HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2021 May 14.

Published in final edited form as:

Cell Rep. 2021 March 02; 34(9): 108789. doi:10.1016/j.celrep.2021.108789.

Functional genetics of human gut commensal *Bacteroides thetaiotaomicron* reveals metabolic requirements for growth across environments

Hualan Liu[1,9], **Anthony L. Shiver**[2,9], **Morgan N. Price**[1], **Hans K. Carlson**[1], **Valentine V. Trotter**[1], **Yan Chen**[3], **Veronica Escalante**[4], **Jayashree Ray**[1], **Kelsey E. Hern**[8], **Christopher J. Petzold**[3], **Peter J. Turnbaugh**[4,5], **Kerwyn Casey Huang**[2,5,6], **Adam P. Arkin**[1,7], **Adam M. Deutschbauer**[1,8,10,*]

[1]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[2]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

[3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[4]Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA 94143, USA

[5]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

[6]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA

[7]Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA

[8]Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA 94720, USA

[9]These authors contributed equally

[10]Lead contact

SUMMARY

Harnessing the microbiota for beneficial outcomes is limited by our poor understanding of the constituent bacteria, as the functions of most of their genes are unknown. Here, we measure the

*Correspondence: amdeutschbauer@lbl.gov.

AUTHOR CONTRIBUTIONS

Conceptualization, H.L., M.N.P., A.P.A., and A.M.D.; software, M.N.P.; investigation, H.L., A.L.S., M.N.P., H.K.C., V.V.T., Y.C., V.E., J.R., and K.E.H.; writing – original draft, H.L., A.L.S., K.C.H., and A.M.D.; writing – review & editing, all authors; funding acquisition, C.J.P., P.J.T., K.C.H., A.P.A., and A.M.D.; resources, A.L.S. and K.C.H.; supervision, C.J.P., P.J.T., K.C.H., A.P.A., and A.M.D.

SUPPLEMENTAL INFORMATION

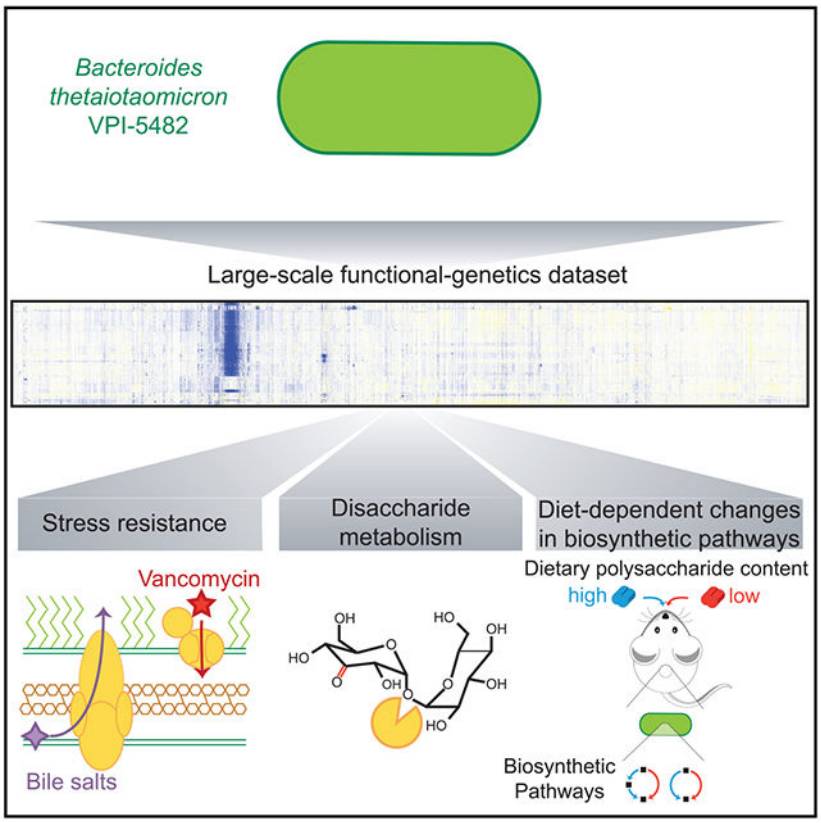
Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2021.108789.

DECLARATION OF INTERESTS

The authors declare no competing interests.

growth of a barcoded transposon mutant library of the gut commensal *Bacteroides thetaiotaomicron* on 48 carbon sources, in the presence of 56 stress-inducing compounds, and during mono-colonization of gnotobiotic mice. We identify 516 genes with a specific phenotype under only one or a few conditions, enabling informed predictions of gene function. For example, we identify a glycoside hydrolase important for growth on type I rhamnogalacturonan, a DUF4861 protein for glycosaminoglycan utilization, a 3-keto-glucoside hydrolase for disaccharide utilization, and a tripartite multidrug resistance system specifically for bile salt tolerance. Furthermore, we show that *B. thetaiotaomicron* uses alternative enzymes for synthesizing nitrogen-containing metabolic precursors based on ammonium availability and that these enzymes are used differentially *in vivo* in a diet-dependent manner.

Graphical Abstract



In brief

Liu et al. report large-scale, genome-wide fitness assays for the model human gut commensal *Bacteroides thetaiotaomicron*, both *in vitro* and *in vivo*. These data provide insights into the bacterium's carbohydrate catabolism, stress responses, and adaptation to diet-dependent ammonium availability in the gut.

INTRODUCTION

The human gut microbiota plays important roles in health and disease (Cho and Blaser, 2012; Zitvogel et al., 2017). It is anticipated that the human gut microbiota can be manipulated to treat a number of diseases (Guinane and Cotter, 2013). However, these efforts will be limited without a greater understanding of the functional potential of the bacteria that comprise the human microbiota. A primary bottleneck is that approximately half of protein-coding genes do not yet have a known molecular function (Heintz-Buschart and Wilmes, 2018).

One of the most well-studied and prevalent members of the human gut microbiota is the anaerobe *Bacteroides thetaiotaomicron* VPI-5482 (Wexler and Goodman, 2017). Pioneering work revealed the broad range of complex polysaccharides and simple carbohydrates that *B. thetaiotaomicron* can degrade and ferment in the gut (Kotarski and Salyers, 1984; Salyers and O'Brien, 1980; Salyers et al., 1977). Genome analysis found that many carbon degradation systems are encoded within polysaccharide utilization loci (PULs) (Xu et al., 2003), which typically contain a cluster of genes involved in the regulation, transport, and catabolism of polysaccharides (Bjursell et al., 2006). The most well-studied PUL in *B. thetaiotaomicron* is the starch utilization system (Sus), encoded by *susRABCDEFG* (Martens et al., 2009). 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, computational analyses have identified >80 predicted PULs in the *B. thetaiotaomicron* genome (Martens et al., 2008) but the majority are experimentally uncharacterized.

Although traditional single-gene genetics and enzymology are useful for elucidating protein function, such approaches are often time-consuming and laborious. Thus, large-scale functional-genetic methods are attractive for the characterization of genes of unknown function (Gray et al., 2015). Methods based on transposon-site sequencing (TnSeq) are useful for assaying the importance of many genes in parallel under multiple conditions (van Opijnen and Camilli, 2013). For example, transposon insertion sequencing (INSeq) applied to *B. thetaiotaomicron* revealed genes required for colonizing mice (Goodman et al., 2009). Subsequent application of INSeq to multiple *Bacteroides* species identified species-specific genes important *in vivo* (Wu et al., 2015). The development of large, multi-condition gene-phenotype datasets has proven useful for inferring gene function in a broad range of organisms (Deutschbauer et al., 2011; Hillenmeyer et al., 2010; Nichols et al., 2011), although this approach has not been applied to gut commensals.

Using a barcoded variant of TnSeq (RB-TnSeq) (Wetmore et al., 2015), we performed hundreds of *in vitro* genome-wide fitness assays in *B. thetaiotaomicron*, including in the presence of simple sugars, complex polysaccharides, and antibiotics. We further profiled fitness during mono-colonization of germ-free mice. This rich dataset provides experimental evidence for many predicted gene functions in *B. thetaiotaomicron* and links genes that previously lacked informative annotations to specific processes including polysaccharide degradation, disaccharide catabolism, and bile salt tolerance. Our data also revealed that

alternative enzymes are used to synthesize nitrogen-containing precursor metabolites in response to diet-dependent availability of ammonium in the gut.

RESULTS

A gene-phenotype map of *B. thetaiotaomicron*

To enable high-throughput genetics in *B. thetaiotaomicron* VPI-5482, we generated a genome-wide RB-TnSeq library in this bacterium. The final library contains 315,668 uniquely barcoded mutants with insertions that we could confidently map to a single location in the genome (Figure 1A). To enable follow-up studies, we generated a 40-plate archived collection of individual transposon-insertion mutants in a 96-well microplate format by sorting cells using flow cytometry (Shiver et al., 2019).

To validate the utility of the mutant library, we grew it in a defined medium with either glucose or xylose as the sole carbon source. For each experiment, we measured the abundance of the DNA barcodes by sequencing (BarSeq) before and after selection, and we defined the fitness of a gene as the average $\log_2$ change in relative abundance of its mutants (Figure 1B). Negative and positive values mean that the mutants in that gene were less and more fit, respectively, than the average strain. Many genes involved in amino acid biosynthesis were important for fitness in either carbon source because the medium lacked most amino acids. Genes from the xylose utilization pathway (BT0792, xylulose kinase; BT0793, xylose isomerase; BT0794, xylose transporter) were important for growth on xylose only (Figure 1B). This example illustrates the biological consistency of fitness assays using our mutant library.

To create a gene-phenotype compendium, we performed 523 genome-wide fitness assays that passed metrics for internal consistency (Wetmore et al., 2015; STAR methods), and all analyses in this study are derived from these experiments. Many experiments were in defined Varel-Bryant (VB) media with a single carbon source. In addition, we performed hundreds of experiments with an inhibitory concentration of a stressor such as an antibiotic, biocide, or bile salt. Finally, we measured the *in vivo* fitness of each mutant during mono-colonization of germ-free mice. We successfully quantified fitness for 4,055 protein-coding genes in each experiment. Among the genes with no fitness data, we estimate that 378 are likely essential for viability in the brain heart infusion supplemented (BHIS) medium used to generate the library (Table S1). Of these genes, 202 (53%) were also identified as candidate essential genes in a previous study of transposon insertion mutants grown in a rich medium similar to BHIS medium (Goodman et al., 2009).

We identified 516 genes with a specific phenotype (Table S2), which we defined as a gene-condition combination with $|\text{fitness}| > 1$ that was statistically significant and for which the gene lacked a significant fitness change in most other experiments (STAR methods). A specific phenotype often links a gene to a condition that is directly related to its function (Price et al., 2018). For example, BT3788 is a SusC homolog located in the PUL for mannan utilization (Cuskin et al., 2015); across all 48 carbon sources profiled, BT3788 was important only for fitness on mannan (Figure 1C). BT4507 encodes a class A serine β-lactamase (Stentz et al., 2015); mutants in this gene were impaired on two β-lactam

antibiotics (ceftriaxone and cephalothin) but not under any of the other conditions profiled, including other antibiotics (Figure 1D).

A second method for inferring gene function is co-fitness (Price et al., 2018), a measure of the degree to which pairs of genes share similar fitness profiles across a large genetic dataset (Table S3). To illustrate this approach, we used co-fitness to identify the enzyme responsible for the first step of arginine biosynthesis in *B. thetaiotaomicron*. Bacteroides species use a non-canonical pathway for synthesizing arginine, in which intermediates are succinylated instead of acetylated (Cabrera-Luque, 2010; Shi et al., 2006). A majority of the enzymes known to be involved in arginine synthesis in *Escherichia coli* (Caldara et al., 2008) have clear homology to enzymes in *B. thetaiotaomicron*; a notable exception is the first step, in which N-acetylglutamate synthase appears to have been replaced by an evolutionarily unrelated N-succinylglutamate synthase. BT3761 has been predicted to fill this role, in part because it is in a 5-gene operon of arginine biosynthetic genes. In addition, BT3761 belongs to the GNAT family of N-acyltransferases and is 59% identical (protein sequence) to an unusual N-acetyl-glutamate synthase from *Caldithrix abyssi* (Cabrera-Luque, 2010; Kublanov et al., 2017). The four most co-fit genes with BT3761 are involved in arginine biosynthesis (BT3759, homologous to *argC* in *E. coli*; BT3733, *argH*; BT3760, *argG*; and BT3395, *argB*). Furthermore, BT3761 was critical for growth on minimal media without amino acids, and its fitness defect was rescued through addition of arginine (Figure 1E). Additionally, quantitative real-time reverse-transcriptase PCR (RT-PCR) revealed that BT3761 RNA levels in wild-type cells were repressed 6.4-fold upon addition of 3 mM arginine (95% confidence interval: 5.3- to 7.8-fold repression). These results support the hypothesis that BT3761 participates in arginine biosynthesis as an N-succinylglutamate synthase.

Overall, we identified a functional association (a specific phenotype, or co-fitness of ≥0.8) for 633 of the 4,055 protein-coding genes for which we collected data. Genes with a functional association were more likely to have a specific Clusters of Orthologous Groups (COG) functional category (Tatusov et al., 2003) assigned than genes with no functional association (55% versus 39%; $p = 10^{-15}$, Fisher's exact test). Nonetheless, we were able to identify a functional association for 270 genes without a specific COG function category. According to the PaperBLAST database (Price and Arkin, 2017), 174 of these genes have not been previously discussed in the literature, nor do they have close homologs (>70% amino acid identity) to other *Bacteroides* species.

Functional-genetic screen links PULs to specific substrates

Fitness profiling across multiple substrates provided a broad view of the importance of each PUL component to carbon catabolism. The 48 carbon sources we assayed represent a range of substrate complexity, including 14 monosaccharides and their derivatives, 12 disaccharides, 1 trisaccharide, 2 tetrasaccharides, 3 oligosaccharides (>4 monomers), 2 sugar alcohols, and 14 complex polysaccharides. As an illustration, we re-examined the Sus system by comparing its phenotypes during growth on starch; the component polysaccharides of starch (amylopectin and amylose); the oligosaccharide α-cyclodextrin; and the simpler break-down products glucose, maltose, maltotetraose, and maltohexaose

(Figure 1F). The SusCD transporter proteins were important on all saccharides of a length of ≥4 but were largely dispensable for growth on maltose and glucose, which is in agreement with previous findings that SusD importance varies depending on the size of glucose polymers (Koropatkin et al., 2008).

The three glycoside hydrolases (GHs; SusA, SusB, and SusG) were variably important across substrates. SusA was previously characterized as a neopullulanase that cleaves α-1,4 linkages in pullulan, a polymer of maltotriose subunits, and was shown to be mildly important for growth on starch (D'Elia and Salyers, 1996; Reeves et al., 1996). In our data, SusA was only important for growth on α-cyclodextrin, a cyclic oligosaccharide of 6 glucose monomers linked by α-1,4 glycosidic bonds (Figure 1F). SusA homologs in other species are often annotated as a cyclomaltodextrinase (enzyme that linearizes circular oligosaccharides), and homologs are required for growth on α-cyclodextrin in *Caulobacter crescentus* and *Echinocola vietnamensis* KMM 6221 (Price et al., 2018). These data suggest that SusA also has cyclomaltodextrinase activity, as previously proposed (Smith and Salyers, 1991).

SusB is a GH97 family member that hydrolyzes multiple linkages, including α-1,4 bonds (Kitamura et al., 2008). In previous work, a Δ*susB* mutant had negligible growth defects on starch, raising doubts about the role of SusB in the use of starch and related compounds (D'Elia and Salyers, 1996; Reeves et al., 1996). We found that insertions in *susB* had fitness defects on several forms of starch (Figure 1F), motivating us to re-examine the role of SusB in polysaccharide utilization. To confirm that we were measuring a bona fide phenotype of *susB* and not polar effects on downstream genes like *susCD*, we generated a clean deletion. The strongest fitness defects of *susB* insertions were during growth on α-cyclodextrin, and we found that Δ*susB* indeed had a growth defect on this substrate compared to glucose (Figures 1G and 1H). We used quantitative real-time RT-PCR to test if the *susB* mutation disrupted the expression of *susCD* during growth on α-cyclodextrin. Instead, we found that expression levels of *susC* and *susD* were 2.7-fold higher in Δ*susB* cells than in wild-type cells (*susC*: 2.7-fold, 95% confidence interval: 2.0–3.7; *susD*: 2.7-fold, 95% confidence interval: 2.1–3.4).

