Journal of Bacteriology

| Editor's Pick | Bacteriology | Full-Length Text

Determinants of raffinose family oligosaccharide use in *Bacteroides* species

Anubhav Basu,[1] Amanda N. D. Adams,[1,2] Patrick H. Degnan,[3] Carin K. Vanderpool[1,4]

AUTHOR AFFILIATIONS See affiliation list on p. 20.

ABSTRACT *Bacteroides* species are successful colonizers of the human colon and can utilize a wide variety of complex polysaccharides and oligosaccharides that are indigestible by the host. To do this, they use enzymes encoded in polysaccharide utilization loci (PULs). While recent work has uncovered the PULs required for the use of some polysaccharides, how *Bacteroides* utilize smaller oligosaccharides is less well studied. Raffinose family oligosaccharides (RFOs) are abundant in plants, especially legumes, and consist of variable units of galactose linked by α-1,6 bonds to a sucrose (glucose α-1-β-2 fructose) moiety. Previous work showed that an α-galactosidase, BT1871, is required for RFO utilization in *Bacteroides thetaiotaomicron*. Here, we identify two different types of mutations that increase *BT1871* mRNA levels and improve *B. thetaiotaomicron* growth on RFOs. First, a novel spontaneous duplication of *BT1872* and *BT1871* places these genes under the control of a ribosomal promoter, driving high *BT1871* transcription. Second, nonsense mutations in a gene encoding the PUL24 anti-sigma factor likewise increase *BT1871* transcription. We then show that hydrolases from PUL22 work together with BT1871 to break down the sucrose moiety of RFOs and determine that the master regulator of carbohydrate utilization (BT4338) plays a role in RFO utilization in *B. thetaiotaomicron*. Examining the genomes of other *Bacteroides* species, we found homologs of BT1871 in a subset and showed that representative strains of species with a BT1871 homolog grew better on melibiose than species that lack a BT1871 homolog. Altogether, our findings shed light on how an important gut commensal utilizes an abundant dietary oligosaccharide.

IMPORTANCE The gut microbiome is important in health and disease. The diverse and densely populated environment of the gut makes competition for resources fierce. Hence, it is important to study the strategies employed by microbes for resource usage. Raffinose family oligosaccharides are abundant in plants and are a major source of nutrition for the microbiota in the colon since they remain undigested by the host. Here, we study how the model commensal organism, *Bacteroides thetaiotaomicron* utilizes raffinose family oligosaccharides. This work highlights how an important member of the microbiota uses an abundant dietary resource.

KEYWORDS raffinose family oligosaccharide, melibiose, *Bacteroides*, α-galactosidase, *BT1871*, polysaccharide utilization loci

The gut microbiome plays an important role in human health and development by facilitating energy extraction from food (1), synthesis of vitamins (2, 3), protection from pathogen colonization (4, 5), enhancing the immune system (6, 7), and modulating gut-brain communication (8). Although diverse, the human microbiome inhabiting the colon is dominated by two major phyla, namely the *Bacillota* and *Bacteroidota* (9, 10). Members of the *Bacteroidota* owe their success in this niche to their ability to break down and use complex host and plant-derived polysaccharides. They accomplish this

Editor George O'Toole, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA

Address correspondence to Carin K. Vanderpool, cvanderp@illinois.edu.

The authors declare no conflict of interest.

See the funding table on p. 20.

Received 3 June 2024
Accepted 6 September 2024
Published 27 September 2024

using paralogous gene clusters known as polysaccharide utilization loci (PULs) that encode glycan sensing, transport, and degrading enzymes (11–15). The prototypical PUL is characterized by the presence of homologs of SusC (an outer membrane oligosaccharide transporter) and SusD (an outer membrane oligosaccharide-binding protein) found in the starch utilization system (Sus) (16–18). They also contain other membrane-bound and periplasmic proteins required to bind, break down, and transport complex carbohydrates (19, 20).

Each PUL is specialized for degrading distinct polysaccharides and hence PUL gene expression is tightly regulated. Hybrid two-component systems (HTCS) and extracytoplasmic function (ECF) sigma factor/anti-sigma pairs are the most common regulators of PULs in *Bacteroides* (21, 22). HTCS combine the sensor kinase and response regulator proteins of a classical two-component system into a single polypeptide spanning the inner membrane. The recognition domain of the HTCS senses a unique oligosaccharide signal usually 2–8 subunits in length which causes sensor domain autophosphorylation and transfer of the phosphate to the regulator domain that promotes transcription of target genes (21, 23–28). PULs regulated by ECF sigma/anti-sigma pairs are usually associated with the breakdown of host-derived polysaccharides (29). Transport of oligosaccharides through the SusC-like protein is coupled to conformational changes in the anti-sigma, resulting in the release of the cognate sigma factor from the inner membrane and transcription of target genes (14, 30).

Bacteroides thetaiotaomicron, a model gut microbe, has over 100 PULs and dedicates ~18% of its genome to encoding functions for carbohydrate sensing and usage (31, 32). In recent years, many studies have probed the genetic and mechanistic details of how long-chain polysaccharides are broken down and utilized by *B. thetaiotaomicron* (27–29, 33–36). However, less work has focused on how *B. thetaiotaomicron* or other *Bacteroides* species can use smaller oligosaccharides. One such group of oligosaccharides is the raffinose family oligosaccharides (RFOs). RFOs are soluble carbohydrates and are functionally α-D-galactosyl derivatives of sucrose, a disaccharide of glucose and fructose [Fig. S1 and (31)]. Raffinose is the simplest RFO and is a trisaccharide where a single galactose is α-1,6 linked to the glucose moiety of sucrose. Longer RFOs such as stachyose and verbascose contain two and three galactose residues, respectively (Fig. S1). RFOs are highly abundant in the seeds of many crops, particularly in members of the legume family such as soybean, lentils, and chickpea (37–39). In plants, RFOs function in storage and translocation (40–42), stress tolerance (43–45), seed germination, and desiccation tolerance (46–49). RFOs are indigestible by humans since we lack the α-galactosidases required to break them down (50). Hence, RFOs in the diet pass undigested to the colon where they are utilized by a variety of microbes (51–53).

Recent studies have revealed the importance of RFOs in human health as prebiotics because they can modulate the abundance of beneficial bacteria in the gut (51, 52, 54). Because of their ability to influence the microbiota, it is important to elucidate the mechanism by which gut bacteria can utilize RFOs. Various *Bifidobacterium* species, *Bacillus subtilis,* and *Streptococcus pneumoniae* contain dedicated operons for sensing, transporting, and breaking down RFOs (53, 55–57). Other microbes such as *Escherichia coli and Enterococcus faecium* harbor these functions on plasmids (58, 59). However, homologs of these systems are absent in *Bacteroides* species. Previous work from our lab found that an α-galactosidase encoded by *BT1871* in PUL24 of *B. thetaiotaomicron* is important for RFO utilization *in vitro* (60). Deletion of *BT1871* led to decreased growth on RFOs as the sole carbon source.

In this study, we show that the efficiency of RFO use in *B. thetaiotaomicron* is limited by low levels of expression of *BT1871*, encoding an α-galactosidase that breaks the α-1,6 bond between glucose and galactose in RFOs. We found two different types of mutations that promote higher *BT1871* transcription to increase growth on RFOs. First, we serendipitously identified *B. thetaiotaomicron* strains with a novel duplication involving *BT1871* that leads to substantially better growth on RFOs compared to strains with only a single copy of *BT1871*. Second, we demonstrated that disruption of *BT1876,*

encoding the PUL24 anti-sigma factor, also increases *BT1871* transcript levels and leads to better RFO utilization. We then established that full RFO degradation by *B. thetaiotaomicron* requires the PUL24 α-galactosidase BT1871 as well as sucrases encoded by genes in PUL22. Investigation of regulatory mechanisms controlling RFO use revealed that the master regulator of carbohydrate utilization, BT4338, is required for growth on RFOs through control of the expression of *BT1871*. Finally, we show that *BT1871* homologs in other *Bacteroides* species are also important for their ability to use the disaccharide melibiose. Taken together, our findings reveal key players that help *B. thetaiotaomicron* and other *Bacteroides* species utilize RFOs.

RESULTS

A novel duplication of *BT1871* confers a growth advantage to *B. thetaiotaomicron* growing on RFOs

Previously, we found that deletion of *rbpB* (*BT1887*), coding for the RNA binding protein RbpB, caused a growth defect in *B. thetaiotaomicron* growing on RFOs as the sole carbon source (60). This phenotype was associated with reduced levels of the *BT1871-BT1872* mRNA in the *rbpB* mutant strain compared to the parent strain, though the regulatory mechanism responsible for this association was not determined. *BT1871* is an α-galactosidase capable of breaking down RFOs and its deletion also caused growth defects on RFOs (60).

