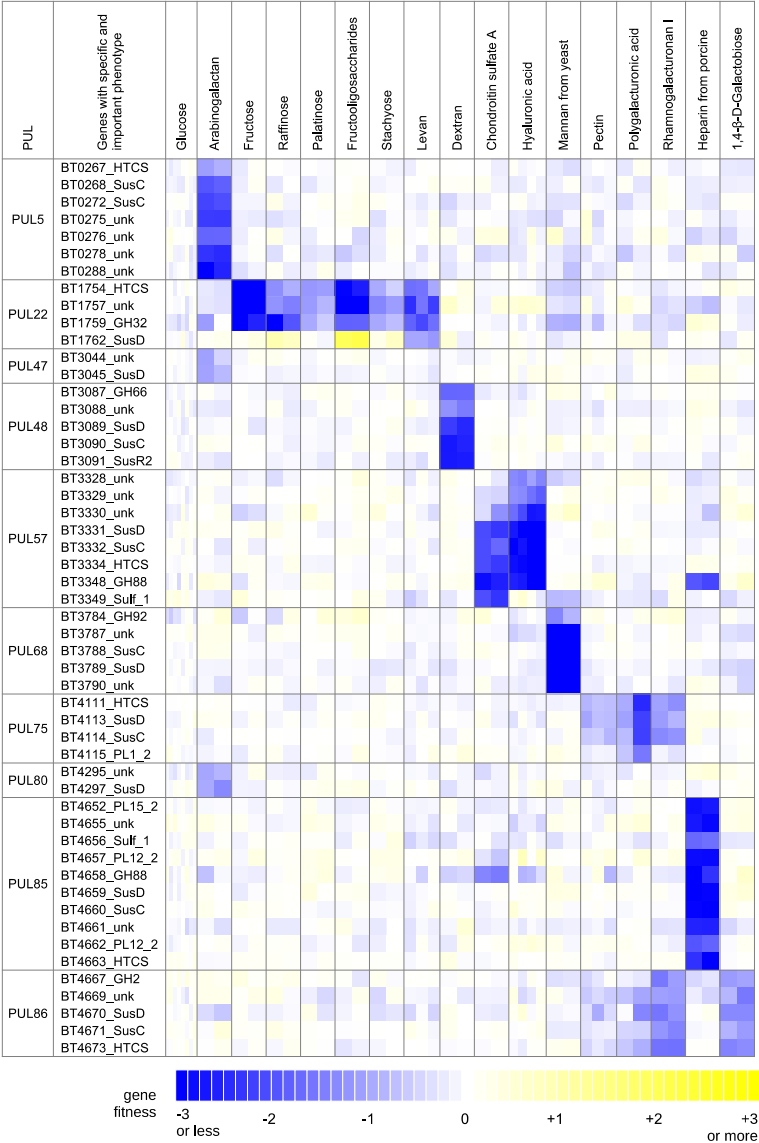
Figure 3. Fitness data associates PULs with substrates. A heatmap of fitness data for select genes with important (fitness < -1) and specific phenotypes during growth on various carbon sources is shown. We present the fitness data for each replicate experiment separately. HTCS: AraC-like hybrid two-component system; unk: unknown; GH: glycoside hydrolase; PL: polysaccharide lyase. The labels indicate gene families and annotations are derived from PULDB [39].