SusG breaks down extracellular starch into simpler oligosaccharides (Koropatkin and Smith, 2010). Consistent with this function, in our fitness data, *susG* was highly important for growth on starch and amylopectin and to a lesser extent on amylose and α-cyclodextrin (Figure 1F). Although the outer membrane proteins SusE and SusF have been demonstrated to bind starch (Cameron et al., 2012), their precise function remains unclear. We found that these proteins were only important for growth on amylopectin, suggesting that SusEF preferentially bind this component of starch or they are important for stabilization of the SusG-starch complex, as previously proposed (Foley et al., 2018; Tuson et al., 2018). Taken together, our re-evaluation of the Sus based on genetics data across multiple substrates adds insight into this well-studied system.

To globally associate PULs to carbon substrates, we identified instances in which at least one gene in a given PUL had a specific and important phenotype (fitness, <−1). Based on these criteria, we associated 20 of the 93 known PULs (Terrapon et al., 2018) to 32

substrates (Table S4; Figure 2). We identified specific and important phenotypes for ≥3 genes from multiple PULs on their predicted or known polysaccharides (Table S4; Figure 2), as follows: PUL5 on arabinogalactan (Cartmell et al., 2018), PUL22 on fructans (Sonnenburg et al., 2010), PUL48 on dextran (Rogers et al., 2013), PUL57 on chondroitin sulfate (CS) and hyaluronic acid (HA) (Raghavan et al., 2014), PUL68 on mannan (Cuskin et al., 2015), PUL75 on pectic glycans (Luis et al., 2018), PUL85 on heparin (Cartmell et al., 2017), and PUL86 on type I rhamnogalacturonan (Luis et al., 2018).

In addition to confirming the results of previous studies, our screen provided genetic evidence for the prediction that, besides PUL5, PUL47 and PUL80 are also important for arabinogalactan utilization (Martens et al., 2011). The arabinogalactan phenotypes for genes in PUL47 and PUL80 were mild (fitness values ranged from −0.7 to −1.4), but similar phenotypes were observed for ≥3 genes in each PUL, and the genes in these clusters did not have clear phenotypes under any other conditions profiled (Figures 3A and 3B). Even though these genes did not meet the criteria for a specific phenotype, mutants in these genes had statistically significant defects, all with $|t| > 4$, and an estimated false discovery rate of <5% (STAR methods). These data demonstrate that efficient breakdown of certain complex polysaccharides like arabinogalactan requires multiple PULs. Additional genes (PUL65 and BT3674–BT3687) are involved in removing glucuronic acid units from gum Arabic arabinogalactan (Cartmell et al., 2018), but these genes had weak phenotypes (fitness, > −0.5) during growth on larch wood arabinogalactan, consistent with the low glucuronic acid content of larch wood arabinogalactan (Willför et al., 2002).

Our analysis also linked PULs to specific monosaccharides and disaccharides. For example, the SusCD homologs of PUL15 (BT1118 and BT1119) were important on galacturonic acid and 1,4-β-D-galactobiose (Figure 3C). Consistent with Glowacki et al. (2020) results, components of PUL37 were only important for growth on D-ribose, in particular a hypothetical protein (BT2802; fitness, −4.3), two predicted ribokinases (BT2803 and BT2804), and a sugar transporter (BT2809) (Figure 3D). The SusCD homologs in this cluster were also mildly important on D-ribose (BT2805 fitness, −1.3; BT2806 fitness, −1.4). Finally, a putative glucosidase (BT3567), SusD-like protein (BT3568), and SusC-like protein (BT3569) from PUL61 were required for optimal growth on the disaccharide laminaribiose (Figure 3E). Although the PULs of *B. thetaiotaomicron* have typically been studied in the context of polysaccharides, our data suggest that some PULs evolved to catabolize simpler substrates. In support of this view, PUL37 and PUL61 contain the core SusCD proteins but do not have homologs of SusEF or hydrolases expected to act on complex substrates.

Identification of polysaccharide degradation genes

Although the degradation pathways for some polysaccharides are well described in *B. thetaiotaomicron*, prior work found that genes important in glycan utilization can be located outside the corresponding PUL (Benjdia et al., 2011; Raghavan et al., 2014). By examining genes with specific and important phenotypes, we identified three genes for polysaccharide utilization. The first example is in metabolism of type I rhamnogalacturonan (RG-I), a polysaccharide containing repeats of a disaccharide of galacturonic acid and rhamnose. A

recent study proposed the involvement of five GH28-family glycosyl hydrolases (BT4123, BT4146, BT4149, BT4153, and BT4155) in breaking down RG-I (Luis et al., 2018). In our fitness data, only BT4149 was important for utilizing RG-I, and this phenotype was mild (fitness, −1.3). Rather, we found that another GH28-family member, BT4187, was more important for growth on RG-I (fitness, −2.2) (Figure 3F). In support of the importance of BT4187 in RG-I degradation, the operon containing BT4187 is predicted to be in the same regulon as PUL75 (Ravcheev et al., 2013), which is both important for fitness (Figure 2) and induced during growth on RG-I (Martens et al., 2011). Although it is possible that the other hydrolases are functionally redundant, our results demonstrate that the loss of BT4187 alone results in a strong fitness defect during growth on RG-I.

The next two examples associate genes of previously unknown function with utilization of host-derived glycosaminoglycans (GAGs), which are structural components of animal tissues and are composed of repeating disaccharide units of a hexuronic acid linked to an amino sugar. *B. thetaiotaomicron* is capable of using GAGs as nutrients, including CS, HA, and heparin (Cartmell et al., 2017; Ndeh and Gilbert, 2018; Raghavan and Groisman, 2015). CS and HA differ by their repeating disaccharide; CS contains sulfated N-acetylgalactosamine and glucuronic acid, and HA contains N-acetylglucosamine and glucuronic acid. Previous studies proposed a model for CS and HA catabolism by *B. thetaiotaomicron* involving PUL57 and the CS lyase BT4410 (Martens et al., 2008; Raghavan et al., 2014). In our data, components of PUL57 were indeed required for optimal growth on CS and HA (Figure 3G), whereas other components of PUL57 were important on one substrate only. BT3349 (a sulfatase) was only important on CS, as expected given that only CS contains sulfated sugars. The polysaccharide lyase family PL29 enzyme encoded by BT3328 was only important for fitness on HA, despite the activity of this enzyme against both HA and CS *in vitro* (Ndeh et al., 2018).

The first GAG utilization gene we identified was BT4411, which was important for growth on HA but not on CS (Figure 3G). BT4411 is predicted to be in the same operon as the CS lyase BT4410, which was also more important for fitness on HA than on CS. BT4411 contains a domain of unknown function (DUF4627) that has significant similarity to CBM_4_9 (carbohydrate binding domain PF02018) (El-Gebali et al., 2019). The BT4410–BT4411 operon is conserved in several other *Bacteroides* species, and both proteins have putative signal peptides (Yu et al., 2010), suggesting that BT4410 and BT4411 function together in HA catabolism.

Finally, we found that BT3233 was important for growth on the following 3 GAGs: CS (mean fitness, −1.6), HA (mean fitness, −0.7), and heparin (mean fitness, −2.2) (Figure 3G). BT3233 is a member of the uncharacterized protein family DUF4861 (El-Gebali et al., 2019) and belongs to a superfamily of putative carbohydrate binding domains (galactose mutarotase-like domain, InterPro IPR011013) (Mitchell et al., 2019). Homologs of BT3233 are mostly confined to *Bacteroides* species, including a number of human commensals. Using an archived transposon mutant, we confirmed that BT3233 is important for growth on CS (Figure 3H,I). Although its precise function is unknown, our data suggest that BT3233 plays an important role in GAG catabolism.

Disaccharide catabolism through a putative 3-keto intermediate

Our fitness assays revealed that components of a gene cluster (BT2157, BT2159, and BT2160) were important for catabolizing the disaccharides trehalose (glucose-α-1,1-glucose), leucrose (glucose-α-1,5-fructose), palatinose (glucose-α-1,6-fructose), and maltitol (glucose-α-1,4-glucitol) and the trisaccharide raffinose (galactose-α-1,6-glucose-β-1,2-fructose) (Figure 4A). BT2160 encodes a SusR-family transcription factor that likely regulates the cluster (Ravcheev et al., 2013). BT2156 did not have a significant phenotype under any assayed conditions. *B. thetaiotaomicron* contains a second gene cluster (BT4446–BT4449) that is highly similar to BT2156–BT2159 (Figure S1A), except that BT4446–BT4449 does not contain a SusR-family regulator nor did it have a significant phenotype in our data. BT2158 is 98% identical at the nucleotide level to BT4448; hence, we were not able to unambiguously map transposon insertions within these genes by using our standard parameters. To determine if BT2158 was important for fitness, we re-ran our transposon insertion mapping with less stringency by only requiring one nucleotide difference to assign insertions to BT2158 and BT4448, which successfully generated fitness data for both genes. Although BT4448 did not have strong phenotypes, BT2158 was important for growth on trehalose, leucrose, palatinose, maltitol, and raffinose (fitness, ≥ −1.3 in all experiments on these substrates). Furthermore, BT2158 had the highest co-fitness with BT2160 ($r = 0.67$). To confirm these results, we generated individual deletions of BT2158 and BT4448 and assayed growth of these mutants on glucose or trehalose. As expected, ΔBT4448 did not exhibit a growth defect on either substrate (Figures S1E and S1F), whereas ΔBT2158 did not grow on trehalose (Figure S1F). These results demonstrate that BT2157–BT2160 is important for fitness on similar substrates and likely act in the same pathway. A recent genetic and biochemical study found that components of BT2156–BT2160 are also capable of transforming glucosinolates (which are thioglycosides, so there is some chemical similarity to glycosides) to produce isothiocyanates (Liou et al., 2020).

To explore this operon, we examined the protein sequences. BT2156 encodes a predicted lolE-like epimerase, BT2157 encodes a DUF1080-containing protein, and BT2158 and BT2159 encode MviM-like dehydrogenases. DUF1080 has structural similarity to endo-1,3-1,4-β-glucanases (El-Gebali et al., 2019) and is likely a family of hydrolases. Interestingly, homologs of BT2156–BT2159 are chromosomally clustered and are important for catabolizing disaccharides and other glycosides in *C. crescentus, Pedobacter* sp. GW460-11-11-14-LB5, and *E. vietnamensis* KMM 6221 (Figures S1 and S2) (Price et al., 2018). Thus, we considered that *B. thetaiotaomicron* and these other species might catabolize glycosides through similar mechanisms. In *C. crescentus*, homologs of BT2156–BT2159 are in a genome cluster with the glucoside 3-dehydrogenase LacABC (Figure S1B), and these homologs share similar phenotypes on glycosides as LacABC (Figure S2A). LacABC is the first step in lactose catabolism by a 3-keto-lactose intermediate (Arellano et al., 2010). Similarly, sucrose degradation in *Agrobacterium tumefaciens* proceeds through a 3-keto-sucrose intermediate (Bernaerts and De Ley, 1960; Schuerman et al., 1997), and in *Sinorhizobium meliloti*, degradation of various disaccharides occurs through 3-keto intermediates (Ampomah et al., 2013). *A. tumefaciens* and *S. meliloti* use the dehydrogenase ThuAB (unrelated to LacABC) to form 3-keto-disaccharides (Bernaerts and De Ley, 1960; Schuerman et al., 1997). The 3-keto-disaccharides are thought to be hydrolyzed to a hexose

and a 3-keto-hexose (Hayano and Fukui, 1970), followed by reduction of the 3-keto-hexose (Hayano et al., 1973), but the genes involved are not known.

We hypothesized that *B. thetaiotaomicron* degrades glycosides by 3-keto intermediates through activity of BT2157–BT2160, even though *B. thetaiotaomicron* does not have LacABC or ThuAB. We focused on the utilization of trehalose; BT2157–BT2160 had relatively strong phenotypes on trehalose, and homologs of some of these proteins are important for trehalose utilization in *Pedobacter* sp. GW460-11-11-14-LB5 and *E. vietnamensis* KMM 6221 (Figures S2B and S2C). First, we investigated whether expression of this operon was induced on trehalose by using proteomics of stationary-phase cultures. We observed a ≥9-fold increase in abundance for three of the target proteins (BT2157–BT2159) during growth on trehalose compared to glucose, suggesting that this system is regulated by trehalose availability (Figure 4B). Next, we generated a deletion of BT2160 and assayed its ability to grow on trehalose. Similar to ΔBT2158, BT2160 was required for growth on trehalose (Figure 4C). Conversely, neither gene was required for growth on glucose (Figure 4D). Lastly, to provide evidence that the degradation pathway proceeds by a 3-keto-trehalose intermediate, we determined whether the DUF1080 hydrolase encoded by BT2157 was specific to 3-keto-disaccharides. We purified BT2157 and used *Agrobacterium rhizogenes* to produce a supernatant containing 3-keto-trehalose. BT2157 hydrolyzed >70% of the 3-keto-trehalose in the supernatant within 1 min (Figure 4E). In contrast, the concentration of trehalose in the supernatant only decreased by ~14% after 10 min of incubation with BT2157. BT2157 appears to hydrolyze 3-keto-trehalose *in vitro* in the absence of other products of the BT2156–BT2160 operon and more rapidly than the glucosinolates glucoraphanin or glucobrassicin (Liou et al., 2020). Our data demonstrate that BT2157–BT2160 is the primary route of trehalose catabolism in *B. thetaiotaomicron*, that the expression of these proteins is induced by trehalose, and that BT2157 preferentially hydrolyzes 3-keto-trehalose, thereby supporting a model that trehalose degradation proceeds through a 3-keto-glycoside intermediate.

Identification of an efflux system for bile salts

The human gut involves exposure to growth-promoting complex polysaccharides and also to bile salts, which can inhibit bacterial growth. Consequently, commensal bacteria have evolved strategies to cope with bile salts, including removal by efflux pumps, deconjugation by bile salt hydrolases (BSHs), and oxidation and epimerization of hydroxyl groups by hydroxysteroid dehydrogenases (HSDHs) (Begley et al., 2005). BSH and HSDH homologs have been identified in *B. thetaiotaomicron* (Ridlon et al., 2006), and BT2086 has BSH activity *in vitro* and *in vivo* (Yao et al., 2018).

To identify *B. thetaiotaomicron* genes involved in bile salt tolerance, we examined fitness in the presence of inhibitory concentrations of the primary bile acid chenodeoxycholic acid, the secondary bile acids lithocholic acid or deoxycholic acid, or cholic-acid-derived bile salts (a mixture of sodium cholate and sodium deoxycholate, which we refer to generically as “bile salts”). From these data, we identified a four-gene operon, BT2792–BT2795, important for fitness under deoxycholic acid, chenodeoxycholic acid, and bile salt stress (Figure 5A). These genes were not important when stressed with lithocholic acid, at least under the

concentrations assayed. Based on homology (Overbeek et al., 2005), these proteins form a tripartite multidrug efflux system. Interestingly, these genes did not have a significant phenotype under any other conditions profiled (Figure 5A), including dozens of diverse antibiotics, suggesting that this efflux system is specific to bile acids and bile salts. Closely related efflux systems are present in other *Bacteroides* species (Dehal et al., 2010), although the importance of those genes for bile salt tolerance remains to be determined.

To confirm the involvement of this operon in bile salt tolerance, we assayed the growth of archived transposon mutants. BT2793, BT2794, and BT2795 mutants all had longer lag phases and reduced growth rates relative to controls when stressed with 0.5 mg/mL bile salts (Figures 5B and 5C). To determine if this efflux system is induced by the stressor, we performed targeted proteomic analysis by using wild-type cells grown in the presence of various concentrations of bile salts. BT2792, BT2794, and BT2795 were all induced in the presence of bile salts, and the amount of induction increased with higher bile salt concentrations (Figures 5D and 5E), consistent with previous observations of bile-salt-mediated upregulation of efflux pumps in other human gut commensals (Lin et al., 2005; Prouty et al., 2004; Pumbwe et al., 2007; Ruiz et al., 2013). Taken together, our data show that the efflux system encoded by BT2792–BT2795 is induced by and important for resistance to bile salts.