During the construction of *rbpB* mutations in different backgrounds, we noticed that some mutant isolates had no growth defects on RFOs while other mutants showed growth defects similar to the original *rbpB* mutant. To better understand this phenotypic variability, we performed long- and short-read whole-genome sequencing of the two types of mutant isolates and the parent strain. We found that the parent strain had a duplication of the *BT1871-BT1872* locus which is not present in the NCBI reference genome (NC_004663.1). This duplication involves the promoter and 5′ end of a 16S rRNA gene, an insertion sequence (IS3) transposable element (*BT1869-BT1870*) and *BT1871* and *BT1872* (Fig. 1A). The duplication places *BT1871* and *BT1872* downstream of a ribosomal promoter. Once we had defined the structure of the duplication, we were able to isolate wild-type strain derivatives that had lost the duplication. Comparing transcript levels of *BT1871* in strains with (dupl+) or without (dupl−) the duplication, we found that dupl+ strains had substantially higher *BT1871* transcript levels (Fig. 1B, WT dupl+ and Δ*rbpB* dupl+) compared to the strains without the duplication (Fig. 1B, WT dupl− and Δ*rbpB* dupl−). Importantly, the phenotypes we observed previously, including growth on the RFO subunit melibiose, that we previously attributed to loss of *rbpB,* were instead caused by the loss of the duplication. This is evident when comparing the growth of the wild-type (dupl+) and two *rbpB* mutant isolates, one with (dupl+) and one without (dupl−) the duplication, on melibiose (Fig. 1C). Both wild-type dupl+ and Δ*rbpB* dupl+ strains grew well on melibiose (Fig. 1C), consistent with their high levels of expression of *BT1871* (Fig. 1B), which encodes the α-galactosidase that breaks the melibiose disaccharide bond (61, 62). By contrast, the wild-type dupl− and Δ*rbpB* dupl− strains grew significantly slower (mean growth rate of 0.66 h^{-1} for dupl+ strains compared to 0.35 h^{-1} for dupl− strains, $P < 0.01$ by two-tailed t-test) on melibiose (Fig. 1C) consistent with greatly reduced levels of *BT1871* expression observed in these strains (Fig. 1B). Hence, we conclude that the growth phenotype of *B. thetaiotaomicron* on RFOs is linked to the duplication status and expression of the *BT1871* gene and is not linked to the function of RbpB.

Mutations in the anti-sigma gene of PUL24 confer better RFO utilization in *B. thetaiotaomicron* through increased transcription of *BT1871*

To understand more about the role of *BT1871* in carbon source use and determine whether the duplication impacts the use of carbon sources beyond RFOs, we monitored the growth of wild-type dupl+ and dupl− cells on a variety of carbon sources. We

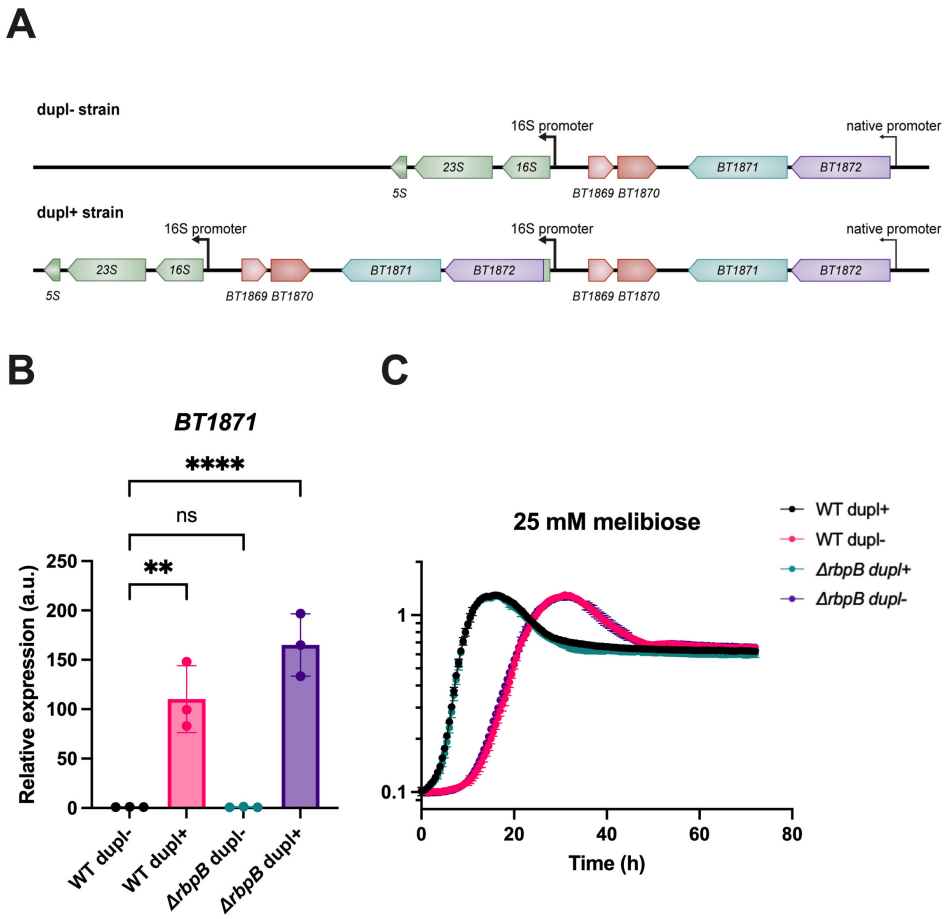

FIG 1 Duplication of the *BT1871-BT1872* locus provides *B. thetaiotaomicron* a growth advantage on melibiose. (A) Structure of the *BT1871-BT1872* (*BT1871*—PUL24 α-galactosidase, *BT1872*—PUL24 β-glucosidase) duplication in a dupl+ strain (bottom) compared to a strain without the duplication (dupl−, top). Red gene models represent genes in an IS3 element transposon while green gene models represent genes in a ribosomal RNA operon. (B) qRT-PCR to measure *BT1871* mRNA levels in the indicated strains grown to mid-log phase in rich (TYG) media. The bars represent the mean and SD of $n = 3$ biological replicates. All values are normalized to the level in WT dupl−. Statistical significance was determined by one-way ANOVA. $^{**}P < 0.01$; $^{****}P < 0.0001$. (C) Growth curves of the indicated strains on minimal media with melibiose as the sole carbon source. The points and error bars represent the mean and SD of $n = 3$ biological replicates.

identified the sugar α-methyl galactoside (AMG) as a carbon source that supported slow growth of dupl+ cells (mean growth rate of 0.10 h^{-1}) while dupl− cells could not grow even after 72 hours of incubation (Fig. 2A). Since AMG has the same galactose-α-1,6 bond found in melibiose and RFOs and *BT1871* is important for growth on RFOs, we hypothesized that the higher levels of BT1871 in the dupl+ strain were responsible for the observed phenotype on AMG.

To better understand mechanisms of regulation of *BT1871-1872* and determine whether duplications or other mutations conferring increased growth on AMG and RFOs could be readily generated *de novo*, we grew 10 replicates each of 8 independent colonies of dupl− cells in media with AMG as the sole carbon source. After about 45 hours, we observed growth in 12 wells (Fig. 2B). Upon transfer of the cultures to fresh media containing AMG, dupl− cells from the wells where growth had initially occurred were able to grow with decreased lag times and increased growth rates as compared to the parent dupl− strain, and most isolates grew faster on AMG than dupl+ cells (Fig. 2B, right panel). To confirm that these putative mutants have a stable, heritable phenotype, we saved six independent isolates and tested their growth on glucose, AMG, and melibiose. The isolates grew similarly to the parent WT dupl− strain on minimal media

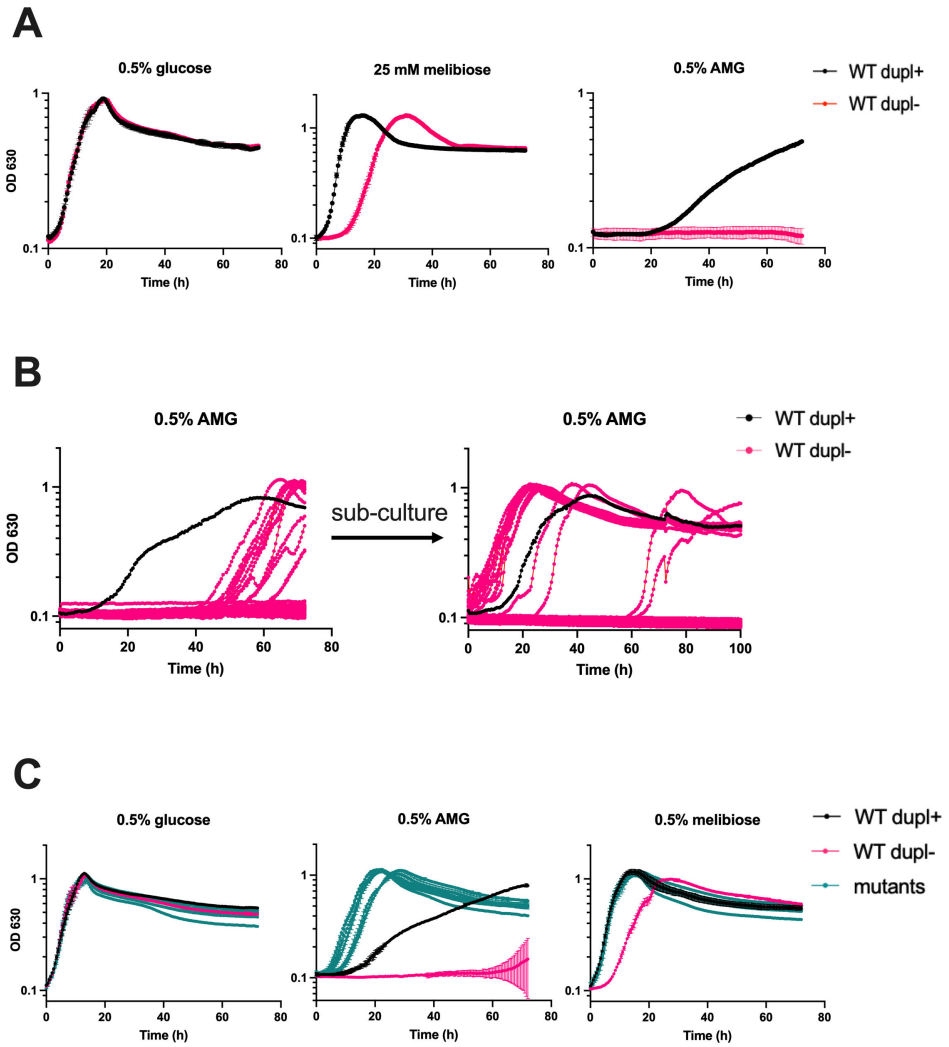

FIG 2 *B. thetaiotaomicron* mutants derived from WT dupl− cells grow on AMG. (A) Growth curves of WT dupl+ and WT dupl− strains on glucose, melibiose, and AMG. On glucose, the WT dupl− strain had a mean growth rate of 0.50 h^{-1} compared to 0.54 h^{-1} for the isolates (P = 0.49 by two-tailed t-test) (B) Growth curves of WT dupl− cells inoculated as described in the text in 96-well plates containing minimal media with AMG as the sole carbon source (left panel). Subculture of strains to fresh media shows that isolates that began growing between 40 and 60 hours show no lag (right panel). New isolates arise again in some wells. (C) Growth curves of WT dupl+, WT dupl−, and six independent mutants of WT dupl− that can grow on AMG as the sole carbon source. In (A) and (C), points and error bars represent the mean and SD of n = 3 biological replicates. For each growth curve, the sugar used as the sole carbon source is indicated at the top along with the concentration used.