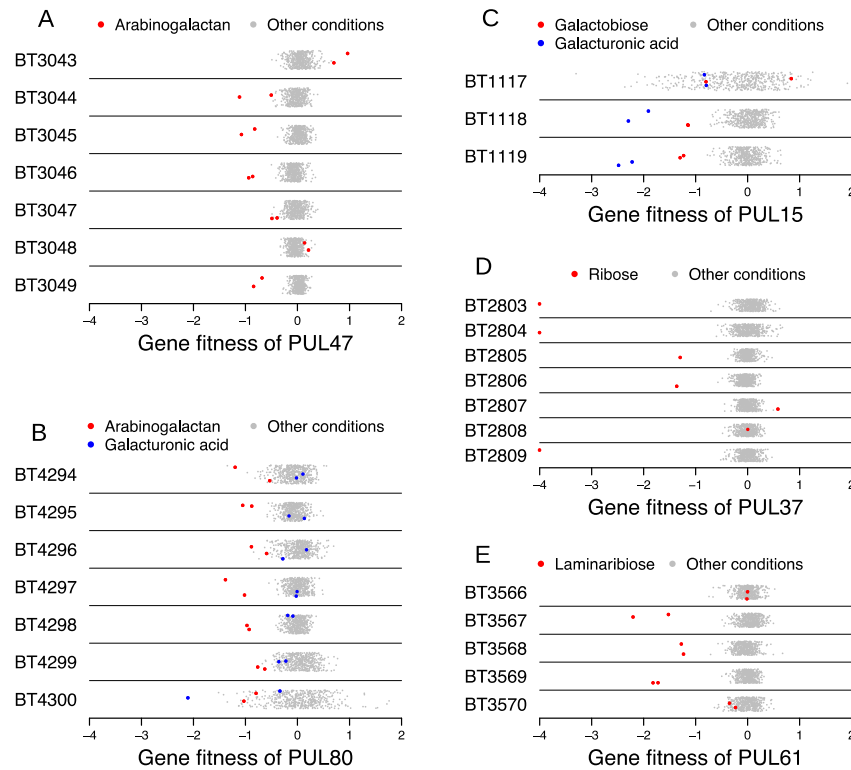
1004
1005
1006 **Figure 4. Fitness data for select PULs provides genetic evidence for novel substrates.** For
1007 the indicated PUL, we show the fitness data for all component genes with certain carbon source
1008 conditions highlighted. The y-axis is random, and values less than -4 are shown at -4.
1009
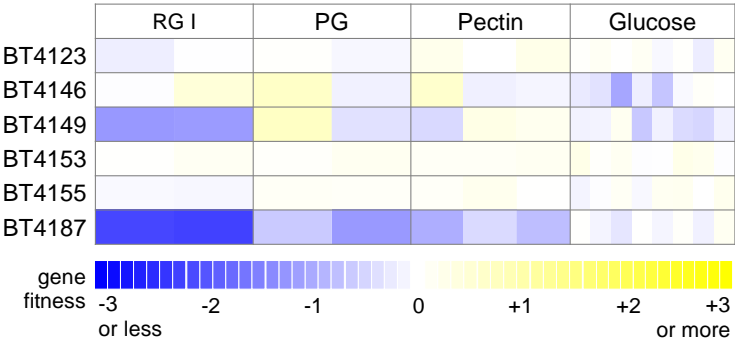
1010
1011
Figure 5. BT4187 is important for type I rhamnogalacturonan utilization. Mutant fitness data for six predicted GH28-family glycosyl hydrolases on type I rhamnogalacturonan (RG-I), polygalacturonic acid (PG), pectin, and glucose. Each replicate experiment is shown separately.
1015
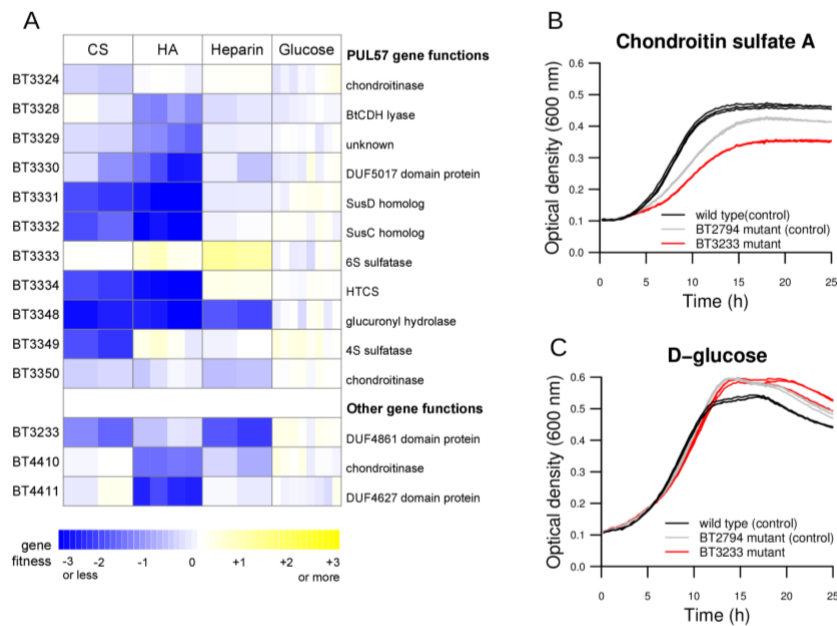
1016
1017 **Figure 6. BT3233 and BT4410 are important for glycosaminoglycan utilization.** (A) Mutant
1018 fitness data for the utilization of chondroitin sulfate A (CS), hyaluronic acid (HA), heparin, and
1019 glucose. Replicate experiments are displayed separated in the heatmap. (B,C) Growth curves of
1020 wild-type *B. thetaiotaomicron*, a BT3233 transposon mutant, and a BT2794 transposon mutant
1021 in (B) CS and (C) glucose (in which no phenotypes were predicted). The BT2794 mutant was
1022 used as a control, as this gene does not have a phenotype during growth on CS, unlike BT3233.