Association of antibiotics with transport systems

The use of antimicrobial compounds to treat infection means that gut commensals are regularly exposed to these inhibitors; thus, we examined fitness with different classes of inhibitors including antibiotics and biocides. Given the prevalence of efflux as a mechanism of antibiotic resistance in bacteria (Webber and Piddock, 2003), we focused on transport-related proteins with strong phenotypes. For example, BexA (BT0588) belongs to the multidrug and toxic compound extrusion (MATE) family and has been shown to be responsible for efflux of the fluoroquinolone antibiotics norfloxacin and ciprofloxacin in *B. thetaiotaomicron* (Miyamae et al., 2001). In our dataset, BT0588 mutants exhibited a mild fitness defect with ciprofloxacin (minimum fitness, −1.6). Insertions in BT0588 resulted in a more pronounced phenotype with another fluoroquinolone (lomefloxacin) or with paraquat (Figure 6A). In another example, we found that the RND transporter system encoded by BT3337–BT3339 was important for fusidic acid tolerance (Figure 6B) and to a lesser extent for tolerating cefoxitin and the antipsychotic drug chlorpromazine.

Lastly, we identified a two-gene operon, BT1439–BT1440, with high positive fitness when grown in the presence of vancomycin (Figure 6C), a glycopeptide antibiotic that inhibits peptidoglycan biosynthesis in the periplasm. Although vancomycin is usually thought to target Gram-positive bacteria, it will inhibit Gram-negative bacteria if it passes through the outer membrane (Antonoplis et al., 2019). Vancomycin is commonly used for treating *Clostridium difficile* infections (CDIs) and is often taken orally. Fecal levels of vancomycin in CDI patients can exceed 2 mg/mL during treatment (Gonzales et al., 2010), which is above the concentrations used in our screen (0.009–0.019 mg/mL). Thus, our data suggest that vancomycin use in humans can impact *B. thetaiotaomicron* in the gut (Basolo et al., 2020). BT1439 is a SusD homolog and BT1440 is a SusC homolog, and these genes had

high co-fitness ($r = 0.75$), indicating that they likely act together. Although SusCD complexes are best known for translocating carbohydrate substrates, no growth defect was found on any of the carbohydrates assayed. To validate our data, we constructed a BT1439 deletion mutant and found that it was more resistant to vancomycin than the parent (Figures 6D, 6E, and S3). Interestingly, the vancomycin phenotype of BT1439–BT1440 was only observed in BHIS medium and not in VB medium (Figure 6C). This medium-specific phenotype could be due to altered regulation or to vancomycin binding to another compound in the growth medium (such as the peptides in BHIS).

B. thetaiotaomicron uses alternative biosynthetic enzymes to adapt to ammonium availability *in vivo*

We were interested in whether data from our *in vitro* chemical-genetic screen could illuminate fitness phenotypes *in vivo*. To generate *in vivo* fitness data, we monocolonized germ-free mice with our barcoded library; collected fecal samples at days 2, 4, and 7 post-gavage; and quantified gene fitness using BarSeq. We first compared our *in vivo* data to those of a previous study that colonized mice with a set of *Bacteroides* species including *B. thetaiotaomicron* VPI-5482, in both a low-fat, high-plant polysaccharide (LF/HPP) and a high-fat, high-sugar (HF/HS) diet (Wu et al., 2015). Our fitness data were more correlated with the LF/HPP ($r = 0.52$, 95% confidence interval: 0.46–0.58) (Figure 7A) than the HF/HS diet ($r = 0.33$, 95% confidence interval: 0.25–0.40), as expected based on the mouse diet in our *in vivo* experiment (STAR methods). This correlation in part reflects the nutrients available *in vivo* to *B. thetaiotaomicron*, as there was significant overlap in genes related to amino acid biosynthesis (odds ratio = 28, 95% confidence interval: 9–92) and carbohydrate-active enzymes (CAZymes) (odds ratio = 40, 95% confidence interval: 15–102) (Figure 7B). Thus, our data are consistent with the previous study of *B. thetaiotaomicron* fitness *in vivo* in the context of a high-plant-polysaccharide diet.

Although insertions in amino acid biosynthesis genes had immediate deleterious effects on growth in our dataset, the fitness defects of CAZymes tended to be smaller at first and increased in magnitude over time (Figure 7C, top), likely reflecting the foraging of *B. thetaiotaomicron* on a diverse set of carbohydrates in the gut (Sonnenburg et al., 2005). Nonetheless, we identified significant growth defects *in vivo* for genes in multiple PULs, including the arabinan-activated PUL7 (BT0350, BT0351, and BT0367), and predicted PUL27 (BT2262–BT2264) (Figure 7C, bottom); PUL27 is constitutively expressed *in vitro* (Glenwright et al., 2017) but has no clearly defined substrates and was not important for growth under any conditions from our *in vitro* screen. The *in vivo* phenotypes of PUL27 suggest that it functions in the uptake of a diet- or host-derived nutrient that has yet to be identified.

The biosynthetic pathways for arginine and lysine were critical for growth *in vivo*, as previously observed (Goodman et al., 2009; Wu et al., 2015). Our *in vitro* data suggested that for both pathways, alternative enzymes are used depending on ammonium availability. The first example is synthesis of meso-diaminopimelate (mDAP), a precursor to lysine and peptidoglycan. Two alternative pathways for producing mDAP from 2,3,4,5-tetrahydropicolinate are conserved among the *Bacteroides* (Figure 7D). The dehydrogenase

branch, catalyzed in a single step by Ddh (BT1979), uses free ammonium as a nitrogen source. The transaminase branch, catalyzed in two steps by DapL (BT0547) and DapF (BT0548), uses glutamate as a nitrogen source. The functions of DapL and Ddh have been verified biochemically using enzymes from *B. fragilis* (Hudson et al., 2008, 2011), which have 90% and 95% amino acid identity to the *B. thetaiotaomicron* enzymes, respectively. In *Corynebacterium glutamicum*, the dehydrogenase branch preferentially operates when ammonium is abundant; flux through the dehydrogenase branch is correlated with ammonium levels (Sonntag et al., 1993), and mutants in the transaminase branch require high levels of ammonium (Wehrmann et al., 1998). In our *B. thetaiotaomicron* library, *ddh* was important for fitness during growth in minimal media with ammonium but had no phenotype when ammonium was replaced with amino acids or other organic nitrogen sources (Figure 7E). Instead, insertions in *dapLF* were deleterious under these low ammonium conditions (Figure 7E), suggesting that relative flux through the two branches depends on availability of free ammonium in *B. thetaiotaomicron* as well (Figure 7D).

The second example of use of alternative enzymes is carbamoyl phosphate synthase (CPS), which synthesizes a nitrogen-containing precursor (carbamoyl phosphate) to arginine and pyrimidine nucleotides. In eubacteria and eukaryotes, CPS is composed of two subunits; the glutaminase subunit (CarA) liberates ammonium from glutamine and funnels it to the synthetase subunit (CarB), where the ammonium combines with other reactants to form carbamoyl phosphate (Figure 7D; Anderson and Meister, 1965). To date, all eubacterial CPSs have been found to be glutamine dependent. However, eubacterial CarB functions as an ammonium-dependent CPS when purified alone (Trotta et al., 1971), and mutations in the glutaminase subunit of the mitochondrial CPS that participate in the urea cycle make the enzyme ammonium dependent (Saeed-Kothe and Powers-Lee, 2003). *B. thetaiotaomicron* carries a two-gene locus, BT0556–BT0557, encoding a full CPS (CarAB) as well as a lone synthetase subunit BT3866 (CarB2) with no clear glutaminase partner (Figure 7D). Unexpectedly, the fitness defects of insertions in *carAB* were limited to minimal media without added ammonium (Figure 7E). With ammonium present, insertions in *carAB* had no phenotype, whereas *carB2* was very important for fitness (Figure 7E). There were no genes that were both co-fit with *carB2* and had clear sequence homology to a CPS glutaminase subunit. In particular, *carA* (BT0556) was co-fit with *carB* (BT0557) but not *carB2* ($r = 0.85$ and $r = -0.02$, respectively). In other bacteria with two CPSs, one enzyme supplies carbamoyl phosphate for arginine biosynthesis, whereas the second is responsive to needs for pyrimidine biosynthesis (Paulus and Switzer, 1979), regardless of the nitrogen source. We propose that in *B. thetaiotaomicron*, the use of CarAB or CarB2 depends primarily on ammonium availability, with CarAB acting as a prototypical glutamine-dependent enzyme and CarB2 using ammonium to directly synthesize carbamoyl phosphate (Figure 7D).

In our mouse experiment, both *carAB* and *dapLF* were critical for growth *in vivo* (Figure 7E). This observation is consistent with results of a previous study, in which *carAB* and *dapL* insertions in *B. thetaiotaomicron* VPI-5482 were less fit in a LF/HPP diet (similar to our *in vivo* experiment), whereas *carB2* and *ddh* insertions were less fit in a HF/HS diet (Wu et al., 2015). Likewise, homologs of *carB2* and *ddh* were important for fitness in a HF/HS diet but not in a LF/HPP diet in *B. thetaiotaomicron* 7330 (Wu et al., 2015). The phenotypes of *carB2* and *ddh* were also greatly diminished in *B. ovatus* ATCC8483 when an

arabinoxylan supplement was added to the diet (Wu et al., 2015). Because the diet was otherwise held constant, this last result shows that addition of microbiota-accessible carbohydrates is sufficient to switch *Bacteroides ovatus* from use of Ddh and CarB2 to DapLF and CarAB and suggests that it is the presence of microbiota-accessible carbohydrates in a LF/HPP diet (in contrast to a HF/HS diet) that has the same effect on other *Bacteroides* species.

We propose the following model to synthesize these observations (Figure 7F). The fermentation of protein by the colonic microbiota is a major source of ammonium in the gut (Wrong et al., 1985), but the presence of fermentable carbohydrates significantly reduces the amount of ammonium produced (Birkett et al., 1996; Mortensen, 1992). Thus, when *B. thetaiotaomicron* grows in the context of a diet high in plant polysaccharides, it ferments the undigested carbohydrates and ammonium concentration in the gut remains low. In this environment, DapLF and CarAB are the preferred enzymes. Conversely, when *B. thetaiotaomicron* grows in a low-plant-polysaccharide environment, protein fermentation plays a larger role in energy generation and excess ammonium is liberated from the catabolic deamination of amino acids (Allison and Macfarlane, 1989; Dai et al., 2011; Smith and Macfarlane, 1996). *B. thetaiotaomicron* then takes advantage of the excess ammonium by directly incorporating it into nitrogen-containing compounds using Ddh and CarB2. Both *carAB* and *dapLF* are found in the high-affinity ammonium locus of *B. thetaiotaomicron* VPI-5482, and their expression is induced *in vitro* when ammonium is limiting (Iakiviak, 2017), suggesting that transcriptional control in response to ammonium levels is a critical regulatory mechanism controlling flux through the alternative pathways.

DISCUSSION

A small fraction of the gene content of the human microbiome has been experimentally characterized, and there are large knowledge gaps in our molecular understanding of these health-relevant bacteria. To address this knowledge gap, we performed hundreds of genome-wide fitness experiments in *B. thetaiotaomicron* and used the data to reannotate the functions of 40 proteins (Table S5). Nineteen of these proteins were annotated as “hypothetical protein” in RefSeq (as of April 2019), and many of the other proteins had vague annotations such as “MFS transporter.” These reannotations are also available in the Fitness Browser (Price et al., 2018) and PaperBLAST (Price and Arkin, 2017). Our large dataset contains phenotypes for hundreds of additional genes and should be a valuable resource for gene function inference and hypothesis generation. To facilitate these future studies, the *B. thetaiotaomicron* fitness data are publicly available for comparative analyses in the Fitness Browser at https://fit.genomics.lbl.gov.

Much of the research to date on *B. thetaiotaomicron* has focused on polysaccharide degradation by PULs. Although our analysis was able to link 20 PULs to carbon sources, these findings represent <25% of the predicted PULs in the genome. Why so few? First, *B. thetaiotaomicron* may contain two or more PULs that act on the same substrate, and this redundancy would mask the impact of any single-gene mutation. Second, the PUL component(s) may perform its activity extracellularly, such that other mutants in the pooled library can complement any growth deficiency *in trans*. In addition to assaying mutants

individually, approaches for performing genome-wide fitness assays within droplets hold the potential for eliminating complementation by other mutants (Thibault et al., 2019). Third, some PULs may not be expressed under our growth conditions or are involved in the breakdown of polysaccharides that we did not profile. For example, transcriptional profiling connected some PULs with N-linked glycans (Briliūtė et al., 2019; Cao et al., 2014). Some of these substrates may be activated only within the host; interestingly, we identified a phenotype for components of PUL27 within mice but not in any *in vitro* assays. Fourth, some PULs may have evolved to consume simpler sugars and oligosaccharides. For example, we identified a PUL (BT3567–BT3569) that was important only for fitness on the disaccharide laminaribiose, suggesting that chemical-genetic profiling with additional simple carbon sources may reveal functions for some PULs. Lastly, the definition of a PUL is very minimal and only requires a SusCD gene pair. Some of these minimal PULs may have other cellular roles. For instance, we showed that a two-gene PUL (BT1439–BT1440) confers susceptibility to vancomycin in BHIS (Figure 6D).

We showed that growth on trehalose and other disaccharides required a 3-keto-disaccharide hydrolase (BT2157). This enzyme has a predicted signal peptide (Almagro Armenteros et al., 2019), implying that these disaccharides are oxidized in the periplasm. However, *B. thetaiotaomicron* does not encode the periplasmic 3-ketoglycoside dehydrogenases LacABC or ThuAB (Ampomah et al., 2013; Arellano et al., 2010; Miyazaki et al., 2018). Moreover, *B. thetaiotaomicron* has a limited electron transport chain that ends at menaquinol:fumarate oxidoreductase (Fischbach and Sonnenburg, 2011), yielding succinate as the reduced end product, which probably precludes the operation of cytochrome *c*-dependent dehydrogenases such as LacABC or ThuAB. Instead, oxidation to a 3-keto-disaccharide is probably performed by BT2158, which was important for growth on trehalose and several other disaccharides but not for growth on 3-keto-trehalose (Figure 4A). BT2158 has a putative Tat export signal (Almagro Armenteros et al., 2019) and is distantly related (Pfam: PF01408) to periplasmic glucose-fructose oxidoreductase (Gfo). Gfo has a tightly bound NADP cofactor that is exported together with the folded protein by the Tat system (Halbig et al., 1999), and it oxidizes glucose to gluconolactone while reducing fructose to sorbitol (Kingston et al., 1996). Hence, we propose that BT2158 uses a tightly bound NAD(P) cofactor to oxidize disaccharides to 3-keto-disaccharides while reducing another as yet unidentified sugar.

Integrating data from gnotobiotic mouse models and our chemical-genetic screen, we presented evidence that *B. thetaiotaomicron* remodels its biosynthetic pathways to directly incorporate ammonium into nitrogen-rich compounds when ammonium is plentiful and uses alternative nitrogen sources when ammonium is scarce. Furthermore, we showed that this capacity is likely an adaptation to diet-dependent ammonium fluctuations in the colon. Interestingly, the genes encoding the ammonium-independent enzymes are all located in the high-affinity ammonium (HAA) locus. Expression of genes in the HAA locus is induced under low-ammonium conditions (Iakiviak, 2017), providing a simple regulatory mechanism for the transition between ammonium-dependent and -independent pathways. We found that the HAA locus carries the genes for a canonical glutamine-dependent carbamoyl phosphate synthetase (CPS). Meanwhile, an orphan CPS synthetase subunit (CarB2) is critical for growth in high-ammonium environments. The lack of support for any glutaminase partner for CarB2 in our dataset suggests that CarB2 functions as an ammonium-dependent CPS

(Figure 7D). Prokaryotic ammonium-dependent CPSs have been described previously, but these proteins are restricted to Archaea (Legrain et al., 1995; Popa et al., 2012), leading to a long-standing hypothesis that eubacteria only encode glutamine-dependent CPSs. Further experimental confirmation that CarB2 is an ammonium-dependent CPS will be an important future step in filling this knowledge gap.

Our approach can be applied generally to uncover gene functions in the human microbiome. In particular, barcoding of mutant libraries enables a rapid assessment of gene importance across many conditions at low cost (Price et al., 2018). Nevertheless, challenges still need to be overcome, including the accelerated development of genetic tools for non-model species in the human microbiota (Liu and Deutschbauer, 2018; Peters et al., 2019) and of scalable phenotypic assays that more accurately reflect the natural ecology of these species. In addition, conservation of a phenotype in multiple species is a powerful indicator of gene function (Price et al., 2018), and the generation of similar gene-phenotype maps in related bacteria will expand the scope of our analysis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Adam Deutschbauer (amdeutschbauer@lbl.gov).

Materials availability—Plasmids, strains, and mutant libraries generated in this study are available from the lead contact with a completed Material Transfer Agreement (MTA).