with glucose and significantly better than the parent strain on minimal media with AMG (mean growth rate of 0.35 h^{-1} for the isolates while the WT dupl− strain failed to grow) and melibiose (mean growth rate of 0.28 h^{-1} for the WT dupl− strain compared to 0.53 h^{-1} for the isolates, $P < 0.01$ by two-tailed t-test) (Fig. 2C). Interestingly, the isolates also grew better than WT dupl+ on AMG (mean growth rate of 0.35 h^{-1} for the isolates compared to 0.07 h^{-1} for the WT dupl+ strain, $P < 0.01$ by two-tailed t-test). Thus, the isolates demonstrate phenotypes that may be caused by a stable genetic change.

To identify the putative mutation(s), we performed long- and short-read whole-genome sequencing on three independent isolates and found that each had a different nonsense mutation in gene *BT1876* (Table S1). *BT1876* encodes the anti-sigma factor paired with the sigma factor *BT1877* in PUL24 (Fig. 3A). To independently test whether inactivation of the BT1876 anti-sigma leads to better growth on RFOs, we made an in-frame deletion of *BT1876* in a WT dupl− background. The *BT1876* mutant strain grew

similar to the WT dupl− strain on glucose and significantly better than the WT dupl− strain on AMG (mean growth rate of 0.39 h^{-1} for the Δ*BT1876* strain while the WT dupl− strain failed to grow) and melibiose (mean growth rate of 0.29 h^{-1} for the WT dupl− strain compared to 0.55 h^{-1} for the Δ*BT1876* strain, $P < 0.01$ by two-tailed t-test) (Fig. 3B). The Δ*BT1876* strain grew slightly better than the WT dupl− strain on raffinose (mean growth rate of 0.17 h^{-1} for the WT dupl− strain compared to 0.23 h^{-1} for the Δ*BT1876* strain, $P < 0.01$ by two-tailed t-test). On stachyose, the Δ*BT1876* strain had a slightly slower initial growth rate than the WT dupl− strain (mean growth rate of 0.11 h^{-1} for the WT dupl− strain compared to 0.08 h^{-1} for the Δ*BT1876* strain, $P < 0.01$ by two-tailed t-test) but reached a higher maximal OD by the end of the experiment (mean maximal OD_{630} of 0.53 for the WT dupl− strain compared to 0.69 for the Δ*BT1876* strain) (Fig. 3B). Both raffinose and stachyose have sucrose moieties that are degraded by other enzymes and do not depend on expression levels of PUL24 genes including *BT1871* (60). Complementation of the *BT1876* mutant with a single copy of *BT1876* in *trans* under its native promoter restored growth on AMG and RFOs to the WT dupl− level. (Fig. 3B). Hence, loss of the anti-sigma BT1876 enhances the growth of *B. thetaiotaomicron* on RFOs.

We hypothesized that in the *BT1876* mutant strains, transcription of the α-galactosidase-encoding *BT1871* may be higher than in the WT dupl− strain due to constitutive activity of the sigma factor BT1877. To test this, we used RT-qPCR to measure *BT1871* mRNA levels in WT dupl−, the *BT1876* mutant, and complemented mutant strains grown on RFOs as the sole carbon source. We found that in the *BT1876* mutant, *BT1871* expression increased by ~20-fold on melibiose and ~1,450-fold on raffinose compared to the WT dupl− strain (Fig. 3C). The *BT1871* transcript levels in the complemented strain were similar to the levels in WT dupl− (Fig. 3C).

Two additional lines of evidence confirmed that increased expression of *BT1871* (α-gal) is responsible for better growth of the *BT1876* (anti-sigma) mutant strain on RFOs. First, a *BT1876 BT1871* double mutant strain grew similar to a *BT1871* single mutant strain on AMG, melibiose, and raffinose (Fig. S2A). This shows that activation of other PUL24 genes in the *BT1876* mutant strain does not confer better growth on RFOs if *BT1871* is absent. Second, a *BT1871* mutant strain is unable to grow on AMG as the sole carbon source and does not give rise to variants that acquire the ability to grow (Fig. S2B).

Collectively, these data suggest that there are at least two different mechanisms that provide *B. thetaiotaomicron* with increased fitness during growth on RFOs. One mechanism involves duplication of *BT1871* (α-gal) and the other is the mutation of *BT1876* (anti-sigma). Both mechanisms result in strongly increased transcription of *BT1871*, encoding an α-galactosidase that breaks the α-1,6 bond of RFOs (60–62). For the remainder of this study, we used a wild-type dupl− strain background to assess other determinants of RFO utilization. For simplicity, from now on we will refer to this strain background as "WT."

Other determinants of RFO utilization in *B. thetaiotaomicron*

To determine whether genes outside of PUL24 contribute to RFO utilization, we first tested whether any α-galactosidases other than BT1871 can cleave the α-1,6 bond in RFOs. A recent study showed that when gnotobiotic mice are colonized with a *B. thetaiotaomicron* transposon mutant library, many of the strains that take over the population have transposon insertions upstream of genes encoding α-galactosidases, including *BT1871* (64). The authors hypothesized that increased transcription of the downstream α-galactosidase genes due to readthrough from the transposon gave these mutants an advantage in mice fed a diet rich in raffinose and melibiose. *BT2851* and *BT3131* were two other genes encoding α-galactosidases for which upstream transposon insertions led to an *in vivo* fitness advantage. To test whether either of these genes contributes to *B. thetaiotaomicron* RFO use, we constructed single deletion mutations in the WT background and found that strains lacking either *BT2851* or *BT3131* had no growth defects on RFOs as compared to the WT strain (Fig. S3A). This suggests that the BT2851 and BT3131 α-galactosidases do not contribute meaningfully to RFO utilization.