1023 The reduced growth of the BT2794 control relative to wild-type likely reflects the burden of
1024 expressing the erythromycin antibiotic resistance gene during growth in CS.
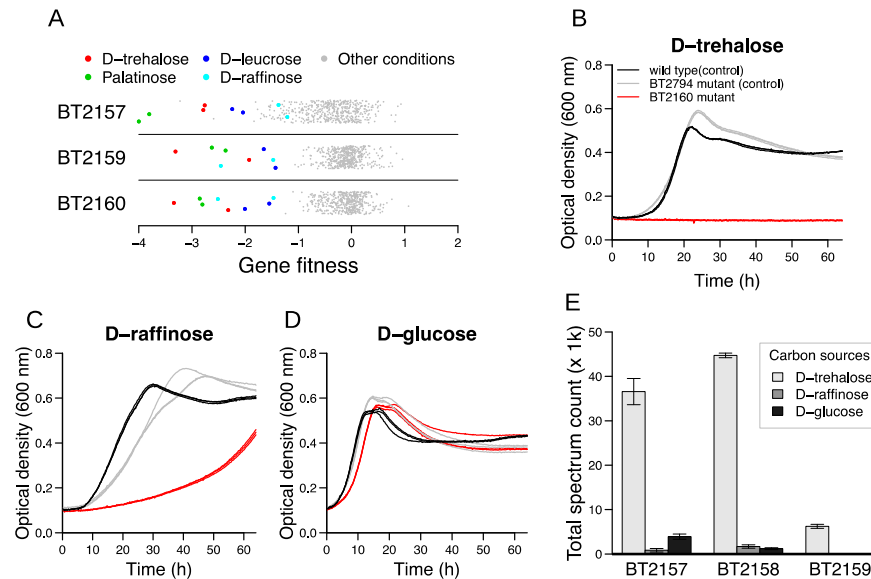1025
1026

1027

Figure 7. Identification of a gene cluster for di- and trisaccharide catabolism in *B. thetaiotaomicron*. (A) Gene fitness values for BT2157, BT2159, and BT2160 under all tested conditions. The *y*-axis is random. (B-D) Growth curves of wild-type *B. thetaiotaomicron*, a BT2160 transposon mutant, and a BT2794 transposon mutant in defined media with either trehalose (B), raffinose (C), or glucose (D) as the sole carbon source. BT2794 does not have a phenotype under any of these conditions in our pooled fitness experiments, and hence was used as a control. Each of three replicate growth curves are plotted (many replicates are highly overlapping). (E) Relative quantification of three proteins (BT2157-2159) that were predicted to be regulated by BT2160 after growth on the indicated carbon sources. Data are the average of 4 replicates. Error bars represent one standard deviation.