Data and code availability—Fitness data from the 523 successful experiments is available for comparative analysis in the Fitness Browser at https://fit.genomics.lbl.gov (archived at https://figshare.com/articles/dataset/The_Fitness_Browser_Genome-wide_mutant_fitness_data_for_diverse_bacteria_November_2020_release_/13172087/1). In addition, original fitness data (and all associated metadata) have been deposited to figshare: https://figshare.com/articles/dataset/RB-TnSeq_data_for_Bacteroides_thetaiotaomicron_VPI-5482/13171766/1. The software for RB-TnSeq is available at https://bitbucket.org/berkeleylab/feba. The code for generating the figure panels of Figures 7A-7C and 7E and associated statistics in the text is available as a part of a Code Ocean compute capsule (https://codeocean.com/capsule/3569674/tree/v1).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial strains, plasmids, and oligonucleotides used in this study are listed in the Key Resources Table.

Male germ-free (GF) C57BL/6J mice were maintained within the UCSF Gnotobiotic Core Facility (https://gnotobiotics.ucsf.edu) and fed *ad libitum* autoclaved standard chow diet (Lab Diet 5021). The mice were housed in a single gnotobiotic isolator for the duration of

the bacterial mutant library colonization experiment. All mouse experiments were approved by the University of California, San Francisco (UCSF) Institutional Animal Care and Use Committee and performed accordingly.

METHOD DETAILS

Media, compounds and growth conditions—Growth media were purchased from BD. In this study, we used many of the compounds from a previous investigation of bacterial mutant fitness (Price et al., 2018); most were purchased from Sigma-Aldrich. We also purchased and assayed additional compounds specifically relevant to *B. thetaiotaomicron*, such as polysaccharides and bile salts. The source for some compounds is listed in Table S7.

We typically grew *B. thetaiotaomicron* in an anaerobic chamber at 37°C in one of two growth media: Brain Heart Infusion Supplemented (BHIS) medium using Bacto Brain Heart Infusion (BHI) as the base, or Varel-Bryant (VB) defined medium (Varel and Bryant, 1974). Media recipes are provided in Table S7. VB medium contains methionine, so vitamin B12 is not required.

Construction of a barcoded transposon delivery vector for B. thetaiotaomicron—We previously described a mariner transposon delivery vector library pHLL255 (magic pool) containing hundreds of combinations of promoters and erythromycin drug markers (Liu et al., 2018). To identify a suitable vector for mutagenizing *B. thetaiotaomicron*, we first constructed a small mutant library by conjugating with an *Escherichia coli* strain carrying pHLL255. We performed TnSeq on this library to map transposon insertion locations and their associated DNA barcodes, enabling us to infer which vector design in the magic pool resulted in the mutation (Liu et al., 2018). Analysis of this library revealed that the optimal drug marker for mutagenizing *B. thetaiotaomicron* was the erythromycin resistance gene *ermG* (Goodman et al., 2009; Shoemaker et al., 2001), the optimal promoter for expressing the *mariner* transposase was the native *rpoD* (BT1311) promoter (Goodman et al., 2009), and the optimal promoter for expressing *ermG* was the native *dnaE* (BT2230) promoter.

Although this vector design had the highest insertion efficiency relative to the other designs in the magic pool, it still resulted in strand bias whereby the insertion orientation of the transposon was more often in the same transcriptional orientation as the mutated gene. Thus, in addition to constructing the best but potentially suboptimal vector design from the magic pool results (pTGG45), we also constructed another vector (pTGG46) by replacing the *dnaE* promoter in pTGG45 with a *cepA* hybrid promoter from *B. fragilis* RBF-103 (Bayley et al., 2000). To determine which vector exhibited the best combination of high efficiency and low strand bias, we constructed two small mutant libraries with either pTGG45 or pTGG46. TnSeq analysis of these libraries revealed that both had high insertion efficiency, but while the strand bias was again significant for pTGG45 (60% of insertions were in the same orientation as transcription of the mutated gene), there was essentially no strand bias for pTGG46 (49.4%). Thus, we barcoded pTGG46 (pTGG46_NN1) with millions of random 20-nucleotide barcodes using Golden Gate assembly, as previously described (Liu et al., 2018).

Transposon mutant library construction—We carried out transposon mutagenesis by conjugating *B. thetaiotaomicron* recipient cells with *E. coli* donor cells carrying the barcoded transposon vector library pTGG46_NN1 at 1:1 ratio. Specifically, we grew 10 mL of wild-type *B. thetaiotaomicron* VPI-5482 in BHIS overnight at 37°C. The next morning, we recovered a 2 mL freezer stock of strain AMD776 (*E. coli* WM3064 with pTGG46_NN1) in 50 mL LB supplemented with 50 μg/mL carbenicillin and 300 μM diaminopimelic acid (DAP) at 37°C. When the OD_{600} of the *E. coli* donor strain reached ~1.0, we harvested 15 OD_{600} units of the culture and washed the cells three times with fresh BHIS supplemented with DAP. Then, 15 OD_{600} units of *B. thetaiotaomicron* wild-type cells were harvested, mixed with the washed donor cells, and resuspended in a final volume of 60 μL with BHIS supplemented with DAP. The resuspension was spotted onto 0.45-μm gravimetric analysis membrane filters (Millipore) and incubated anaerobically overnight on BHIS agar plates supplemented with DAP at 37°C. The next day, the conjugation mixture was scraped from the membrane and resuspended in 6 mL fresh BHIS medium supplemented with 10 μg/mL erythromycin and plated at different dilutions on BHIS plates supplemented with 10 μg/mL erythromycin. Plates were incubated at 37°C anaerobically for 48 h to let visible colonies develop. We then pooled ~250,000 colonies and grew the pooled library anaerobically in liquid BHIS supplemented with 10 μg/mL erythromycin and 200 μg/mL gentamicin (to further select against the *E. coli* conjugation donor strain as *B. thetaiotaomicron* is naturally resistant to gentamicin). When the culture reached saturation, we made multiple, single-use glycerol stocks of the final library and extracted genomic DNA for TnSeq analysis. To map the genomic locations of the transposon insertions and link these insertions to their associated DNA barcodes, we used our previously described TnSeq protocol (Wetmore et al., 2015). To map insertions in BT2158, we relaxed our mapping stringency and only required one nucleotide difference relative to BT4448 (BT2158 and BT4448 are 98% identical with only 30 mismatches over 1,476 nucleotides). The final mutant library was named Btheta_ML6.

***In vitro* genome-wide mutant fitness assays—**We performed genome-wide mutant fitness assays as described previously (Price et al., 2018). For *B. thetaiotaomicron*, we thawed an aliquot of the full transposon mutant library, inoculated the entire aliquot into 50 mL of BHIS supplemented with 10 μg/mL erythromycin, and grew the library to mid-log phase. We then collected 6 cell pellets of ~1.0 OD_{600} unit each (the “Time0” sample). We used the remaining cells to inoculate competitive growth assays. Nearly all fitness assays were performed in 1.2 mL of growth medium in either a 24-well transparent microplate (Greiner) or in a 96-deepwell plate (Costar) at a starting OD_{600} of 0.02. For carbon and nitrogen source assays in VB, we washed the cells in VB lacking these nutrients two times. We grew cultures until they reached stationary phase, and then collected cell pellets (the “Condition” sample). We also performed a few “survival” experiments in which we incubated the mutant library aerobically for different amounts of time. After this incubation, we inoculated cells into fresh BHIS to select for cells that survived the stress. We extracted genomic DNA from the TimeO and Condition samples in a 96-well microplate format with a QIAamp 96 DNA Qiacube HT kit (QIAGEN). In addition, we performed genome-wide fitness assays using previously described *E. vietnamensis* KMM6221 and *Pedobacter* sp.

GW460-11-11-14-LB5 barcoded transposon mutant libraries (Price et al., 2018) with the carbon sources substrates lactulose, maltitol, methyl-β-D-galactopyranoside, and lactitol.

We performed barcode sequencing (BarSeq) as previously described (Price et al., 2018, 2019). For most experiments, we performed BarSeq with indexed P2 oligos and a mix of non-indexed P1 oligos with variable lengths of Ns (2-5) to stagger the reads. For some experiments, we used BarSeq oligos with both P1 and P2 indexed to minimize the impact of incorrectly assigned indexes in Illumina HiSeq4000 runs (Sinha et al., 2017). Strain and gene fitness scores were calculated as previously described (Wetmore et al., 2015).

Construction of deletion strains and individual mutant growth assays—We constructed unmarked, in-frame gene deletions using a previously established approach (Mishra and Imlay, 2013). Growth curves of individual mutants were carried out in a 96-well plate with either an Epoch2 (BioTek) or Infinite F200 (Tecan) plate reader housed in an anaerobic chamber. Note that optical densities in microplates are typically lower than those in a standard cuvette. To calculate IC_{50} values for ΔBT1439 and the parent on vancomycin, we used the drc package in R (Ritz et al., 2015). Specifically, we used drm() with the LL.4() model and defined the response as the average OD_{600} at the time that uninhibited cells (no vancomycin) reached their maximum OD_{600} value.

Growth and cell harvesting for RT-qPCR assay of BT3761 repression by arginine—An overnight culture of *B. thetaiotaomicron* VPI-5482 Δ*tdk* in 5 mL of tryptone-yeast extract-glucose (TYG) broth was grown at 37°C in a temperature-controlled Model 2000 incubator (COY Lab Products) in an anaerobic chamber (custom build, COY Lab Products). This overnight culture was pelleted and washed in VB medium three times before being inoculated into several wells of VB (the control condition) and VB supplemented with 3 mM arginine (the experimental condition). Biological replicates were grown in a 2-mL 96-deepwell plate (PlateOne, USA Scientific, Cat. No. 1896-2110). The atmospheric gas was a blend of 85% nitrogen, 10% carbon dioxide, and 5% hydrogen (Praxair, Cat. No. BI-NICDHYC4). OD_{600} was monitored using an Epoch2 (BioTek) microplate reader inside the anaerobic chamber, and cultures were periodically diluted into fresh media to keep the OD_{600} below 0.05 so that cells remained in log phase. After ~24 h of growth, 1.5 mL were transferred into 3 mL of RNAprotect® Bacterial Reagent. The cell suspension was centrifuged for 30 min at 4,000*g* and 4°C in a Sorvall Legend XTR centrifuge (Thermo Scientific). The supernatant was removed, cell pellets were air-dried at room temperature for 5 min, and pellets were flash frozen in liquid nitrogen and stored at −80°C until RNA was extracted the next day.

Growth and cell harvesting for RT-qPCR assay of potential polar effects of Δ*susB*—Overnight cultures of *B. thetaiotaomicron* VPI-5482 Δ*tdk* (the parental control) and *B. thetaiotaomicron* VPI-5482 Δ*tdk* ΔBT3703 (Δ*susB*, the experimental strain) were grown in 5 mL VB with α-cyclodextrin as the sole carbon source at 37°C in a temperature-controlled incubator inside an anaerobic chamber. Cultures were then inoculated into separate wells of VB with α-cyclodextrin and grown in log phase via periodic dilution into fresh medium to keep $OD_{600} < 0.05$. Details of growth conditions (plasticware and other equipment) are the same as the arginine repression experiment above. After ~24 h of growth,

1 mL was transferred into 2 mL of RNAprotect® Bacterial Reagent and the cell suspension was centrifuged for 30 min at 4°C and 3,428*g* in a 5920R centrifuge (Eppendorf). The supernatant was removed, cell pellets were dried at room temperature for 5 min, and pellets were stored at −80°C (without flash freezing) until the RNA was extracted (within 1 week).

RNA isolation and quality control for RT-qPCR experiments—RNA was isolated using the RNeasy® Mini Kit (QIAGEN, Cat. No. 74104) with an on-column DNase I digestion using the RNase-free DNase Set (QIAGEN, Cat. No. 79254). RNA isolation was performed following the kit handbook, with no modifications. We used 2-mercaptoethanol from Sigma-Aldrich (Cat. No. M3148-100ML)to make fresh RLT buffer (from the QIAGEN kit) for every RNA isolation experiment. Every sample was eluted into 50 μL of RNase-free water from the kit. RNA concentration and purity were quantified using UV-vis absorbance spectra measured on a NanoDrop One C (Thermo Scientific, software version 1.4.2). The yield (concentration of RNA) and purity (A260/A280, A260/A230) are reported for every sample in Table S6. After quantification, RNA samples were diluted to their final concentration using values reported from the UV-vis spectrophotometer and stored at −20°C. After using the RNA in RT-qPCR experiments, the integrity of RNA in the remaining volume of samples was quantified using a Fragment Analyzer (Advanced Analytical Technologies, Inc.) and a high-sensitivity RNA kit (15 nt) (Agilent, Cat. No. DNF-472-0500). The RNA Quality Number, as determined by PROsize v. 3.0.16, is reported in Table S6.

Reaction conditions for RT-qPCR experiments—All RT-qPCR experiments were performed using the iTaq Universal SYBR® Green One-Step Kit (Bio-Rad, Cat. No. 1725151). In every experiment, we used a 10 μL final volume with 4 μL of input RNA and 0.3 μM primer. For the arginine-repression experiment, 22 ng of total RNA were used. For Δ*susB* polarity experiments, 100 ng of total RNA were used. The RT-qPCR reaction was performed according to the kit handbook instructions for Bio-Rad CFX96 machines, without modification. Briefly, a single-step incubation for 10 min at 50°C allowed for reverse transcription to proceed, followed by a single-step incubation at 95°C for 1 min to denature the DNA and activate the polymerase. These steps were followed by 40 cycles of a 10 s denaturation step (at 95°C) and a 20 s annealing and extension hybrid step (at 60°C). Finally, we performed a melt-curve analysis by incrementing from 65°C to 95°C in 0.5°C steps, at 5 s per step. Samples were analyzed in Hard-Shell PCR Plates, white shell/white well (Bio-Rad, Cat. No. HSP9655) and run on a Bio-RAD® CFX96 machine using Bio-Rad CFX Manager v. 3.1. In both experiments, three biological replicates were analyzed, with one technical replicate per sample (n = 6). At least one no-template control was run on the same plate for each primer set. A no-reverse transcriptase experiment was run for each sample using the experimental target primers (BT3761 for the arginine expression experiment, and *susC* and *susD* for the *susB* polarity experiment). We designed primers for the amplification of 15 reference target genes in *B. thetaiotaomicron* in addition to the 3 experimental targets and the traditional reference target of 16S rDNA (Table S6). We determined the amplification efficiency of all of these primer pairs using 10-fold RNA dilution experiments (Table S6). We also found the best combination of reference genes for both experiments by running preliminary experiments using a large set of reference genes. A set of reference

genes were chosen that exhibited high stability between experiments and controls and that were distributed across multiple biological processes.

Targeted proteomics—For carbon source experiments, we grew cultures of wild-type *B. thetaiotaomicron* in VB with 20 mM D-trehalose, D-raffinose, or D-glucose overnight to OD_{600} values between 1.0 and 1.5. For bile salt experiments, we grew wild-type *B. thetaiotaomicron* in VB with 20 mM D-glucose supplemented with 0, 0.125, 0.25, or 0.5 mg/mL bile salts for 20 h to OD_{600} values between 0.5 and 0.9. For all experiments, we harvested 2 OD_{600} units of cells and washed two times with 10 mM NH_4HCO_3 before extracting total protein using a methanol-chloroform precipitation method (Chen et al., 2019a). Protein concentration was determined using the Bio-Rad DC protein assay (Lot #5000112) following the manufacturer's protocol, and the same amount of proteins was subjected to standard reduction, alkylation, and trypsin digestion procedures. A targeted selected reaction monitoring (SRM) method was developed to quantify unique peptides of proteins using an in-house constructed *B. thetaiotaomicron* spectral library (Chen et al., 2019b).

SRM targeted proteomic assays were performed on an Agilent 6460 QQQ mass spectrometer system coupled to an Agilent 1290 UHPLC system (Agilent Technologies, Santa Clara, CA). Twenty micrograms of peptides from each sample were separated on an Ascentis Express Peptide C18 column (2.7-mm particle size, 160-Å pore size, 5-cm length × 2.1-mm inside diameter (ID), coupled to a 5-mm × 2.1-mm ID guard column with the same particle and pore size, operating at 60°C; Sigma-Aldrich) operating at a flow rate of 0.4 mL/min via the following gradient: initial conditions were 98% solvent A (0.1% formic acid), 2% solvent B (99.9% acetonitrile, 0.1% formic acid). Solvent B was increased to 40% over 5 min, and then to 80% over 0.5 min, and held for 2 min at a flow rate of 0.6 mL/min, followed by a ramp back down to 2% solvent B over 0.5 min where it was held for 1 min to re-equilibrate the column to original conditions. The eluted peptides were ionized via an Agilent Jet Stream ESI source operating in positive-ion mode with the following source parameters: gas temperature = 250°C, gas flow = 13 L/min, nebulizer pressure = 35 psi, sheath gas temperature = 250°C, sheath gas flow = 11 L/min, capillary voltage = 3500 V, nozzle voltage = 0 V. Data were acquired using Agilent MassHunter v. B.08.02. Acquired SRM data were analyzed by Skyline v. 3.70 (MacCoss Lab Software).