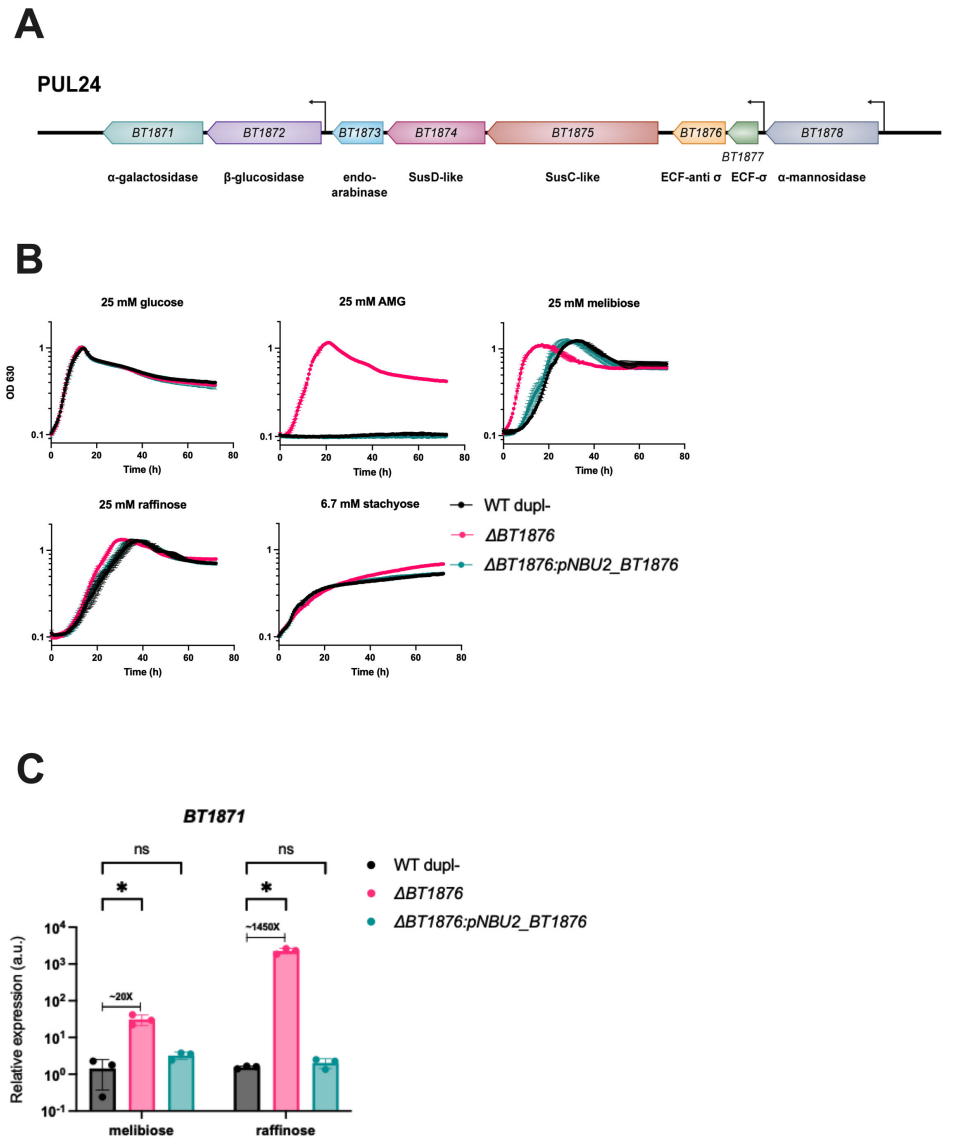

FIG 3 A *BT1876* mutant strain grows better on RFOs than the WT strain due to increased *BT1871* expression. (A) Genomic organization of *B. thetaiotaomicron* PUL24. The location of known transcription start sites [derived from reference (63) in the text] is indicated by bent arrows. (B) Growth curves of WT, Δ*BT1876* anti-sigma mutant, and a complemented strain. On glucose, the WT dupl− strain had a mean growth rate of 0.51 h^{-1} compared to 0.52 h^{-1} for the Δ*BT1876* strain (P = 0.75 by two-tailed t-test). On melibiose, the WT dupl− strain had a mean growth rate of 0.29 h^{-1} compared to 0.29 h^{-1} for the complemented strain (P = 0.60 by two-tailed t-test). On raffinose, the WT dupl− strain had a mean growth rate of 0.17 h^{-1} compared to 0.17 h^{-1} for the complemented strain (P = 0.84 by two-tailed t-test), whereas on stachyose the growth rates and maximal OD were similar (mean growth rate of 0.11 h^{-1} for the WT dupl− strain compared to 0.11 h^{-1} for the complemented strain, P = 0.18 by two-tailed t-test; mean maximal OD_{630} of 0.53 for the WT dupl− strain compared to 0.53 for the complemented strain). Points and error bars represent the mean and SD of n = 3 biological replicates. For each growth curve, the sugar used as the sole carbon source is indicated at the top along with the concentration used. (C) qRT-PCR showing relative *BT1871* mRNA levels in WT, *BT1876* anti-sigma mutant, and the complemented mutant strain. The bars depict the mean and SD of n = 3 biological replicates. All values are normalized to the WT grown on melibiose. Differences between WT and the *BT1876* mutant are significant based on two-way ANOVA. *P < 0.05.

Full degradation of the RFOs raffinose and stachyose requires an α-galactosidase activity (BT1871 in *B. thetaiotaomicron*) to break the α-1,6 glucose-galactose linkage and at least one other enzyme to break the α-1-β−2 glucose-fructose bond of the sucrose moiety (Fig. S1). *B. thetaiotaomicron BT1871* mutant strains fail to grow on melibiose

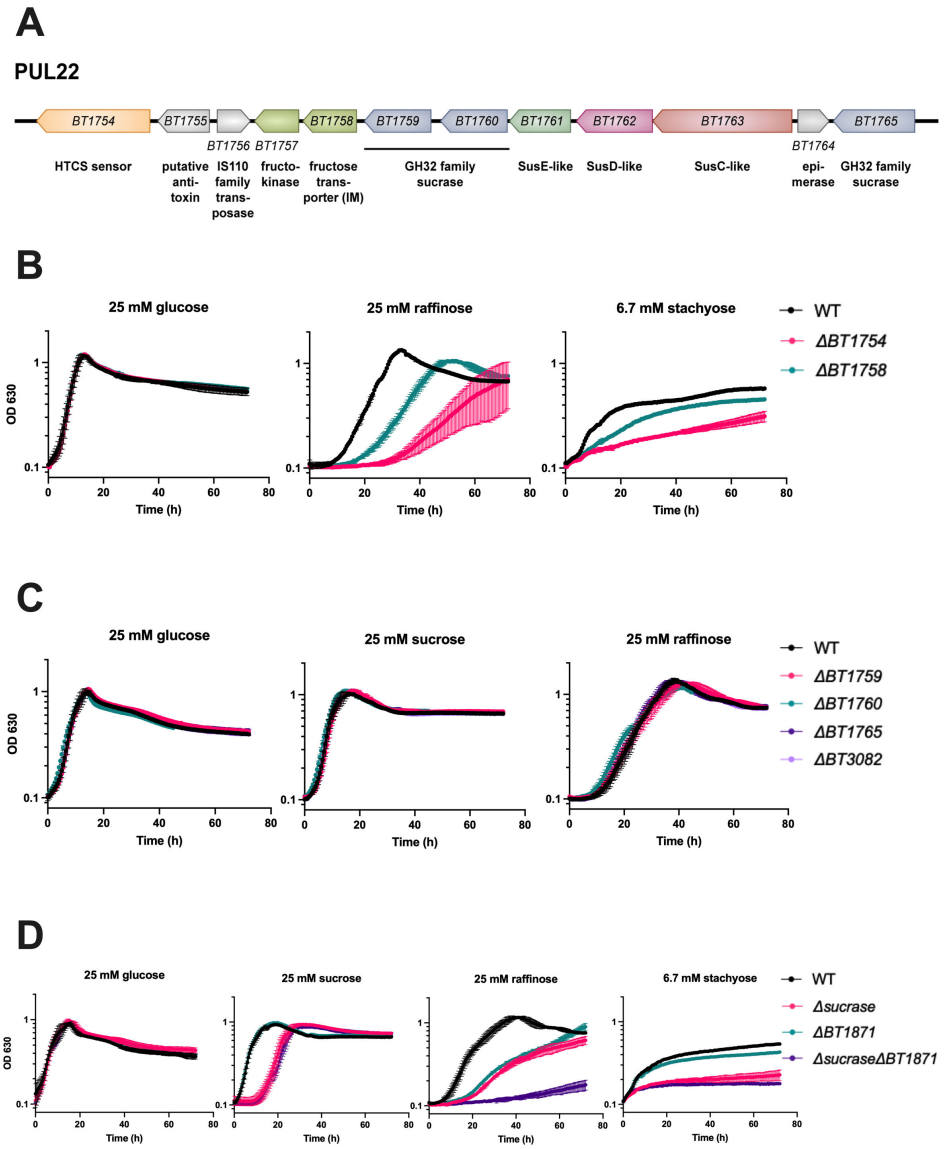

FIG 4 PUL22 is important for RFO utilization in *B. thetaiotaomicron* and its GH32 family sucrases act redundantly to promote RFO utilization. (A) Genomic organization of *B. thetaiotaomicron* PUL22. (B) Growth of WT, Δ*BT1754* (sensor kinase), and Δ*BT1758* (putative fructose transporter) on glucose and raffinose. (C) Growth curves of individual GH32 sucrase mutant strains on glucose, sucrose, and raffinose. *BT1759, BT1760,* and *BT1765* are part of PUL22. *BT3082* is not part of PUL22 but is coregulated with PUL22 genes in response to fructans. On sucrose, the mean growth rate of the WT strain was 0.55 h^{-1} compared to 0.54 h^{-1} for the Δ*BT1759* strain, 0.57 h^{-1} for the Δ*BT1760* strain, 0.56 h^{-1} for the Δ*BT1765* strain, and 0.52 h^{-1} for the Δ*BT3082* strain ($P > 0.05$ for all individual comparisons to the WT strain by two-tailed t-test). On raffinose, the mean growth rate of the WT strain was 0.20 h^{-1} compared to 0.17 h^{-1} for the Δ*BT1759* strain, 0.16 h^{-1} for the Δ*BT1760* strain, 0.20 h^{-1} for the Δ*BT1765* strain, and 0.18 h^{-1} for the Δ*BT3082* strain ($P > 0.05$ for all individual comparisons to the WT strain by two-tailed t-test). (D) Growth curves of WT, *BT1871* mutant, a quadruple GH32 sucrase mutant (Δ*sucrase*) and the Δ*sucrase* mutant in a *BT1871* mutant background (Δ*sucrase*Δ*BT1871*) on RFOs. On sucrose, the WT strain had a mean growth rate of 0.44 h^{-1} compared to 0.39 h^{-1} for the Δ*sucrase* strain ($P = 0.14$ by two-tailed t-test). On stachyose, the Δ*sucrase* strain had a mean growth rate of 0.14 h^{-1} compared to 0.25 h^{-1} for the Δ*sucrase* Δ*BT1871* strain ($P = 0.14$ by two-tailed t-test). Points and error bars represent the mean and SD of $n = 3$ biological replicates. For each growth curve, the sugar used as the sole carbon source is indicated at the top along with the concentration.

(which only contains the α-1,6 glucose-galactose linkage) (Fig. S2A), but still grow on raffinose and stachyose, presumably using fructose liberated by a sucrase that can break

the α-1-β−2 glucose-fructose bond. A large-scale functional study found that transposon insertions in a *B. thetaiotaomicron* gene cluster (*BT2156-BT2160*) conferred fitness defects on raffinose and a few disaccharides (65). We constructed *BT2157* and *BT2158* single deletion mutations in the WT background and probed their growth on RFOs. Strains with mutations in *BT2157* or *BT2158* had no growth defects on RFOs as compared to the WT strain (Fig. S3B). Thus, *B. thetaiotaomicron* does not require *BT2157* or *BT2158* to catabolize the sucrose liberated from RFOs.