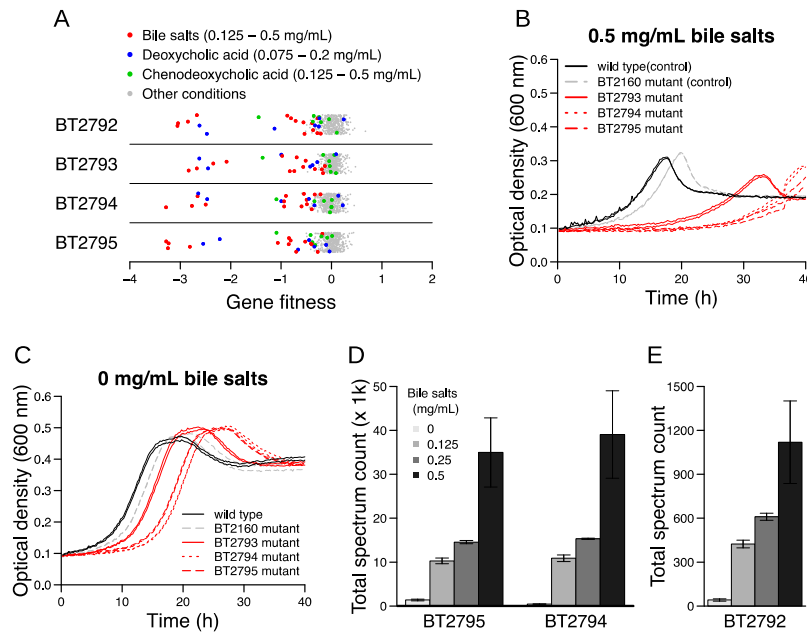
1038

1039

1040
1041
Figure 8. Identification of an efflux system for bile salts in *B. thetaiotaomicron*. (A) Gene
fitness values for BT2792-2795 across all experiments with bile salts and bile acids conditions
highlighted. The highlighted experiments without a significant phenotype correspond to lower
concentrations of the compound. The *y*-axis is random. (B,C) Growth curves of individual *B.
thetaiotaomicron* transposon mutant strains in VB medium with (B) 0.5 mg/mL bile salts or (C)
no added stress. The BT2793-2796 mutants all have extended lag times compared with
controls. For each strain/growth condition, we show two replicate growth curves. (D,E) Relative
abundance of proteins (D) BT2794 and BT2795 or (E) BT2792 from wild-type cells grown in
BHIS with varying concentrations of bile salts demonstrate their upregulation by bile salts in a
concentration-dependent manner. Data are averages of four replicate experiments and error
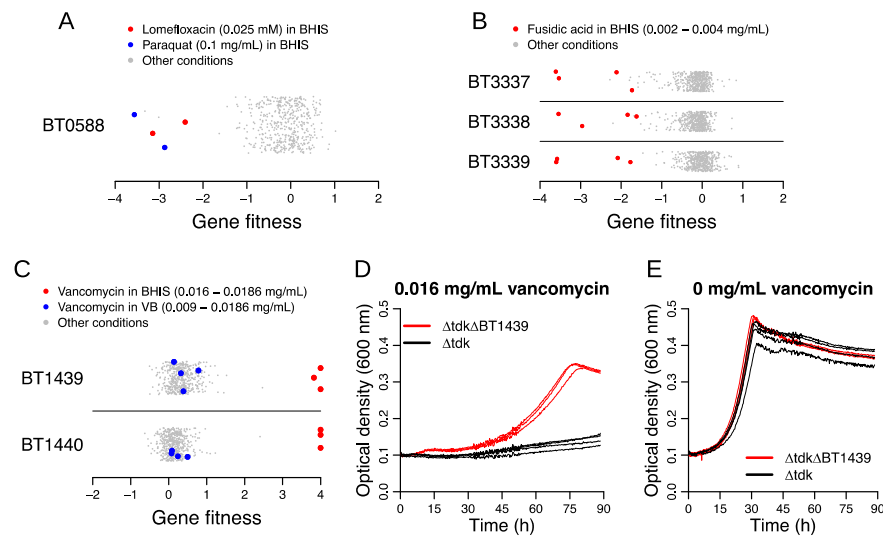bars represent one standard deviation.

Figure 9. Identification of antibiotic resistance genes in *B. thetaiotaomicron.* (A-C) Gene fitness values for BT0588 (also known as BexA) (A), BT3337-3339 (B), and BT1439-1440 (C) under all tested conditions. The *y*-axis in these plots is random. (D,E) Growth curves of a BT1439 deletion and the parental strain (Δ*tdk*) in the presence (D) or absence (E) of vancomycin in BHIS medium. The BT1439 deletion strain exhibits more growth than the parental strain in the presence of vancomycin. Three replicate growth curves per strain/condition are displayed.