Production of 3-keto-trehalose—Single colonies of *Agrobacterium rhizogenes* K599 were inoculated in a minimal medium (MM) from Kurowski and Darbyshire (1978) with 2% (w/v) trehalose as a carbon source and supplemented with 0.5 g/L yeast extract, and grown overnight at 30°C shaking at 250 rpm. Cells were then harvested by centrifugation, washed twice with 0.1 M phosphate-buffer (pH 7.0) and inoculated into 50 mL of MM+2% trehalose without yeast extract. The next day, an aliquot of this culture in MM was used to inoculate 200 mL of MM+2% trehalose (1:20 ratio). Samples were removed periodically, centrifuged, and filter-sterilized to obtain cell-free supernatants. Consumption of trehalose was monitored using the Trehalose Assay kit (Megazyme) and the formation of keto-sugar was determined by the alkaline method from Fukui and Hayano (1969). The cell-free 3-keto-trehalose-

containing culture supernatant used for fitness and enzymatic assays was obtained after 29 h of growth and contained 17.4 mM of trehalose and 7.9 mM 3-keto-trehalose.

Expression and purification of recombinant BT2157—The sequence encoding BT2157 was PCR-amplified from genomic DNA, cloned into a linearized pET32a plasmid using Gibson Assembly, and transformed into *E. coli* chemically competent DH5-α cells (New England Biolabs). The plasmid sequence was verified by Sanger sequencing and used to transform the production strain *E. coli* BL21 DE3. For protein production, *E. coli* BL21 DE3 cells expressing N-terminal His-tagged BT2157 were inoculated from single colonies into LB with 100 μg/mL ampicillin and grown at 37°C overnight. Production cultures were started by inoculating 10 mL of pre-cultures into 1 L of LB with 100 μg/mL ampicillin. After growth at 37°C with agitation until OD_{600}~0.5, protein expression was induced by the addition of 0.15 mM isopropyl β-D-1 thiogalactopyranoside (IPTG) and cultures were incubated at 37°C for 3 h. Cells were pelleted by centrifugation at 8000_g for 15 min, resuspended in 60 mL of lysis buffer (50 mM potassium phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.8), and stored at −80°C. For protein purification, resuspended cells were thawed on ice and lysed by sonication. Cell debris and unbroken cells were removed by centrifugation at 10,000 *rcf.* for 20 min at 6°C. All purification steps were performed using ice-cold buffers and reagents. Protein purification was performed as previously described (Liou et al., 2020), except with a HisTrap FF column (GE Healthcare) replacing the free resin.

In vitro* endpoint assays with recombinant BT2157—In vitro* endpoint assays with purified BT2157 consisted of 50 μL of cell-free *A. rhizogenes* culture supernatant containing 7.9 mM 3-keto-trehalose, 50 μL of purified protein, and reaction buffer (50 mM potassium buffer, pH 7) up to a total volume of 1 mL. Each reaction contained final concentrations of 395 μM 3-keto-trehalose and 28.8 μM his-tagged BT2157. Reactions were initiated by the addition of substrate and incubated at 30°C. Reactions were interrupted by the addition of 3 mL of 0.1 mM NaOH, and after 3 min of incubation, the absorbance at 340 nm was measured for each sample to determine keto-trehalose concentration (Fukui and Hayano, 1969).

Gnotobiotic mouse experiments—Eight GF male 7-week-old C57BL/6J mice were housed in a single gnotobiotic isolator for the duration of the experiment, fed a standard irradiated diet *ad libitum*, split into 3 cages (n = 2-3 mice/cage), and allowed to acclimate for 3 days prior to colonization. The inoculum was prepared by thawing a 1 mL aliquot of the *B. thetaiotaomicron* transposon library, transferring the entire aliquot into 50 mL of BHIS supplemented with 10 μg/mL erythromycin, and incubating anaerobically at 37°C until the cells reached mid-log phase (16 h). Six bacterial cell pellets were collected as Time0 sample references. Each mouse was colonized by oral gavage with 200 μL of the *B. thetaiotaomicron* transposon library. Stool samples were collected on days 2, 4, and 7 post-colonization to assess longitudinal shifts in mutant abundance using BarSeq. Mice were monitored and weighed daily. Genomic DNA from the Time0 cells and stool samples collected over time was extracted using the ZymoBIOMICS 96 MagBead DNA Kit (D4302) and followed by BarSeq, as described above.

QUANTIFICATION AND STATISTICAL ANALYSIS

Fitness data analysis—Overall, we attempted 812 genome-wide fitness assays in this study, of which 523 passed our metrics for internal consistency (Wetmore et al., 2015). We considered a gene fitness value in a single experiment to be statistically significant if |fitness| > 1 and |*t*| > 4, where *t* is fitness/estimated standard error. Across all 523 experiments × 4,055 genes, 31,717 gene fitness values (1.5%) had ~fitness~ > 1 and |*t*| > 4. By contrast, across 68 control comparisons of a Time0 sample to other Time0 samples from the same day, 135 gene fitness values (0.05%) met these thresholds. The estimated false discovery rate for a significant phenotype was 0.05%/1.5% = 3.3%. From these 523 experiments, we identified genes with a specific phenotype in only one or a handful of conditions using previously described criteria (Price et al., 2018): |fitness| > 1 and |*t*| > 5 and 95th percentile(|fitness|) < 1 and |fitness| > 95th percentile(|fitness|)+0.5. For the analysis of the total number of genes with a specific phenotype, we excluded 1 gene that had a specific phenotype in BHIS medium alone. A specific important phenotype has the same criteria as above except that the fitness score has to be negative. For the specific phenotype analysis of PUL genes on carbon sources (Table S4; Figure 2), we used a subset of 492 experiments (we excluded the mouse experiment, and carbon source experiments with lactulose, maltitol, and 3-keto-trehalose) (https://figshare.com/articles/Large-scale_genetic_characterization_of_the_human_gut_commensal_bacterium_Bacteroides_thetaiotaomicron/7742855/1). We identified pairs of genes with highly correlated patterns of phenotypes across all 523 successful experiments (cofitness) using the linear (Pearson) correlation and a threshold of 0.8 (Price et al., 2018). Most data analyses to infer the putative functions of genes were performed with the Fitness Browser (https://fit.genomics.lbl.gov).

One caveat in most transposon mutagenesis studies is complications due to polarity, whereby insertion in upstream genes of an operon can disrupt the expression and function of downstream genes. Relatedly, it is also possible for the promoter of the drug resistance gene used to select mutants to drive expression of downstream genes (if the transposon is oriented in the same direction as the operon). The transposon that we used to construct our library only contains the drug marker cassette (promoter and gene) and the barcode region; there is no obvious transcriptional terminator. Thus, we expect that if a transposon lands in the same transcriptional orientation as the operon, then downstream genes will be overexpressed. To quantify the potential for polarity in our dataset, we identified instances in which fitness values calculated from plus-strand transposon insertions differed significantly from those calculated from minus-strand insertions. Specifically, we considered gene-condition phenotypes to potentially be due to polar effects if: (1) the absolute value of the fitness calculated using the standard approach with all insertions was > 1, (2) there were at least two strains in each orientation with sufficient data, (3) the average fitness on the two strands differed by ≥ 1, and (4) an unpaired t test on the two sets of fitness values gave $p < 0.01$. Among 80,737 gene-experiment pairs with |fitness| > 1 and ≥ 2 insertions on both strands, we identified 2,248 instances (3%) of potential polarity (Table S9), including 208 genes covered by ≥ 2 separate experiments. Thus, while polarity does appear to occur, the majority of significant phenotypes in our dataset are consistent regardless of transposon orientation. [REMOVED HYPERLINK FIELD]

Data analysis for RT-qPCR experiments—Cq values were determined using the Bio-Rad CFX Manager software running the Bio-RAD® CFX96 Real-Time System. Csv files exported by Bio-Rad CFX Manager were manually converted into qbase-format files and input into qbase+ v. 3.2 for analysis. All biological and technical replicates were treated as equivalent in the analysis. Amplification efficiencies of reference and target genes were calculated for every primer pair and used as part of the analysis. Final fold-changes and 95% confidence intervals were calculated from the CNRQ values and standard errors output by qbase+, with the assumption that measurement errors were log-normal. M and CV values for reference gene stability, along with Cq values of no-template controls and the difference in Cq values of target genes with and without reverse-transcriptase, are reported in Table S6.

Data sources for enrichment analysis—Data on predicted amino acid biosynthesis genes in the *B. thetaiotaomicron* VPI-5482 genome were downloaded from the BioCyc database (https://biocyc.org/; Karp et al., 2019). The genes for predicted carbohydrate active enzyme (CAZymes) in *B. thetaiotaomicron* VPI-5482 were downloaded from the Carbohydrate Active Enzymes database (http://www.cazy.org/; Lombard et al., 2014).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The *B. thetaiotaomicron* VPI 5482 Δ*tdk* strain and the plasmid pExchange-tdk were kindly provided by the James A. Imlay lab at the University of Illinois at Urbana-Champaign. We thank Surabhi Mishra and Mikhail Iakiviak for advice on constructing in-frame deletion mutants. Special thanks to Jessie Turnbaugh and the UCSF Gnotobiotics Core facility staff for technical assistance. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH instrumentation grant S10 OD018174. This work was supported by the Allen Discovery Center at Stanford on Systems Modeling of Infection (to A.L.S. and K.C.H.); National Science Foundation grant 1650113 (to V.E.); National Institutes of Health R01 HL122593 (to P.J.T.); Damon Runyon Cancer Research Foundation grant DRR-42-16 (to P.J.T.); and Laboratory Directed Research and Development (LDRD) funding from Berkeley Laboratory, provided by the Director (Office of Science) of the US Department of Energy (DOE) under contract DE-AC02-05CH11231. This work used resources of the Joint BioEnergy Institute supported by the Office of Science and Office of Biological and Environmental Research of the US DOE under contract DE-AC02-05CH11231. P.J.T. and K.C.H. are Chan Zuckerberg Biohub Investigators.

REFERENCES

Allison C, and Macfarlane GT (1989). Influence of pH, nutrient availability, and growth rate on amine production by Bacteroides fragilis and Clostridium perfringens. Appl. Environ. Microbiol 55, 2894–2898. [PubMed: 2560361]

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, and Nielsen H (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol 37, 420–423. [PubMed: 30778233]

Ampomah OY, Avetisyan A, Hansen E, Svenson J, Huser T, Jensen JB, and Bhuvaneswari TV (2013). The thuEFGKAB operon of rhizobia and agrobacterium tumefaciens codes for transport of trehalose, maltitol, and isomers of sucrose and their assimilation through the formation of their 3-keto derivatives. J. Bacteriol 195, 3797–3807. [PubMed: 23772075]

Anderson PM, and Meister A (1965). Evidence for an activated form of carbon dioxide in the reaction catalyzed by Escherichia coli carbamyl phosphate synthetase. Biochemistry 4, 2803–2809. [PubMed: 5326356]

Antonoplis A, Zang X, Wegner T, Wender PA, and Cegelski L (2019). Vancomycin-Arginine Conjugate Inhibits Growth of Carbapenem-Resistant *E. coli* and Targets Cell-Wall Synthesis. ACS Chem. Biol 14, 2065–2070. [PubMed: 31479234]

Arellano BH, Ortiz JD, Manzano J, and Chen JC (2010). Identification of a dehydrogenase required for lactose metabolism in Caulobacter crescentus. Appl. Environ. Microbiol 76, 3004–3014. [PubMed: 20190087]

Basolo A, Hohenadel M, Ang QY, Piaggi P, Heinitz S, Walter M, Walter P, Parrington S, Trinidad DD, von Schwartzenberg RJ, et al. (2020). Effects of underfeeding and oral vancomycin on gut microbiome and nutrient absorption in humans. Nat. Med 26, 589–598. [PubMed: 32235930]

Bayley DP, Rocha ER, and Smith CJ (2000). Analysis of cepA and other Bacteroides fragilis genes reveals a unique promoter structure. FEMS Microbiol. Lett 193, 149–154. [PubMed: 11094294]

Begley M, Gahan CGM, and Hill C (2005). The interaction between bacteria and bile. FEMS Microbiol. Rev 29, 625–651. [PubMed: 16102595]

Benjdia A, Martens EC, Gordon JI, and Berteau O (2011). Sulfatases and a radical S-adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the prominent human gut symbiont, Bacteroides thetaiotaomicron. J. Biol. Chem 286, 25973–25982. [PubMed: 21507958]

Bernaerts MJ, and De Ley J (1960). The structure of 3-ketoglycosides formed from disaccharides by certain bacteria. J. Gen. Microbiol 22, 137–146. [PubMed: 13799611]

Birkett A, Muir J, Phillips J, Jones G, and O'Dea K (1996). Resistant starch lowers fecal concentrations of ammonia and phenols in humans. Am. J. Clin. Nutr 63, 766–772. [PubMed: 8615362]

Bjursell MK, Martens EC, and Gordon JI (2006). 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. [PubMed: 16968696]

Briliūtė J, Urbanowicz PA, Luis AS, Baslé A, Paterson N, Rebello O, Hendel J, Ndeh DA, Lowe EC, Martens EC, et al. (2019). Complex N-glycan breakdown by gut Bacteroides involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. Nat. Microbiol 4, 1571–1581. [PubMed: 31160824]

Cabrera-Luque JM (2010). Discovery of novel pathways of microbial arginine biosynthesis. Dissertation (Children's National Medical Center).

Caldara M, Dupont G, Leroy F, Goldbeter A, De Vuyst L, and Cunin R (2008). Arginine biosynthesis in Escherichia coli: experimental perturbation and mathematical modeling. J. Biol. Chem 283, 6347–6358. [PubMed: 18165237]

Cameron EA, Maynard MA, Smith CJ, Smith TJ, Koropatkin NM, and Martens EC (2012). Multidomain Carbohydrate-binding Proteins Involved in Bacteroides thetaiotaomicron Starch Metabolism. J. Biol. Chem 287, 34614–34625. [PubMed: 22910908]

Cao Y, Rocha ER, and Smith CJ (2014). Efficient utilization of complex N-linked glycans is a selective advantage for Bacteroides fragilis in extraintestinal infections. Proc. Natl. Acad. Sci. USA 111, 12901–12906. [PubMed: 25139987]

Cartmell A, Lowe EC, Baslé A, Firbank SJ, Ndeh DA, Murray H, Terrapon N, Lombard V, Henrissat B, Turnbull JE, et al. (2017). How members of the human gut microbiota overcome the sulfation problem posed by glycosaminoglycans. Proc. Natl. Acad. Sci. USA 114, 7037–7042. [PubMed: 28630303]

Cartmell A, Muñoz-Muñoz J, Briggs JA, Ndeh DA, Lowe EC, Baslé A, Terrapon N, Stott K, Heunis T, Gray J, et al. (2018). A surface endogalactanase in Bacteroides thetaiotaomicron confers keystone status for arabinogalactan degradation. Nat. Microbiol 3, 1314–1326. [PubMed: 30349080]

Chen Y, Guenther JM, Gin JW, Chan LJG, Costello Z, Ogorzalek TL, Tran HM, Blake-Hedges JM, Keasling JD, Adams PD, et al. (2019a). Automated "Cells-To-Peptides" Sample Preparation Workflow for High-Throughput, Quantitative Proteomic Assays of Microbes. J. Proteome Res 18, 3752–3761. [PubMed: 31436101]

Chen Y, Vu J, Thompson MG, Sharpless WA, Chan LJG, Gin JW, Keasling JD, Adams PD, and Petzold CJ (2019b). A rapid methods development workflow for high-throughput quantitative proteomic applications. PLoS One 14, e0211582. [PubMed: 30763335]

Cho I, and Blaser MJ (2012). The human microbiome: at the interface of health and disease. Nat. Rev. Genet 13, 260–270. [PubMed: 22411464]

Cuskin F, Lowe EC, Temple MJ, Zhu Y, Cameron E, Pudlo NA, Porter NT, Urs K, Thompson AJ, Cartmell A, et al. (2015). Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. Nature 517, 165–169. [PubMed: 25567280]

D'Elia JN, and Salyers AA (1996). Effect of regulatory protein levels on utilization of starch by Bacteroides thetaiotaomicron. J. Bacteriol 178, 7180–7186. [PubMed: 8955400]

Dai Z-L, Wu G, and Zhu W-Y (2011). Amino acid metabolism in intestinal bacteria: links between gut ecology and host health. Front. Biosci 16, 1768–1786.

Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, et al. (2010). MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Res. 38, D396–D400. [PubMed: 19906701]

Deutschbauer A, Price MN, Wetmore KM, Shao W, Baumohl JK, Xu Z, Nguyen M, Tamse R, Davis RW, and Arkin AP (2011). Evidence-based annotation of gene function in Shewanella oneidensis MR-1 using genome-wide fitness profiling across 121 conditions. PLoS Genet. 7, e1002385. [PubMed: 22125499]

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al. (2019). The Pfam protein families database in 2019. Nucleic Acids Res. 47, D427–D432. [PubMed: 30357350]

Fischbach MA, and Sonnenburg JL (2011). Eating for two: how metabolism establishes interspecies interactions in the gut. Cell Host Microbe 10, 336–347. [PubMed: 22018234]

Foley MH, Martens EC, and Koropatkin NM (2018). SusE facilitates starch uptake independent of starch binding in B. thetaiotaomicron. Mol. Microbiol 108, 551–566. [PubMed: 29528148]

Fukui S, and Hayano K (1969). Micro Methods for Determination of 3-Keto-sucrose and 3-Ketoglucose. Agric. Biol. Chem 33, 1013–1017.

Glenwright AJ, Pothula KR, Bhamidimarri SP, Chorev DS, Baslé A, Firbank SJ, Zheng H, Robinson CV, Winterhalter M, Kleinekathöfer U, et al. (2017). Structural basis for nutrient acquisition by dominant members of the human gut microbiota. Nature 541, 407–411. [PubMed: 28077872]

Glowacki RWP, Pudlo NA, Tuncil Y, Luis AS, Sajjakulnukit P, Terekhov AI, Lyssiotis CA, Hamaker BR, and Martens EC (2020). A Ribose-Scavenging System Confers Colonization Fitness on the Human Gut Symbiont Bacteroides thetaiotaomicron in a Diet-Specific Manner. Cell Host Microbe 27, 79–92.e9. [PubMed: 31901520]

Gonzales M, Pepin J, Frost EH, Carrier JC, Sirard S, Fortier L-C, and Valiquette L (2010). Faecal pharmacokinetics of orally administered vancomycin in patients with suspected Clostridium difficile infection. BMC Infect. Dis 10, 363. [PubMed: 21192802]

Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, and Gordon JI (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe 6, 279–289. [PubMed: 19748469]

Gray AN, Koo B-M, Shiver AL, Peters JM, Osadnik H, and Gross CA (2015). High-throughput bacterial functional genomics in the sequencing era. Curr. Opin. Microbiol 27, 86–95. [PubMed: 26336012]

Guinane CM, and Cotter PD (2013). Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. Therap. Adv. Gastroenterol 6, 295–308.

Halbig D, Wiegert T, Blaudeck N, Freudl R, and Sprenger GA (1999). The efficient export of NADP-containing glucose-fructose oxidoreductase to the periplasm of Zymomonas mobilis depends both on an intact twin-arginine motif in the signal peptide and on the generation of a structural export signal induced by cofactor binding. Eur. J. Biochem 263, 543–551. [PubMed: 10406965]

Hayano K, and Fukui S (1970). Alpha-3-ketoglucosidase of Agrobacterium tumefaciens. J. Bacteriol 101, 692–697. [PubMed: 5438043]

Hayano K, Tsubouchi Y, and Fukui S (1973). 3-Ketoglucose reductase of Agrobacterium tumefaciens. J. Bacteriol 113, 652–657. [PubMed: 4144143]

Heintz-Buschart A, and Wilmes P (2018). Human gut microbiome: function matters. Trends Microbiol. 26, 563–574. [PubMed: 29173869]

Hillenmeyer ME, Ericson E, Davis RW, Nislow C, Koller D, and Giaever G (2010). Systematic analysis of genome-wide fitness data in yeast reveals novel gene function and drug action. Genome Biol. 11, R30. [PubMed: 20226027]

Hudson AO, Gilvarg C, and Leustek T (2008). Biochemical and phylogenetic characterization of a novel diaminopimelate biosynthesis pathway in prokaryotes identifies a diverged form of LL-diaminopimelate aminotransferase. J. Bacteriol 190, 3256–3263. [PubMed: 18310350]

Hudson AO, Klartag A, Gilvarg C, Dobson RCJ, Marques FG, and Leustek T (2011). Dual diaminopimelate biosynthesis pathways in Bacteroides fragilis and Clostridium thermocellum. Biochim. Biophys. Acta 1814, 1162–1168. [PubMed: 21616177]

Iakiviak M (2017). Analysis of the ammonium assimilation pathways of the human colonic bacterium, Bacteroides thetaiotaomicron. Dissertation (University of Illinois at Urbana-Champaign).

Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, et al. (2019).The BioCyc collection of microbial genomes and metabolic pathways. Brief. Bioinform 20, 1085–1093. [PubMed: 29447345]

Kingston RL, Scopes RK, and Baker EN (1996). The structure of glucose-fructose oxidoreductase from Zymomonas mobilis: an osmoprotective periplasmic enzyme containing non-dissociable NADP. Structure 4, 1413–1428. [PubMed: 8994968]

Kitamura M, Okuyama M, Tanzawa F, Mori H, Kitago Y, Watanabe N, Kimura A, Tanaka I, and Yao M (2008). Structural and functional analysis of a glycoside hydrolase family 97 enzyme from Bacteroides thetaiotaomicron. J. Biol. Chem 283, 36328–36337. [PubMed: 18981178]

Koropatkin NM, and Smith TJ (2010). SusG: a unique cell-membrane-associated alpha-amylase from a prominent human gut symbiont targets complex starch molecules. Structure 18, 200–215. [PubMed: 20159465]

Koropatkin NM, Martens EC, Gordon JI, and Smith TJ (2008). Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. Structure 16, 1105–1115. [PubMed: 18611383]

Kotarski SF, and Salyers AA (1984). Isolation and characterization of outer membranes of Bacteroides thetaiotaomicron grown on different carbohydrates. J. Bacteriol 158, 102–109. [PubMed: 6715279]

Kublanov IV, Sigalova OM, Gavrilov SN, Lebedinsky AV, Rinke C, Kovaleva O, Chernyh NA, Ivanova N, Daum C, Reddy TBK, et al. (2017). Genomic Analysis of *Caldithrix abyssi*, the Thermophilic Anaerobic Bacterium of the Novel Bacterial Phylum *Calditrichaeota*. Front. Microbiol 8, 195. [PubMed: 28265262]

Kurowski WM, and Darbyshire J (1978). The Production of 3-Ketosucrose by *Agrobacterium tumefaciens* in Batch Culture. J. Appl. Chem. Biotechnol 28, 638–640.

Legrain C, Demarez M, Glansdorff N, and Pierard A (1995). Ammonia-dependent synthesis and metabolic channelling of carbamoyl phosphate in the hyperthermophilic archaeon Pyrococcus furiosus. Microbiology 141, 1093–1099. [PubMed: 33820113]

Lin J, Cagliero C, Guo B, Barton Y-W, Maurel M-C, Payot S, and Zhang Q (2005). Bile salts modulate expression of the CmeABC multidrug efflux pump in Campylobacter jejuni. J. Bacteriol 187, 7417–7424. [PubMed: 16237025]

Liou CS, Sirk SJ, Diaz CAC, Klein AP, Fischer CR, Higginbottom SK, Erez A, Donia MS, Sonnenburg JL, and Sattely ES (2020). A metabolic pathway for activation of dietary glucosinolates by a human gut symbiont. Cell 180, 717–728.e19. [PubMed: 32084341]

Liu H, and Deutschbauer AM (2018). Rapidly moving new bacteria to model-organism status. Curr. Opin. Biotechnol 51, 116–122. [PubMed: 29316481]

Liu H, Price MN, Waters RJ, Ray J, Carlson HK, Lamson JS, Chakraborty R, Arkin AP, and Deutschbauer AM (2018). Magic pools: parallel assessment of transposon delivery vectors in bacteria. mSystems 3, e00143–17. [PubMed: 29359196]

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, and Henrissat A (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490–D495. [PubMed: 24270786]

Luis AS, Briggs J, Zhang X, Farnell B, Ndeh D, Labourel A, Baslé A, Cartmell A, Terrapon N, Stott K, et al. (2018). Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides. Nat. Microbiol 3, 210–219. [PubMed: 29255254]

Martens EC, Chiang HC, and Gordon JI (2008). Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe 4, 447–457. [PubMed: 18996345]

Martens EC, Koropatkin NM, Smith TJ, and Gordon JI (2009). Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. J. Biol. Chem 284, 24673–24677. [PubMed: 19553672]

Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW, Henrissat B, Gilbert HJ, Bolam DN, and Gordon JI (2011). Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol. 9, e1001221. [PubMed: 22205877]

Mishra S, and Imlay JA (2013). An anaerobic bacterium, Bacteroides thetaiotaomicron, uses a consortium of enzymes to scavenge hydrogen peroxide. Mol. Microbiol 90, 1356–1371. [PubMed: 24164536]

Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y, El-Gebali S, Fraser MI, et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 47, D351–D360. [PubMed: 30398656]

Miyamae S, Ueda O, Yoshimura F, Hwang J, Tanaka Y, and Nikaido H (2001).A MATE family multidrug efflux transporter pumps out fluoroquinolones in Bacteroides thetaiotaomicron. Antimicrob. Agents Chemother. 45, 3341–3346. [PubMed: 11709306]

Miyazaki R, Yamazaki T, Yoshimatsu K, Kojima K, Asano R, Sode K, and Tsugawa W (2018). Elucidation of the intra- and inter-molecular electron transfer pathways of glucoside 3-dehydrogenase. Bioelectrochemistry 122, 115–122. [PubMed: 29625423]

Mortensen PB (1992).The effect of oral-administered lactulose on colonic nitrogen metabolism and excretion. Hepatology 16, 1350–1356. [PubMed: 1446891]

Ndeh D, and Gilbert HJ (2018). Biochemistry of complex glycan depolymerisation by the human gut microbiota. FEMS Microbiol. Rev 42, 146–164. [PubMed: 29325042]

Ndeh D, Munoz Munoz J, Cartmell A, Bulmer D, Wills C, Henrissat B, and Gray J (2018). The human gut microbe *Bacteroides thetaiotaomicron* encodes the founding member of a novel glycosaminoglycan-degrading polysaccharide lyase family PL29. J. Biol. Chem 293, 17906–17916. [PubMed: 30262663]

Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, Chaba R, Lee S, Kazmierczak KM, Lee KJ, Wong A, et al. (2011). Phenotypic landscape of a bacterial cell. Cell 144, 143–156. [PubMed: 21185072]

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, et al. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. 33, 5691–5702. [PubMed: 16214803]

Paulus TJ, and Switzer RL (1979). Characterization of pyrimidine-repressible and arginine-repressible carbamyl phosphate synthetases from Bacillus subtilis. J. Bacteriol 137, 82–91. [PubMed: 216664]

Peters JM, Koo B-M, Patino R, Heussler GE, Hearne CC, Qu J, Inclan YF, Hawkins JS, Lu CHS, Silvis MR, et al. (2019). Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. Nat. Microbiol 4, 244–250. [PubMed: 30617347]

Peterson J, Garges S, Giovanni M, McInnes P, Wang L, Schloss JA, Bonazzi V, McEwen JE, Wetterstrand KA, Deal C, et al.; NIH HMP Working Group (2009). The NIH human microbiome project. Genome Res. 19, 2317–2323. [PubMed: 19819907]

Popa E, Perera N, Kibédi-Szabó CZ, Guy-Evans H, Evans DR, and Purcarea C (2012). The smallest active carbamoyl phosphate synthetase was identified in the human gut archaeon Methanobrevibacter smithii. J. Mol. Microbiol. Biotechnol 22, 287–299. [PubMed: 23107800]

Price MN, and Arkin AP (2017). PaperBLAST: Text Mining Papers for Information about Homologs. mSystems 2, e00039–17. [PubMed: 28845458]

Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, Kuehl JV, Melnyk RA, Lamson JS, Suh Y, et al. (2018). Mutant phenotypes for thousands of bacterial genes of unknown function. Nature 557, 503–509. [PubMed: 29769716]

Price MN, Ray J, Iavarone AT, Carlson HK, Ryan EM, Malmstrom RR, Arkin AP, and Deutschbauer AM (2019). Oxidative pathways of deoxyribose and deoxyribonate catabolism. mSystems 4, e00297–18.

Prouty AM, Brodsky IE, Falkow S, and Gunn JS (2004). Bile-salt-mediated induction of antimicrobial and bile resistance in Salmonella typhimurium. Microbiology (Reading) 150, 775–783. [PubMed: 15073288]

Pumbwe L, Skilbeck CA, Nakano V, Avila-Campos MJ, Piazza RMF, and Wexler HM (2007). Bile salts enhance bacterial co-aggregation, bacterial-intestinal epithelial cell adhesion, biofilm formation and antimicrobial resistance of Bacteroides fragilis. Microb. Pathog 43, 78–87. [PubMed: 17524609]

Raghavan V, and Groisman EA (2015). Species-specific dynamic responses of gut bacteria to a mammalian glycan. J. Bacteriol 197, 1538–1548. [PubMed: 25691527]

Raghavan V, Lowe EC, Townsend GE 2nd, Bolam DN, and Groisman EA (2014). Tuning transcription of nutrient utilization genes to catabolic rate promotes growth in a gut bacterium. Mol. Microbiol. 93, 1010–1025. [PubMed: 25041429]

Ravcheev DA, Godzik A, Osterman AL, and Rodionov DA (2013). Polysaccharides utilization in human gut bacterium Bacteroides thetaiotaomicron: comparative genomics reconstruction of metabolic and regulatory networks. BMC Genomics 14, 873. [PubMed: 24330590]

Reeves AR, D'Elia JN, Frias J, and Salyers AA (1996). A Bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch. J. Bacteriol 178, 823–830. [PubMed: 8550519]

Ridlon JM, Kang D-J, and Hylemon PB (2006). Bile salt biotransformations by human intestinal bacteria. J. Lipid Res 47, 241–259. [PubMed: 16299351]

Ritz C, Baty F, Streibig JC, and Gerhard D (2015). Dose-Response Analysis Using R. PLoS One 10, e0146021. [PubMed: 26717316]

Rogers TE, Pudlo NA, Koropatkin NM, Bell JSK, Moya Balasch M, Jasker K, and Martens EC (2013). Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. Mol. Microbiol 88, 876–890. [PubMed: 23646867]

Ruiz L, Margolles A, and Sánchez B (2013). Bile resistance mechanisms in Lactobacillus and Bifidobacterium. Front. Microbiol 4, 396. [PubMed: 24399996]

Saeed-Kothe A, and Powers-Lee SG (2003). Gain of glutaminase function in mutants of the ammonia-specific frog carbamoyl phosphate synthetase. J. Biol. Chem 278, 26722–26726. [PubMed: 12738780]

Salyers AA, and O'Brien M (1980). Cellular location of enzymes involved in chondroitin sulfate breakdown by Bacteroides thetaiotaomicron. J. Bacteriol 143, 772–780. [PubMed: 6782076]

Salyers AA, Vercellotti JR, West SE, and Wilkins TD (1977). Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. Appl. Environ. Microbiol 33, 319–322. [PubMed: 848954]

Schuerman PL, Liu JS, Mou H, and Dandekar AM (1997). 3-Ketoglycoside-mediated metabolism of sucrose in E. coli as conferred by genes from Agrobacterium tumefaciens. Appl. Microbiol. Biotechnol 47, 560–565. [PubMed: 9210346]

Shi D, Morizono H, Cabrera-Luque J, Yu X, Roth L, Malamy MH, Allewell NM, and Tuchman M (2006). Structure and catalytic mechanism of a novel N-succinyl-L-ornithine transcarbamylase in arginine biosynthesis of Bacteroides fragilis. J. Biol. Chem 281, 20623–20631. [PubMed: 16704984]

Shiver AL, Culver R, Deutschbauer AM, and Huang KC (2019). Rapid ordering of barcoded transposon insertion libraries of anaerobic bacteria. bioRxiv. 10.1101/780593.