To identify other genes that may be involved in RFO use, we performed RNA-seq on WT *B. thetaiotaomicron* grown on raffinose or glucose as the sole carbon source. We found 1,037 genes significantly differentially expressed (absolute fold change >2 and adjusted *P* value < 0.05) on raffinose compared to glucose. Of those, 743 (71.6%) genes were upregulated while 294 (28.4%) were downregulated on raffinose compared to glucose (Fig. S4; Table S2). We noted that *BT1871* (α-gal) was not upregulated in raffinose compared to glucose-grown cells even though it confers a growth advantage. We found that genes in PUL22 were highly upregulated (average fold change of ~776) on raffinose (Fig. 4A; Fig. S4; Table S2). *B. thetaiotaomicron* deploys PUL22 for utilizing fructans such as levan and fructo-oligosaccharides (28). This PUL is induced by fructose through a hybrid two-component sensor kinase BT1754 and contains genes encoding an inner membrane fructose transporter (*BT1758*), a fructokinase (*BT1757*), and three GH32 family enzymes, which include some with sucrase activity (*BT1759*, *BT1760*, and *BT1765*) (28). We hypothesized that these PUL22 sucrases cleave the glucose-fructose bond in RFOs, thus liberating free fructose and activating PUL22 through *BT1754*.

To test the role of PUL22 genes in RFO use, we deleted *BT1754,* which encodes the HTCS activator of PUL22. Indeed, a *B. thetaiotaomicron BT1754* mutant strain has a strong and significant growth defect on raffinose (mean growth rate of 0.22 h^{-1} for the WT strain compared to 0.11 h^{-1} for the Δ*BT1754* strain, $P < 0.01$ by two-tailed t-test) and stachyose (mean growth rate of 0.09 h^{-1} for the WT strain compared to 0.05 h^{-1} for the Δ*BT1754* strain, $P < 0.01$ by two-tailed t-test) (Fig. 4B). To determine whether growth on RFOs requires the PUL22 transporter, GH32 sucrases, or both, we constructed *BT1758* (transporter), and *BT1759, BT1760,* and *BT1765* (putative sucrase) mutant strains. The *BT1758* (transporter) mutant strain had a growth defect on raffinose and stachyose that was less severe than the growth defect of the *BT1754* (HTCS) mutant strain (mean growth rate of 0.17 h^{-1} for the Δ*BT1758* strain on raffinose and 0.10 h^{-1} on stachyose) (Fig. 4B), suggesting that fructose transport via BT1758 is important for growth of *B. thetaiotaomicron* on RFOs. However, mutant strains lacking individual PUL22-encoded sucrases grew as well as the WT parent on both sucrose and raffinose (Fig. 4C).

Another GH32 family enzyme (not encoded in PUL22), BT3082, is produced at high levels in the presence of fructans and has *in vitro* sucrase activity (28). A mutant strain lacking only *BT3082* also had no growth defect on raffinose (Fig. 4C). We hypothesized that the GH32 family sucrases may act redundantly on RFOs since they all break the glucose-fructose bond. To test this, we constructed a *B. thetaiotaomicron* strain lacking all four GH32 enzymes (Δ*BT1759* Δ*BT1760* Δ*BT1765* Δ*BT3082*), which we called the Δ*sucrase* mutant strain. This mutant and the WT strain had a similar growth on sucrose, but the mutant strain had a longer lag phase (mean lag phase of 2.3 hours for the WT strain compared to 14 hours for the Δ*sucrase* strain) (Fig. 4D). Importantly, the Δ*sucrase* mutant strain displayed a significant growth defect on raffinose (mean growth rate of 0.15 h^{-1} for the WT strain compared to 0.10 h^{-1} for the Δ*sucrase* strain, $P < 0.01$ by two-tailed t-test) while on stachyose, even though there was no difference in initial growth rates (mean growth rate of 0.10 h^{-1} for the WT strain compared to 0.14 h^{-1} for the Δ*sucrase* strain, $P = 0.43$ by two-tailed t-test), the WT strain reached a higher maximal OD by the end of the experiment (mean maximal OD_{630} of 0.54 for the WT strain compared to 0.23 for the Δ*sucrase* strain) (Fig. 4D). Combining mutations in *BT1871* (α-gal) and the four GH32 sucrases (Δ*sucrase* Δ*BT1871*) abolished the growth of *B. thetaiotaomicron* on raffinose as the sole carbon source, whereas the Δ*sucrase* Δ*BT1871* and Δ*sucrase* strains grew similarly and very poorly on stachyose (Fig. 4D). Taken together, we conclude that

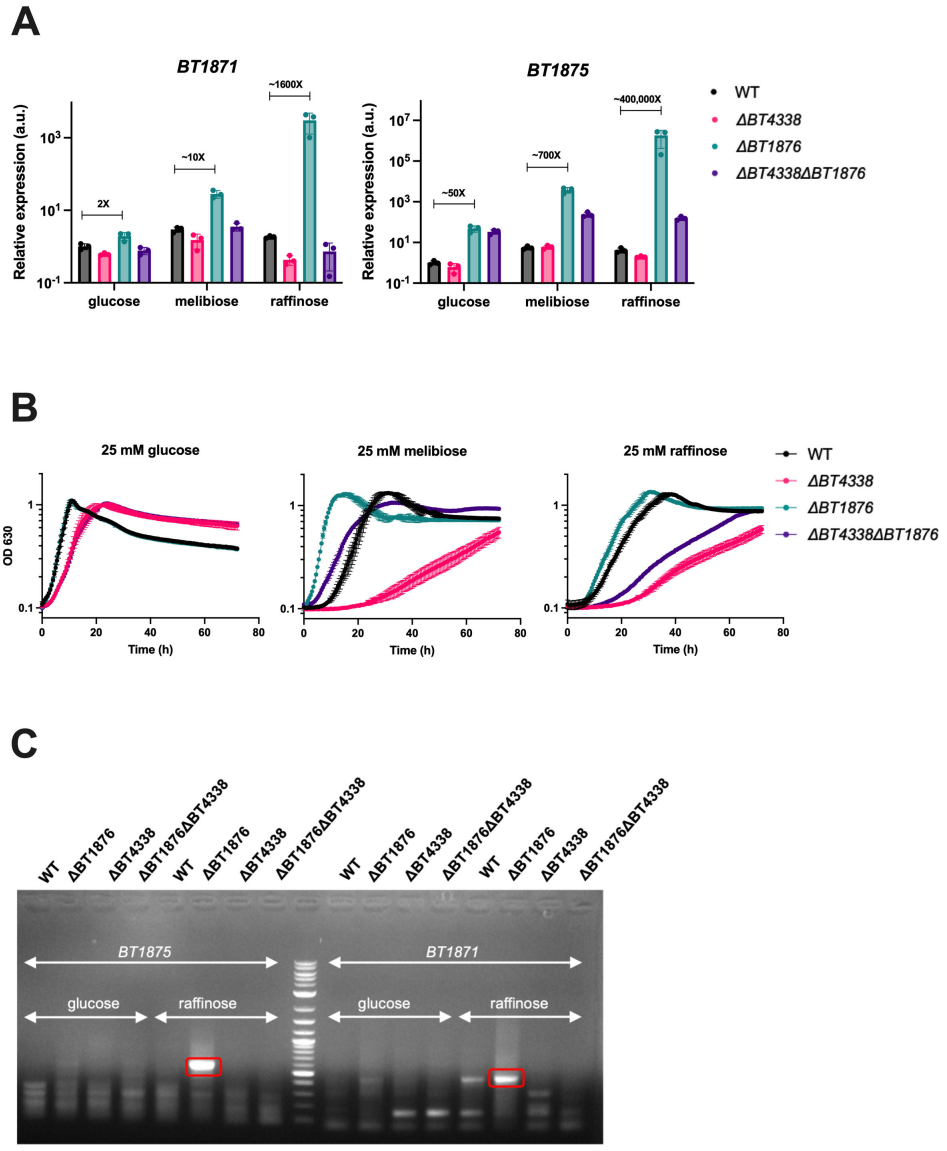

FIG 5 The global regulator *BT4338* is important for RFO utilization in *B. thetaiotaomicron*. (A) qRT-PCR was used to measure relative levels of *BT1875* (PUL24 SusC-like transporter) and *BT1871* (PUL24 α-galactosidase) mRNAs. The media and strains used are indicated. The bars depict the mean and SD of n = 3 biological replicates. All values are normalized to the level in the WT strain grown on glucose. (B) Growth curves of WT, *BT4338* (global regulator of carbon utilization) mutant, *BT1876* (PUL24 anti-sigma) mutant, and their double mutant strains on RFOs. Points and error bars represent the mean and SD of n = 3 biological replicates. For each growth curve, the sugar used as the sole carbon source is indicated at the top along with the concentration used. (C) 5′ RACE to identify 5′ ends for *BT1875* and *BT1871* mRNAs. The bands that were sequenced for identifying novel 5' ends are highlighted in red boxes and the middle lane corresponds to a DNA ladder. Fig. S5 depicts the location of sequenced 5′ ends with respect to the flanking genes.

B. thetaiotaomicron uses enzymes from both PUL22 and PUL24 for utilizing RFOs and the GH32 family sucrases in *B. thetaiotaomicron* have redundant functions with respect to RFO usage.

RFO-dependent activation of PUL24 genes in the *BT1876* anti-sigma mutant requires a global regulator of carbohydrate utilization

We expected that if PUL24 is dedicated to RFO sensing and utilization, the transcription of PUL24 genes would be strongly activated in response to RFOs through inactivation

of the anti-sigma BT1876 and release of the sigma factor BT1877. Our RNA-seq analysis suggested only weak activation of PUL24 genes in response to raffinose (Table S2), so we used RT-qPCR to further explore this. We measured levels of two portions of PUL24 mRNA, *BT1875*, encoding the SusC-like transporter, and *BT1871,* encoding the α-galactosidase important for RFO use. We found that *BT1875* (transporter) transcripts were upregulated 5.5-fold on melibiose and 4.2-fold on raffinose compared to glucose in the WT strain and *BT1871* (α-gal) was upregulated 3-fold in melibiose and 2-fold in raffinose compared to glucose (Fig. 5A, compare black bars). This level of activation is very small compared to what is seen for activation of other PULs by their cognate substrates. For example, genes in PUL22 are upregulated 50-fold to 500-fold in the presence of the cognate substrate levan (28) while genes in PUL7 are upregulated >1,000-fold in the presence of the substrate arabinan (34). We reasoned that the fully induced expression levels of PUL24 genes would be observed in the *BT1876* (anti-sigma) mutant strain irrespective of growth substrate if the sigma factor (BT1877) is the primary regulator of PUL24 gene expression. In the *BT1876* (anti-sigma) mutant strain, we observed a 50-fold upregulation of *BT1875* (transporter) compared to the WT strain when grown on glucose (Fig. 5A). Interestingly, levels of *BT1875* (transporter) transcripts in the *BT1876* (anti-sigma) deletion strain were further increased when cells were grown on melibiose (~700-fold increase compared to WT) and raffinose (~400,000-fold increase compared to WT). This trend of dramatically increased transcript levels in response to RFOs in the *BT1876* (anti-sigma) mutant background was also observed for *BT1871* (Fig. 5A). These data suggest that there are multiple mechanisms of regulation of PUL24 genes relevant to RFO utilization.

B. thetaiotaomicron possesses a conserved protein BT4338, which has some similarity to the CRP transcription factor that controls the catabolite repression response in enteric bacteria (13, 22, 66). BT4338 is required for efficient use of a variety of monosaccharides and polysaccharides *in vitro* (34) and many *B. thetaiotaomicron* genes were shown to be differentially regulated in either rich media or under carbon limiting conditions in a *BT4338* mutant strain compared to the wild-type strain (67). We wondered whether BT4338 plays a role in the activation of PUL24 genes in response to RFOs. To test this, we measured levels of *BT1875* (transporter) and *BT1871* (α-gal) mRNAs in *BT4338* and *BT4338 BT1876* mutant strains (Fig. 5B). There were small decreases in levels of *BT1875* (transporter) and *BT1871* (α-gal) mRNAs in the *ΔBT4338* mutant strain compared to wild type in all the conditions tested (Fig. 5A). In the glucose growth condition, Δ*BT1876* (anti-sigma) and Δ*BT1876* Δ*BT4338* strains had very similar levels of both the transporter and α-gal transcripts. In sharp contrast, the strong RFO-dependent activation observed for both transcripts in the Δ*BT1876* (anti-sigma) mutant was strongly attenuated in the Δ*BT1876* Δ*BT4338* mutant (Fig. 5A). These data suggest that RFO-dependent activation of PUL24 genes in the anti-sigma mutant strain either directly or indirectly depends on BT4338.

We next tested whether the growth of these same strains was impacted by the presence or absence of BT4338. There was a small but significant difference in growth rate between the *BT4338* mutant strains and their respective parent strains on glucose as the sole carbon source (mean growth rate of 0.56 h^{-1} for the WT strain compared to 0.35 h^{-1} for the Δ*BT4338* strain, $P = 0.02$ by two-tailed t-test; mean growth rate of 0.53 h^{-1} for the Δ*BT1876* strain compared to 0.29 h^{-1} for the Δ*BT4338* Δ*BT1876* strain, $P < 0.01$ by two-tailed t-test) (Fig. 5B). In melibiose, where *BT1871* expression levels are strongly correlated with growth [(60) and Fig. 1B and C], the Δ*BT4338* strain was strongly growth inhibited compared to the WT parent (mean growth rate of 0.30 h^{-1} for the WT strain compared to 0.08 h^{-1} for the Δ*BT4338* strain, $P < 0.01$ by two-tailed t-test) (Fig. 5B). This may be explained by the reduced abundance of *BT1871* mRNA observed under this growth condition (Fig. 5A). The growth advantage displayed by the Δ*BT1876* strain compared to the WT parent was eliminated in the Δ*BT1876* Δ*BT4338* strain (Fig. 5B), consistent with reduced *BT1871* mRNA levels observed in the double mutant compared to the Δ*BT1876* parent (Fig. 5A). In raffinose, where genes in both PUL22

(sucrases) and PUL24 (α-gal) are important for growth, we saw the same overall patterns as for melibiose. The Δ*BT4338* strain was strongly growth inhibited compared to the WT parent (mean growth rate of 0.19 h^{-1} for the WT strain compared to 0.09 h^{-1} for the Δ*BT4338* strain, $P < 0.01$ by two-tailed t-test) and the growth advantage displayed by the Δ*BT1876* strain was eliminated in the Δ*BT1876* Δ*BT4338* strain (Fig. 5B). These patterns of growth are consistent with the expression levels of *BT1871* mRNA (Fig. 5A). Taken together, we conclude that *BT4338* is important for RFO utilization in *B. thetaiotaomicron* by modulating (directly or indirectly) transcription of *BT1871* in response to RFOs.

PUL24 genes are expressed from transcription start sites dependent on BT4338 and BT1877

A recent study annotated transcription start sites (TSSs) in *B. thetaiotaomicron* grown in rich media and identified transcription start sites upstream of *BT1877* and *BT1872* within PUL24 [(63), Fig. 3A]. To find putative TSSs in PUL24 that are active during growth on glucose and RFOs, we performed 5′ RACE. We harvested RNA from WT, Δ*BT1876*, Δ*BT4338,* and Δ*BT4338* Δ*BT1876* strains grown on glucose or raffinose and used primers to define 5′ ends near *BT1875* and *BT1871*. In the *BT1876* (anti-sigma) mutant grown on raffinose, we observed a strong PCR band representing a 5′ end upstream of *BT1875* (Fig. 5C). A fainter band of the same size was also observed in the *BT1876* mutant grown on glucose. Importantly, this band was absent in the WT, *BT4338,* and the *BT1876 BT4338* double mutant on both glucose and raffinose. After sequencing, we mapped this 5′ end to a position 291-nt upstream of the *BT1875* coding sequence (Fig. S5A). In fact, this position is inside the 3′ end of the *BT1876* coding region. We found no consensus promoter sequence for the housekeeping sigma factor upstream of this 5′ end, which suggests that this may represent a site for transcription initiation by the PUL24 sigma factor *BT1877*.

Using primers to detect 5′ ends upstream of *BT1871,* we observed a strong signal in the *BT1876* mutant strain grown on raffinose (Fig. 5C). A fainter band of the same size was also observed for the *BT1876* mutant grown on glucose while there was no band for the *BT4338* mutant or *BT4338 BT1876* double mutant on glucose or raffinose. A faint band was also observed for the WT strain grown on raffinose indicating possible *BT1871* transcription even in the presence of the anti-sigma BT1876. This signal represented a site 17-nt upstream of the *BT1871* coding sequence (Fig. S5B). These results are consistent with RFO-dependent activation of transcription at a promoter upstream of *BT1871* that is responsive to BT4338.

We hypothesized that *BT1877* (sigma factor) is responsible for both basal (in the WT strain) and activated (in the Δ*BT1876* strain) transcription of PUL24 genes. However, a *BT1877* mutant strain grew similar to the WT strain on RFOs (Fig. S6A). This suggests that *BT1877* is not required for basal transcription of *BT1871*. A Δ*BT1876* Δ*BT1877* double mutant lost the growth advantage of the *BT1876* single mutant when grown on RFOs and completely failed to grow on AMG (Fig. S6B). These observations suggest that the PUL24 sigma factor *BT1877* is not required for basal transcription of *BT1871* but is responsible for its higher transcript levels in the anti-sigma mutant.

Taken together, our results are consistent with a model where activated transcription of PUL24 genes including *BT1871* in response to RFOs requires the PUL24 sigma factor *BT1877* and the global regulator *BT4338*.

BT1871 is important for RFO utilization in other *Bacteroides* species

To determine whether mechanisms of RFO utilization are conserved across *Bacteroides* species, we identified homologs of *BT1871* that shared >50% identity across >95% of the protein sequence. To assess the importance of these *BT1871* homologs in RFO utilization, we compared the growth of species with *BT1871* homologs namely, *B. thetaiotaomicron*, *Bacteroides ovatus, Bacteroides caccae, Bacteroides faecis, Bacteroides uniformis,* and *Bacteroides intestinalis* (Table S3) with the growth of species lacking a *BT1871* homolog*, Bacteroides eggerthii*, *Bacteroides fragilis,* and *Bacteroides salyersiae*. All these organisms

grew similarly on rich (TYG) medium, except *B. eggerthii* which had a long lag phase. The growth of these species in a minimal medium with glucose was more variable (Fig. 6A and B). Notably, the group of organisms with a *BT1871* homolog grew on melibiose, albeit with different lags and growth rates (Fig. 6A) while all the species lacking a *BT1871* homolog failed to grow on melibiose as the sole carbon source (Fig. 6B). The two species with the most distant homologs namely, *B. uniformis* and *B. intestinalis* (58% and 56% identity to *BT1871,* respectively) had the slowest growth rates and lowest maximum optical densities when grown on melibiose. This result suggests that having a *BT1871* homolog is important for the growth of *Bacteroides* on melibiose. All the tested strains except *B. eggerthii* DSM 20697 showed some degree of growth on raffinose as the sole carbon source (Fig. 6A and B). Species with *BT1871* homologs tended to have a higher maximal OD on raffinose compared to species without a homolog, although there were no significant differences in raffinose growth rates associated with the presence or absence of *BT1871* homolog (Fig. S7A and B). The increased maximal OD may be because the species with *BT1871* homologs can break down the galactose-glucose bond in raffinose as well as the glucose-fructose bond, thereby using all three monosaccharides, whereas species without the *BT1871* homolog are unable to break the galactose-glucose bond and can only use the liberated fructose.