Shoemaker NB, Vlamakis H, Hayes K, and Salyers AA (2001). Evidence for extensive resistance gene transfer among Bacteroides spp. and among Bacteroides and other genera in the human colon. Appl. Environ. Microbiol 67, 561–568. [PubMed: 11157217]

Sinha R, Stanley G, Gulati GS, Ezran C, Travaglini KJ, Wei E, Chan CKF, Nabhan AN, Su T, Morganti RM, et al. (2017). Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. bioRxiv. 10.1101/125724.

Smith EA, and Macfarlane GT (1996). Studies on Amine Production in the Human Colon: Enumeration of Amine forming Bacteria and Physiological Effects of Carbohydrate and pH. Anaerobe 2, 285–297.

Smith KA, and Salyers AA (1991). Characterization of a neopullulanase and an alpha-glucosidase from Bacteroides thetaiotaomicron 95-1. J. Bacteriol 173, 2962–2968. [PubMed: 1708385]

Sonnenburg JL, Xu J, Leip DD, Chen C-H, Westover BP, Weatherford J, Buhler JD, and Gordon JI (2005). Glycan foraging in vivo by an intestine-adapted bacterial symbiont. Science 307, 1955–1959. [PubMed: 15790854]

Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, and Sonnenburg JL (2010). Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. Cell 141, 1241–1252. [PubMed: 20603004]

Sonntag K, Eggeling L, De Graaf AA, and Sahm H (1993). Flux partitioning in the split pathway of lysine synthesis in Corynebacterium glutamicum. Quantification by 13C- and 1H-NMR spectroscopy. Eur. J. Biochem 213, 1325–1331. [PubMed: 8504824]

Stentz R, Horn N, Cross K, Salt L, Brearley C, Livermore DM, and Carding SR (2015). Cephalosporinases associated with outer membrane vesicles released by Bacteroides spp. protect gut pathogens and commensals against β-lactam antibiotics. J. Antimicrob. Chemother 70, 701–709. [PubMed: 25433011]

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, et al. (2003). The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4, 41. [PubMed: 12969510]

Terrapon N, Lombard V, Drula É, Lapébie P, Al-Masaudi S, Gilbert HJ, and Henrissat B (2018). PULDB: the expanded database of Polysaccharide Utilization Loci. Nucleic Acids Res. 46 (D1), D677–D683. [PubMed: 29088389]

Thibault D, Jensen PA, Wood S, Qabar C, Clark S, Shainheit MG, Isberg RR, and van Opijnen T (2019). Droplet Tn-Seq combines microfluidics with Tn-Seq for identifying complex single-cell phenotypes. Nat. Commun 10, 5729. [PubMed: 31844066]

Trotta PP, Burt ME, Haschemeyer RH, and Meister A (1971). Reversible dissociation of carbamyl phosphate synthetase into a regulated synthesis subunit and a subunit required for glutamine utilization. Proc. Natl. Acad. Sci. USA 68, 2599–2603. [PubMed: 4944634]

Tuson HH, Foley MH, Koropatkin NM, and Biteen JS (2018).The Starch Utilization System Assembles around Stationary Starch-Binding Proteins. Biophys. J 115, 242–250. [PubMed: 29338841]

van Opijnen T, and Camilli A (2013). Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nat. Rev. Microbiol 11, 435–442. [PubMed: 23712350]

Varel VH, and Bryant MP (1974). Nutritional features of Bacteroides fragilis subsp. fragilis. Appl. Microbiol 28, 251–257. [PubMed: 4853401]

Webber MA, and Piddock LJV (2003). The importance of efflux pumps in bacterial antibiotic resistance. J. Antimicrob. Chemother 51, 9–11. [PubMed: 12493781]

Wehrmann A, Phillipp B, Sahm H, and Eggeling L (1998). Different modes of diaminopimelate synthesis and their role in cell wall integrity: a study with Corynebacterium glutamicum. J. Bacteriol 180, 3159–3165. [PubMed: 9620966]

Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, and Deutschbauer A (2015). Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. mBio 6, e00306–15. [PubMed: 25968644]

Wexler AG, and Goodman AL (2017). An insider's perspective: Bacteroides as a window into the microbiome. Nat. Microbiol 2, 17026. [PubMed: 28440278]

Willför S, Sjöholm R, Laine C, and Holmbom B (2002). Structural features of water-soluble arabinogalactans from Norway spruce and Scots pine heartwood. Wood Sci. Technol 36, 101–110.

Wrong OM, Vince AJ, and Waterlow JC (1985). The contribution of endogenous urea to faecal ammonia in man, determined by 15N labelling of plasma urea. Clin. Sci. (Lond.) 68, 193–199. [PubMed: 3967466]

Wu M, McNulty NP, Rodionov DA, Khoroshkin MS, Griffin NW, Cheng J, Latreille P, Kerstetter RA, Terrapon N, Henrissat B, et al. (2015). Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut Bacteroides. Science 350, aac5992. [PubMed: 26430127]

Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, and Gordon JI (2003). A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. Science 299, 2074–2076. [PubMed: 12663928]

Yao L, Seaton SC, Ndousse-Fetter S, Adhikari AA, DiBenedetto N, Mina AI, Banks AS, Bry L, and Devlin AS (2018).A selective gut bacterial bile salt hydrolase alters host metabolism. eLife 7, e37182. [PubMed: 30014852]

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, and Brinkman FS (2010). PSORTb3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26, 1608–1615. [PubMed: 20472543]

Zitvogel L, Daillère R, Roberti MP, Routy B, and Kroemer G (2017). Anticancer effects of the microbiome and its products. Nat. Rev. Microbiol 15, 465–478. [PubMed: 28529325]

Highlights

- Identification of specific phenotypes for 516 *Bacteroides thetaiotaomicron* genes
- A 3-keto-glucoside hydrolase is important for disaccharide utilization
- A tripartite multidrug resistance system is important for bile salt resistance
- Use of alternate biosynthetic enzymes depends on ammonium availability in the gut

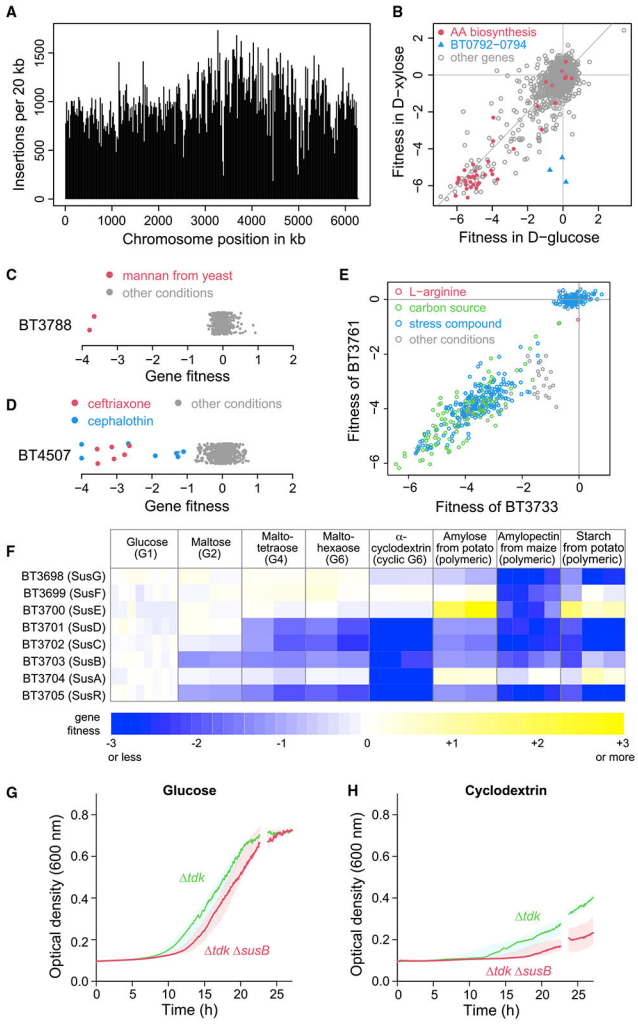


Figure 1. High-throughput genetics in *B. thetaiotaomicron*

(A) Transposon insertion distribution across the chromosome.

(B) Gene fitness values on glucose or xylose as the sole carbon source. The amino acid biosynthesis genes are predicted by TIGR (The Institute for Genomic Research) roles (Peterson et al., 2009).

(C and D) Gene fitness values across all conditions for BT3788 and BT4507. The y axis is random, and fitness values of <−4 are shown at −4. Conditions in which these genes had a specific phenotype are marked.

(E) Comparison of gene fitness values for BT3733 and BT3761 across all experiments. All assays marked “carbon source” were performed in a defined medium minus arginine; the experiment marked “L-arginine” was performed in defined glucose medium supplemented with 0.5 mM L-arginine.

(F) Heatmap of gene fitness values for starch utilization system genes on various glucose-containing carbon sources is shown. Data from each replicate experiment are shown separately. Compounds are ordered from left to right by increasing complexity.

(G and H) Growth curves of Δ*tdk* (control, 15 replicates per condition) and Δ*susB* (18 replicates per condition) in defined medium with glucose (G) or α-cyclodextrin (H). For

each strain and condition, curves are medians, and shaded regions represent the 10^{th}-90^{th} percentile.

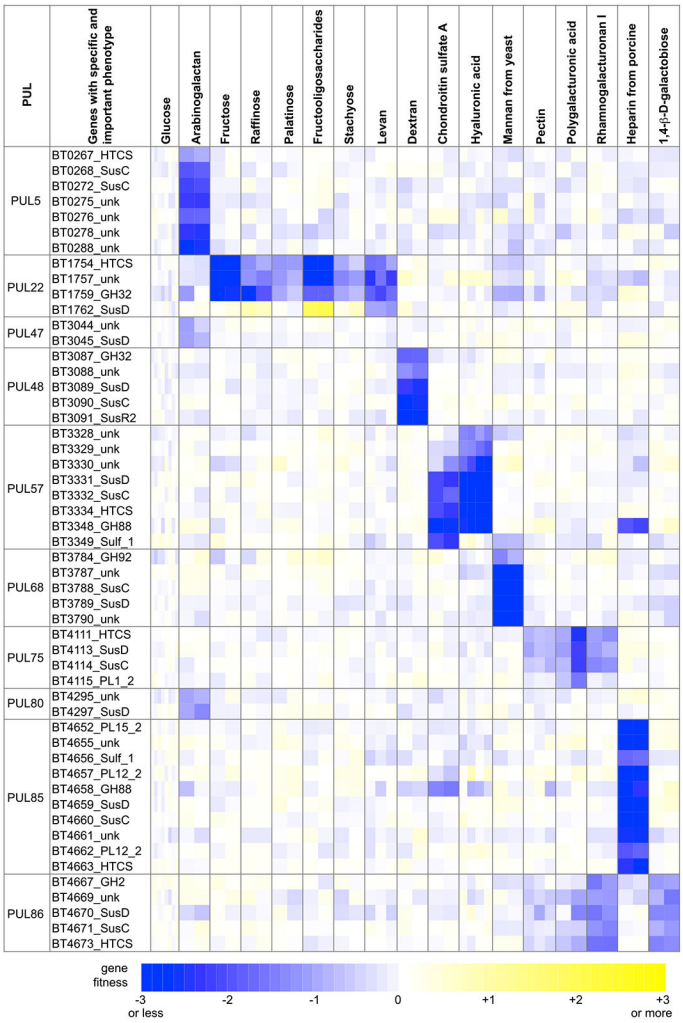


Figure 2. Fitness data associate PULs with substrates

Heatmap of fitness data for genes with important (fitness, <−1) and specific phenotypes on various carbon sources. Data for each replicate experiment are presented separately. HTCS: AraC-like hybrid two-component system; unk: unknown; GH: glycoside hydrolase; PL: polysaccharide lyase. Labels indicate gene families, and annotations are derived from PULDB (The Polysaccharide Utilization Loci database) (Terrapon et al., 2018).

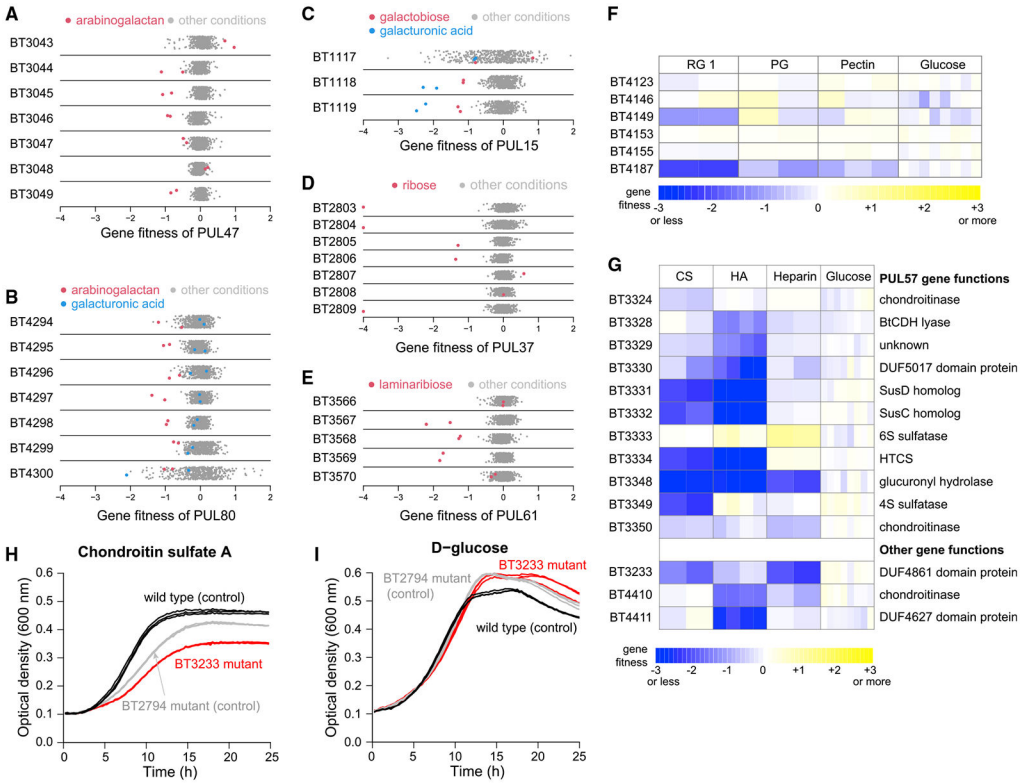


Figure 3. Genetic evidence linking genes to substrate catabolism

(A–E) Fitness data for select PULs. For the indicated PUL, fitness data for all component genes with certain carbon source conditions are highlighted. The y axis is random, and values of <−4 are shown at −4.

(F) Fitness data for six GH28-family glycosyl hydrolases on type I rhamnogalacturonan (RG-I), polygalacturonic acid (PG), pectin, and glucose. Each replicate experiment is shown separately.

(G) Fitness data on chondroitin sulfate (CS), hyaluronic acid (HA), heparin, and glucose.

(H and I) Growth curves of wild type, a BT3233 transposon mutant, and a BT2794 transposon mutant in CS (H) and glucose (I). The BT2794 mutant was used as a control, as this gene does not have a phenotype during growth on CS, unlike BT3233. The reduced growth of the BT2794 mutant control relative to that of the wild type may reflect the burden of expressing the erythromycin resistance gene. Three replicate growth curves per strain/condition are displayed.