To further test the importance of *BT1871* in RFO utilization in other *Bacteroides* species, we deleted the homologs of *BT1871*, namely *BACOVA_02959* and *BACOVA_02767* in *B. ovatus* ATCC 8483. These GH97 family α-galactosidases have 87% and 56% amino acid identity to *BT1871,* respectively (Table S3). Furthermore, *BACOVA_02959* is part of *B. ovatus* ATCC 8483 PUL60 which has synteny to PUL24 in *B. thetaiotaomicron* (32). If these α-galactosidases are necessary and sufficient for cleaving the α-1,6 bond in melibiose, we would expect the double mutant would fail to grow on melibiose as a sole carbon source. However, the double mutant strain (Δ*BACOVA_02959*Δ*BACOVA_02767*) did not show a growth defect on RFOs and grew similar to its parent WT strain (Fig. 6C). It is worth noting that *B. ovatus ATCC 8483* has 10 other GH97 family enzymes (with <50% identity to *BT1871*) that might work redundantly to catabolize RFOs (32).

Finally, we wanted to test whether the heterologous expression of *BT1871* is sufficient to confer better RFO utilization in organisms that lack a homolog. A constitutive *BT1871* expression cassette (*BT1871* expressed by the strong housekeeping sigma^{70} promoter) was constructed and validated by integration into the genome of a *B. thetaiotaomicron BT1871* mutant strain as a single copy insert. The complemented strain (Δ*BT1871_pNBU2_BT1871*) grew significantly better than the *BT1871* mutant strain on RFOs as the sole carbon source (Fig. S8). On melibiose, the mean growth rate of the Δ*BT1871* strain was 0.10 h^{-1} compared to 0.48 h^{-1} for the complemented strain ($P < 0.01$ by two-tailed t-test). On raffinose, the mean growth rate of the Δ*BT1871* strain was 0.06 h^{-1} compared to 0.20 h^{-1} for the complemented strain ($P < 0.01$ by two-tailed *t*-test) and on stachyose the initial growth rates were similar (mean growth rate of 0.15 h^{-1} for the Δ*BT1871* strain compared to 0.13 h^{-1} for the complemented strain, $P = 0.16$ by two-tailed t-test) but the complemented strain reached a higher maximal OD than the *BT1871* mutant (0.41 for the Δ*BT1871* strain compared to 0.53 for the complemented strain). This expression cassette was then integrated into the genomes of *B. fragilis 638R* and *B. eggerthii DSM 20697* which lack a *BT1871* homolog. Constitutive *BT1871* expression conferred the ability to grow on melibiose as the sole carbon source on both species (Fig. 6D). However, the *BT1871*-expressing strains grew similarly to their WT counterpart when grown on raffinose or stachyose as the sole carbon source (Fig. 6D). Taken together, we conclude that *BT1871* homologs are important for catabolizing melibiose in different *Bacteroides* species.

DISCUSSION

In this work, we have shown how the model commensal *B. thetaiotaomicron* utilizes raffinose family oligosaccharides (Fig. 7). When *B. thetaiotaomicron* encounters RFOs in its environment, they are transported across the outer membrane by an unknown

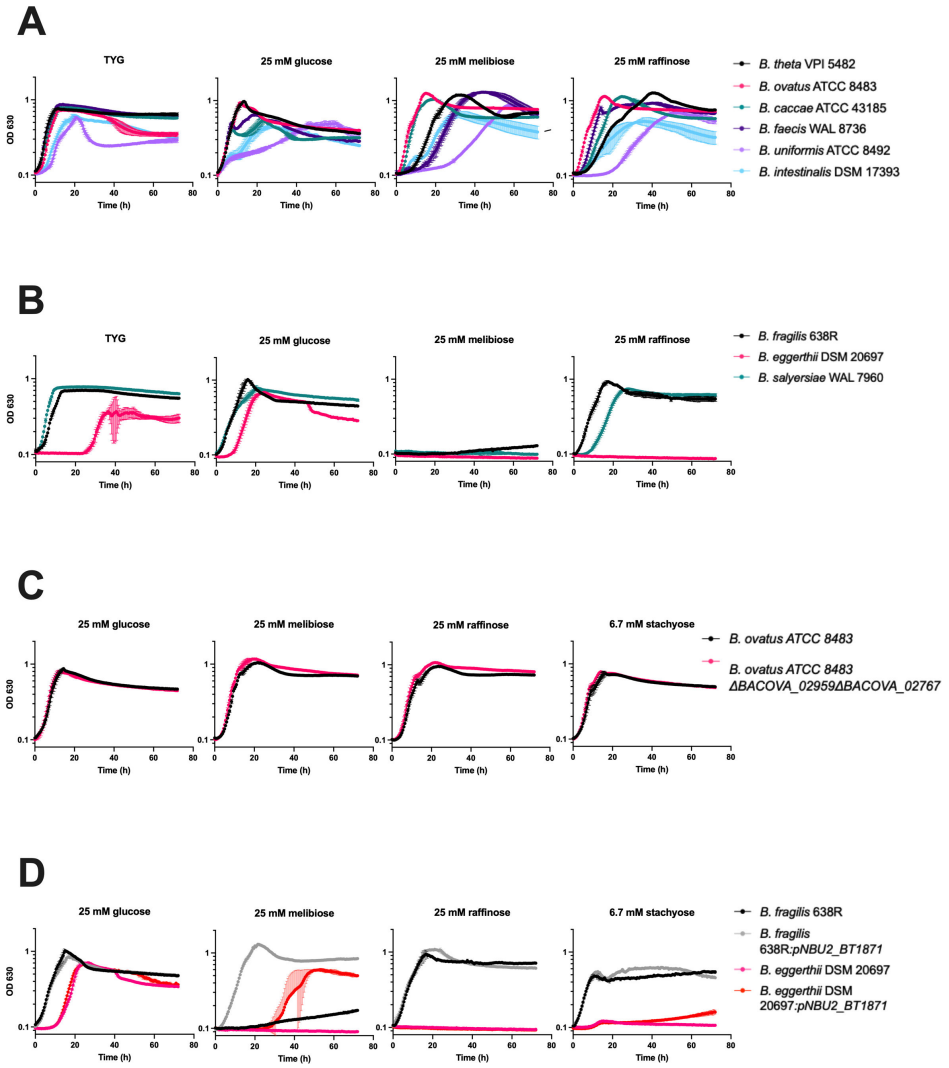

FIG 6 The alpha-galactosidase BT1871 and its homologs are important for melibiose utilization in *Bacteroides* species. (A) Growth curves of *Bacteroides* species containing a *BT1871* homolog on RFOs. (B) Growth curves of *Bacteroides* species lacking a *BT1871* homolog on RFOs. (C) Growth curves of wild-type *B. ovatus* ATCC 8483 (BO) and a mutant strain lacking two *BT1871* homologs on RFOs. On melibiose, the WT strain had a mean growth rate of 0.32 h^{-1} compared to 0.37 h^{-1} for the double mutant strain, $P = 0.06$ by two-tailed t-test. On raffinose, the WT strain had a mean growth rate of 0.32 h^{-1} compared to 0.37 h^{-1} for the double mutant strain, $P = 0.12$ by two-tailed t-test. On stachyose, the WT strain had a mean growth rate of 0.40 h^{-1} compared to 0.51 h^{-1} for the double mutant strain, $P = 0.14$ by two-tailed t-test). (D) Growth curves of *B. fragilis* 638R and *B. eggerthii* DSM 20697 alongside their derived strains expressing *BT1871* from a constitutive promoter. On raffinose, the *B. fragilis* WT strain had a mean growth rate of 0.33 h^{-1} compared to 0.32 h^{-1} for the *BT1871* expressing strain ($P = 0.69$ by two-tailed t-test) while both the strains of *B. eggerthii* failed to grow. On stachyose, the *B. fragilis* WT strain had a mean growth rate of 0.70 h^{-1} compared to 0.57 h^{-1} for the *BT1871* expressing strain ($P = 0.21$ by two-tailed t-test) while both the strains of *B. eggerthii* failed to grow.

In all panels, points and error bars represent the mean and SD of $n = 3$ biological replicates. For each growth curve, the sugar used as the sole carbon source is indicated at the top along with the concentration used.

transporter. In the periplasm, the α-galactosidase BT1871 cleaves the α-1,6 galactoside bonds, liberating galactose while the remaining sucrose moiety is cleaved by PUL22-encoded GH32 family sucrases. Monomeric fructose activates genes in PUL22 *via* the HTCS BT1754, resulting in increased expression of these genes. Specific inner membrane proteins then transport the monomers glucose, galactose, and fructose into the

cytoplasm where they enter central metabolism. *BT1871* is transcribed at a basal level by an unknown sigma factor and this expression is dependent on the global regulator BT4338. Upon deletion of the anti-sigma *BT1876* or in the presence of the unknown natural substrate of PUL24, expression of PUL24 is upregulated. The presence of RFOs and BT4338 promotes high levels of transcription from the novel transcription start sites that are not active undergrowth in rich media (63). BT4338 may also control the expression of PUL22 since ChIP-seq data indicate the presence of several binding sites for BT4338 within PUL22 (67).