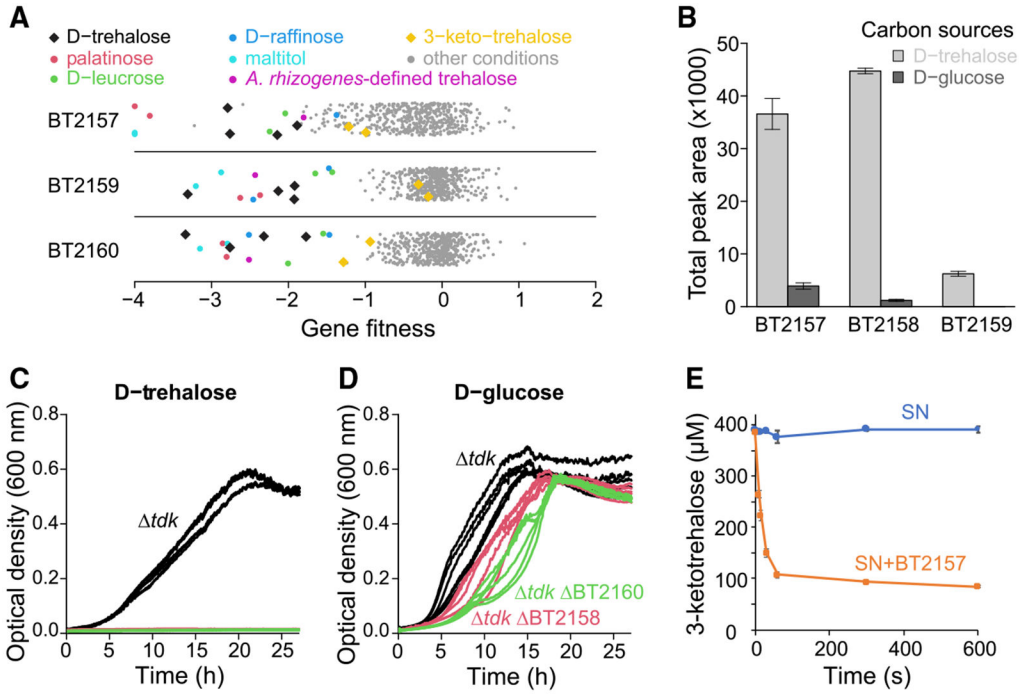


Figure 4. Disaccharide catabolism through a putative 3-keto-glycoside intermediate

(A) Gene fitness values for BT2157, BT2159, and BT2160 under all tested conditions. The y axis is random, and values of <−4 are shown at −4. *A. rhizogenes*-defined trehalose is the base medium used to produce 3-keto-trehalose using *A. rhizogenes*. The final 3-keto-trehalose-containing supernatant contained 7.9 mM 3-keto-trehalose.

(B) Relative quantification of the levels of proteins (BT2157–BT2159) predicted to be regulated by BT2160 after growth on the indicated carbon sources. Data are averages of 4 replicates. Error bars represent 1 standard deviation.

(C and D) Growth curves of deletion mutants (*tdk*, BT2158, and BT2160) with trehalose (C) or glucose as the sole carbon source. Δ*tdk* was used as a control. Six replicate growth curves per strain/condition are displayed, except for Δ*tdk* ΔBT2158/D-glucose (5 replicates).

(E) Enzymatic assay of purified BT2157 on 3-keto-trehalose supernatant (SN). The total 3-keto-trehalose in the reaction system was measured in a 10-min window, with or without the presence of purified BT2157 protein. Data are averages of 3 replicate experiments, and error bars represent 1 standard deviation.

See also Figures S1 and S2.

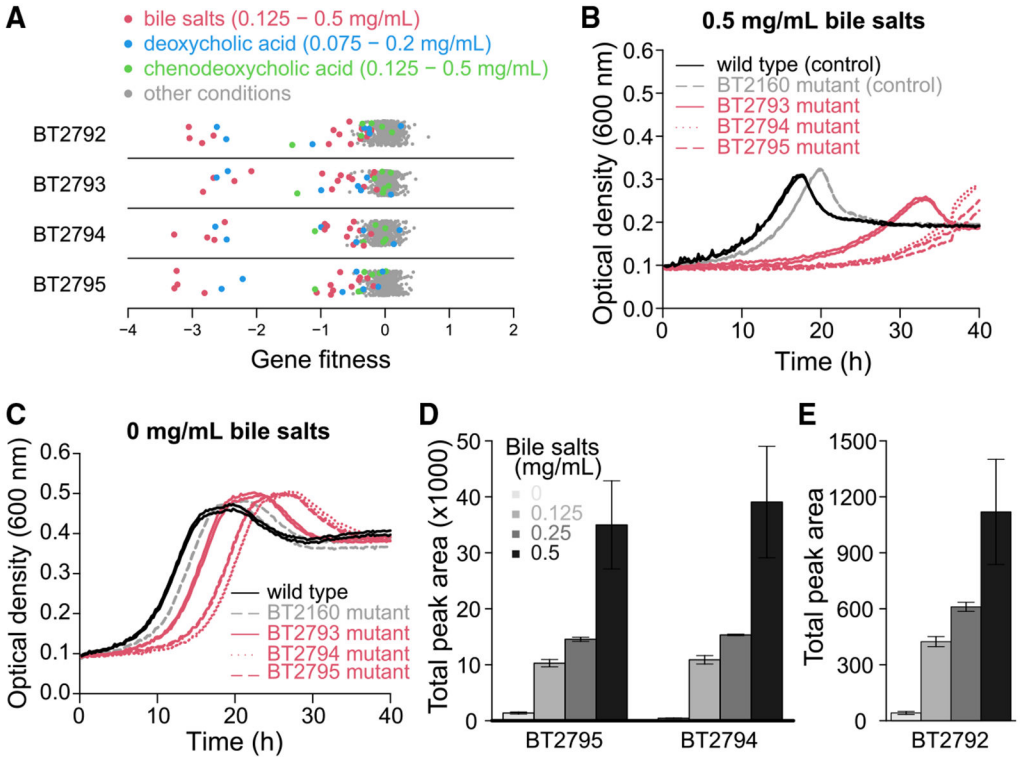


Figure 5. Identification of an efflux system for bile salts

(A) Gene fitness values for BT2792–BT2795 across all experiments, with bile salts and bile acid conditions highlighted. The highlighted experiments without a significant phenotype correspond to lower concentrations of the compound. The y axis is random.

(B and C) Growth curves of individual transposon mutant strains in VB medium with 0.5 mg/mL bile salts (B) or no added stress (C). For each strain/growth condition, two replicate growth curves are shown.

(D and E) Relative abundance of proteins BT2794 and BT2795 (D) or BT2792 (E) from wild-type cells grown in BHIS medium with various concentrations of bile salts. Data are averages of four replicate experiments, and error bars represent 1 standard deviation.

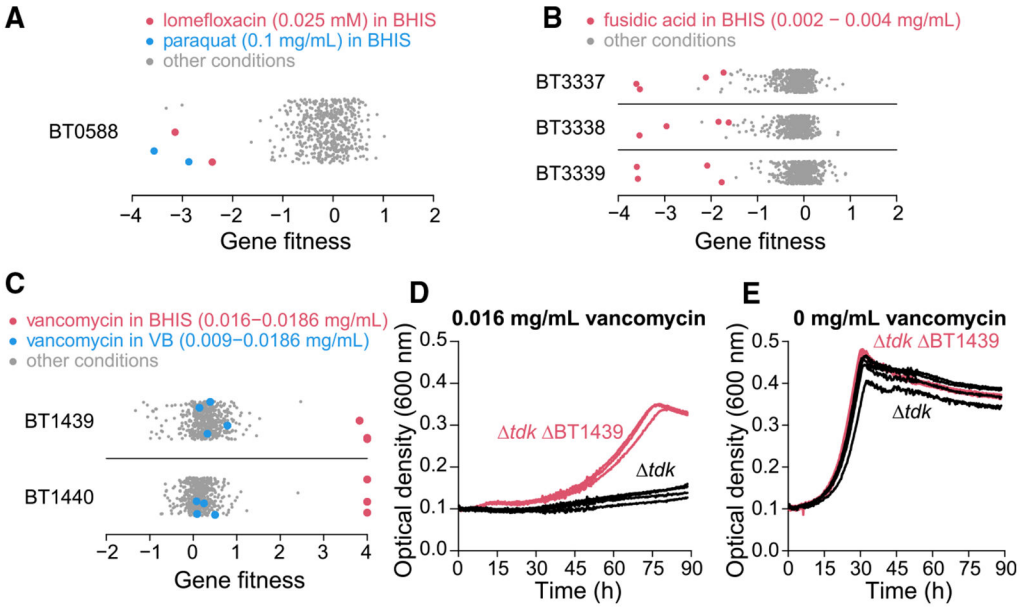


Figure 6. Identification of antibiotic resistance and sensitivity genes

(A–C) Gene fitness values for BT0588 (BexA) (A), BT3337–BT3339 (B), and BT1439–BT1440 (C) under all tested conditions. The y axis is random, and fitness values of > 4 are shown at 4.

(D and E) Growth curves of ΔBT1439 and the parental strain (Δ*tdk*) in the presence (D) or absence (E) of vancomycin in BHIS medium. Three replicate growth curves per strain/condition are shown.

See also Figure S3.

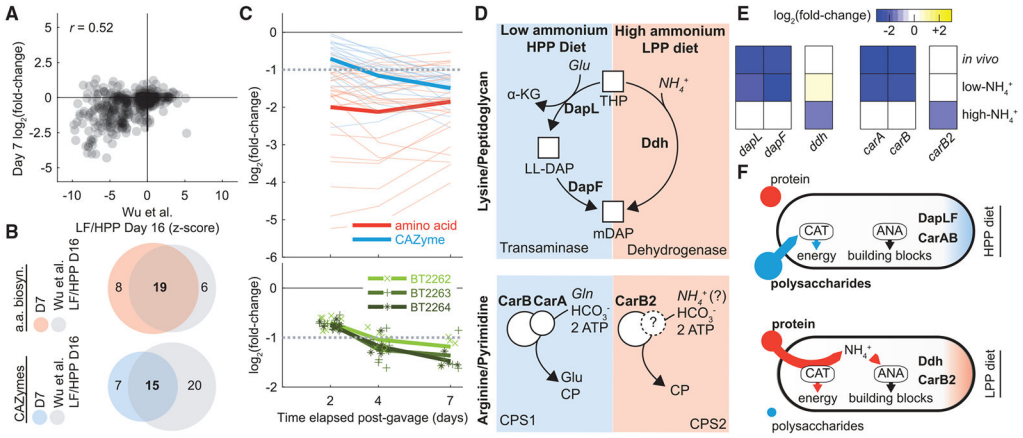


Figure 7. *B. thetaiotaomicron* uses alternative enzymes in amino acid biosynthesis to adapt to ammonium availability *in vitro* and dietary polysaccharide content *in vivo*

Fitness of the *B. thetaiotaomicron* VPI-5482 barcoded library was measured in gnotobiotic mice fed a standard diet.

(A) Comparison to a previous study. Data from Wu et al. (2015) on day 16 post-gavage in a low-fat, high-polysaccharide (LF/HPP) diet are plotted on the x axis (reported as *Z* scores). Gene fitness scores from our study on day 7 post-gavage are plotted on the y axis.

(B) Overlap between our dataset and the LF/HPP diet of Wu et al. (2015) for genes with fitness defects that are associated with amino acid biosynthesis (aa biosyn.; red) and carbohydrate-active enzymes (CAZymes; blue).

(C, top) Fitness data for amino acid biosynthesis and carbohydrate-active enzymes genes are plotted as a function of time post-gavage (thick lines). Values were averaged across mice (n = 4–8). Only genes with a $\log_2$ fold change of ≥−1 on at least one of the sampling days were included. The fitness of each gene is plotted as a thin line. Gray dotted line ($\log_2$ fold change = −1) is shown for comparison. (C, bottom) Fitness of genes in PUL27 (BT2262–BT2264) as a function of time post-gavage. The average across mice for each gene is plotted as a solid line. Individual measurements are plotted with markers; “x,” BT2262; “+,” BT2263; “*,” BT2264.

(D) Diagram of alternative enzymes and pathways used in the synthesis of meso-diaminopimelate (top) and carbamoyl phosphate (CP; bottom). The reactions shown are simplified for clarity and do not list all reactants and products. Reactants that serve as a nitrogen source are italicized. HPP, high plant polysaccharide; LPP, low plant polysaccharide; THP, 2,3,4,5-tetrahydropicolinate; LL-DAP, L,L-diaminopimelate; mDAP, meso-diaminopimelate; Glu, glutamate; α-KG, α-ketoglutarate; Gln, glutamine.

(E) Heatmap of fitness values for the alternative enzymes in (D). The *in vivo* condition is an average of fitness values across all mice and all days. The *in vitro* low-NH_4^+ condition is an average across multiple (n = 8) low-ammonium *in vitro* assays, in which the ammonium component of VB medium was replaced with a single amino acid. The *in vitro* high-NH_4^+ condition is an average across multiple assays (n = 10) with the canonical VB formulation (NH_4^+ as the nitrogen source). The carbon source for all *in vitro* assays was glucose.

(F) Model for the *in vitro* and *in vivo* phenotypes of *dapLF*, *ddh*, *carAB*, and *carB2*. Flux of amino acids (red) and polysaccharides (blue) into *B. thetaiotaomicron* metabolism is shown with colored lines. Availability of nutrients (amino acids and polysaccharides) is represented

by the size of the circle. CAT, catabolic pathways that produce energy; ANA, anabolic pathways that produce building blocks.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial strains		
Escherichia coli strain TransforMax EC100D pir+	https://www.lucigen.com/	ECP09500
Escherichia coli strain 10-beta	https://www.neb.com	C3019H
Escherichia coli strain WM3064	William Metcalf lab, University of Illinois	N/A
Escherichia coli strain BL21 (DE3)	https://www.thermofisher.com/us/en/home.html	EC0114
Bacteroides thetaiotaomicron VPI 5482	https://www.atcc.org:443/	ATCC 29148
B. thetaiotaomicron VPI 5482 Δ*tdk*	Mishra and Imlay, 2013	N/A
Agrobacterium rhizogenes strain K599	Jay Keasling lab, University of California at Berkeley	N/A
B. thetaiotaomicron VPI 5482 *Tn*:BT2160(*ermG*)	This study	N/A
B. thetaiotaomicron VPI 5482 *Tn*:BT2793(*ermG*)	This study	N/A
B. thetaiotaomicron VPI 5482 *Tn*:BT2794(*ermG*)	This study	N/A
B. thetaiotaomicron VPI 5482 *Tn*:BT2795(*ermG*)	This study	N/A
B. thetaiotaomicron VPI 5482 *Tn*:BT3233(*ermG*)	This study	N/A
B. thetaiotaomicron VPI 5482 Δ*tdk* ΔBT1439	This study	N/A
B. thetaiotaomicron VPI 5482 Δ*tdk* ΔBT3703	This study	N/A
B. thetaiotaomicron VPI 5482 Δ*tdk* ΔBT4448	This study	N/A
B. thetaiotaomicron VPI 5482 Δ*tdk* ΔBT2158	This study	N/A
B. thetaiotaomicron VPI 5482 Δ*tdk* ΔBT2160	This study	N/A
Biological samples		
Germ-free (GF) mice	UCSF Gnotobiotic Core Facility (https://gnotobiotics.ucsf.edu)	N/A
Chemicals		
See Table S7 for list and information	N/A	N/A
Critical commercial assays		
QIAamp 96 DNA QIAcube HT kit	QIAGEN	Cat. No. 51331
RNeasy Mini Kit	QIAGEN	Cat. No. 74104
iTaq Universal SYBR Green One-Step Kit	Bio-Rad	Cat. No. 1725151

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ZymoBIOMICS 96 MagBead DNA Kit	ZYMO RESEARCH	Cat. No. D4302
Trehalose Assay Kit	https://www.megazyme.com/	Cat. No. K-TREH
Deposited data		
Fitness browser	Price et al., 2018	https://fit.genomics.lbl.gov
Fitness browser archive	This study	https://figshare.com/articles/dataset/The_Fitness_Browser_Genome-wide_mutant_fitness_data_for_diverse_bacteria_November_2020_release_/13172087/1
Fitness data for *B. thetaiotaomicron*	This study	https://figshare.com/articles/dataset/RB-TnSeq_data_for_Bacteroides_thetaiotaomicron_VPI-5482/13171766/1
Oligonucleotides		
See Tables S6 and S8 for list and information	N/A	N/A
Recombinant DNA		
Plasmid: pTGG45	This study	N/A
Plasmid: pTGG46	This study	N/A
Plasmid: pTGG46_NN1	This study	N/A
Plasmid: pHLL213	Liu et al., 2018	N/A
Plasmid: pHLL264	This study	N/A
Plasmid: pJW66	Liu et al., 2018	N/A
Plasmid: pJW21	Liu et al., 2018	N/A
Plasmid: pHLL215	Liu et al., 2018	N/A
Plasmid: pJW43	Liu et al., 2018	N/A
Plasmid: pHL280	This study	N/A
Plasmid: pExchange-tdk	Mishra and Imlay, 2013	N/A
Plasmid: pExchange-tdk-ΔBT1439	This study	N/A
Plasmid: pExchange-tdk-ΔBT3703	This study	N/A
Plasmid: pExchange-tdk-ΔBT4448	This study	N/A
Plasmid: pExchange-tdk-ΔBT2158	This study	N/A
Plasmid: pExchange-tdk-ΔBT2160	This study	N/A
pKH-BT2157	This study	N/A
Software and algorithms		
Software for RB-TnSeq	Wetmore et al., 2015	https://bitbucket.org/berkeleylab/feba
Software for proteomic analysis	https://skyline.ms/	Skyline v. 3.70