Oligosaccharides are prevalent in a plant-rich diet and are gaining popularity as prebiotics (68–70). Here we have shown that hydrolases from two separate PULs, PUL24 and PUL22, work together to degrade an abundant class of plant oligosaccharides, namely RFOs, in *B. thetaiotaomicron*. This dual function of enzymes encoded by PULs capable of degradation of more complex substrates for catabolism of oligosaccharides is also seen for human milk oligosaccharides (HMOs) which are degraded by *B. thetaiotaomicron* using hydrolases from PULs that usually catabolize mucins (71). This is likely due to structural similarity between some components of HMOs and core mucin structures and the presence of similar glycosidic bonds. Similarly, *B. thetaiotaomicron* can utilize galacto-oligosaccharides using enzymes from its pectin-galactan utilization and mucin utilization PULs (72), while fructo-oligosaccharides and arabino-oligosaccharides are degraded using enzymes from the fructan PUL and arabinan PUL, respectively (28, 34). While the natural substrate for PUL24 is still unknown, the presence of an endo-arabinase (*BT1873*) suggests that the substrate could be a plant-derived glycan since animal (host)-derived glycans do not contain arabinose (73, 74). However, most plant-derived glycan degrading PULs contain an HTCS regulator while PUL24 has a sigma, anti-sigma pair controlling it. Further work needs to be done to elucidate the natural substrate of PUL24.

We showed that deleting the global regulator BT4338 affects the growth on RFOs and the expression of PUL24 genes. However, the mode of action of BT4338 on this PUL is unknown. Does it act as an activator by directly binding to promoters in PUL24 as for *fusA2* (67) or does it indirectly control expression *via* some other activator as it does for *roc* (66)? Recent work has shown that BT4338-activated genes are dominantly silenced by the monosaccharides glucose and fructose (66). It is likely that the low level of PUL24 induction in minimal media glucose is because of the same phenomenon and growth on other substrates such as melibiose or raffinose alleviates this silencing. Interestingly, RNA-seq data showed that there is transcription from the 3′ regions of *BT1876* in nutrient-starved *B. thetaiotaomicron* cells while there was no transcription from this region in glucose-grown cells (75). The transcription under this condition starts very close to where we mapped the putative 5′ end upstream of *BT1875* using 5′ RACE (Fig. S5A). Importantly, starvation is known to activate *BT4338* and induce widespread gene expression changes in *B. thetaiotaomicron*, possibly by changing levels of some metabolite that modulates BT4338 activity. We failed to find any motif resembling the BT4338 consensus binding sequence (wwwTATGTTnTAnAACATAwww) (22) upstream of the *BT1875* 5′ end and a ChIP-seq study did not reveal any BT4338-binding sites upstream of *BT1875* (67). This suggests that BT4338 control of *BT1875* expression may be indirect in nature.

Finally, it is important to think about RFO usage by *Bacteroides* in the context of the densely packed and competitive environment of the gut. As mentioned above, *Bacteroides* rely on enzymes encoded by genes in PULs for RFO degradation. However, other commensals residing in the colon like *Bifidobacterium* and *Lactobacillus* harbor operons that encode dedicated RFO regulation, transport, and metabolism functions (51, 53, 57). This may reflect that *Bacteroides* are better adapted to preferentially degrade longer polysaccharides, whereas other commensals are more efficient degraders of smaller oligosaccharides like RFOs or HMOs. Indeed, *Bifidobacterium animalis* subsp. *lactis Bl-04*, which has a dedicated RFO utilization system, can outcompete *B. ovatus* when grown on raffinose as the sole carbon source (57). Moreover, *in vitro* fermentation of fecal inoculum containing raffinose from healthy human donors led to an increased abundance of

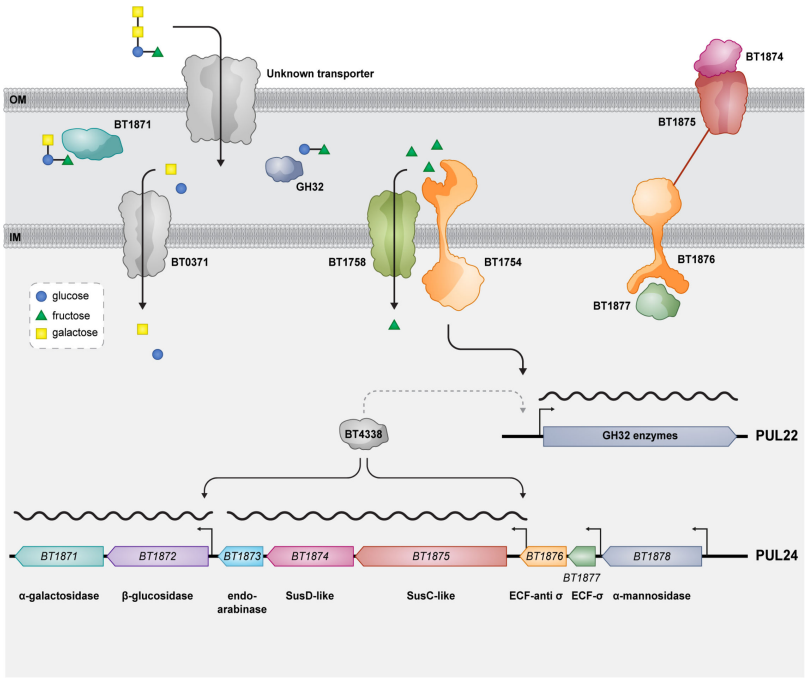

FIG 7 Model for raffinose family oligosaccharide utilization by *B. thetaiotaomicron*. After crossing the outer membrane (OM), RFOs are acted upon by BT1871 which cleaves the α-1,6 galactoside bond between glucose and galactose. Next, GH32 family enzymes act redundantly on the liberated sucrose moiety in the periplasm. The monosaccharide subunits are transported across the inner membrane (IM) by dedicated transporters. Expression of PUL22 genes depends on the HTCS BT1754 which binds fructose and activates PUL22 genes, including the sucrases that cleave the bond between glucose and fructose. The natural signal that induces PUL24 genes, including *BT1871*, is unknown, but basal levels of *BT1871* transcription depend on BT4338. When the PUL24 anti-sigma factor is deleted, there is an RFO and BT4338-dependent upregulation from novel transcription start sites in PUL24 whose approximate locations are indicated by bent arrows. A dashed arrow indicates that BT4338 may also control the expression of PUL22 genes.

Bifidobacteriales and *Lactobacillales* with a concurrent decrease in *Bacteriodales* (54). These observations suggest that in microbial communities where *Bacteroides* species are competing with organisms with dedicated RFO utilization systems, *Bacteroides* may not be the primary RFO users. However, a recent study showed that the acquisition of mutations like the ones we report here may indeed be relevant *in vivo* for *Bacteroides* RFO utilization. The authors of this study found that the expression of genes in PUL24 increased when germ-free mice fed an RFO-rich diet were monocolonized with *B. thetaiotaomicron* (64), and they found evidence of spontaneous duplication of the *BT1871* locus in these mice. Strains with *BT1871* duplications in their study varied with respect to the length of the duplicated region and exact junctions, but all showed the same outcome as we observed: increased expression of *BT1871* due to transcription from a ribosomal promoter driving the duplicated copy. These studies were performed in mice fed a diet rich in melibiose and raffinose where increased expression of *BT1871* would be advantageous. Future functional studies using different diets and competition assays with other gut commensals will be needed to fully elucidate the importance of PUL24 and RFO utilization for *Bacteroides in vivo*.

MATERIALS AND METHODS

Media and growth conditions

All strains, plasmids, and oligonucleotides used in this study are listed in Tables S4 to S6, respectively. *B. thetaiotaomicron* VPI-5482 and other *Bacteroides* strains were grown in a Coy Laboratory Products vinyl anaerobic chamber with an input gas of 20% CO_2, 10% H_2, and 70% N_2 balance. Routine culturing of *Bacteroides* was done in TYG (tryptone, yeast extract, and glucose) broth (76) and on Difco brain heart infusion (BHI) agar plates supplemented with 10% defibrinated horse blood (HB; Quad Five) or 5 mg/L hemin and 2.5 µg/L vitamin K3 (BHIS) at 37°C. *Escherichia coli* strains were grown aerobically at 37°C on BHIS for conjugations and in LB for all other applications. Minimal medium (77) was supplemented with B_{12} (3.75 nM, final; Sigma) and carbohydrates as needed at a final concentration of 25 mM except stachyose which was used at 0.5% wt/vol. When needed, antibiotics were added at the following final concentrations: 100 µg/mL ampicillin (Sigma), 200 µg/mL gentamicin (Goldbio), and 25 µg/mL erythromycin (VWR).

Construction of strains and genetic manipulation

In-frame gene deletions and mutations were generated using the counter-selectable allelic exchange vector pLGB13 (78). Flanking regions of 1 kb on either side of the target gene were PCR amplified using Q5 polymerase (NEB) and cloned into pLGB13 using the NEBuilder HiFi cloning kit (NEB). The pLGB13 plasmid was linearized by double digest with BamHI-HF and SalI-HF restriction enzymes (NEB). Ligated plasmids were transformed into *E. coli* S-17 Lambda pir cells and positive clones were identified using colony PCR (Gotaq green) and sent for whole plasmid sequencing (Plasmidsaurus). Overnight cultures of the recipient *Bacteroides* as well the donor *E. coli* strains were diluted 1:100 and 1:200, respectively, in 5 mL fresh medium and grown for ~6 h. 1 mL of each culture was centrifuged at 5,000 × *g* for 10 min and both pellets were combined by resuspending in 100 µL 1× PBS. The suspension was spread out on BHIS plates and incubated aerobically at 37°C overnight to form a lawn. The next day, the lawn was scraped and resuspended in 5 mL 1× PBS. 5 µL of this suspension was spread on BHIS with erythromycin and gentamicin plates and incubated anaerobically. After 2–3 days, two colonies were selected and restreaked for isolation, and colony PCR was performed to verify cointegrants. For colony PCR, a single colony was lysed for 15 min at 95°C in 20 µL of 25 mM NaOH, 0.2 mM disodium EDTA, pH 12. The lysate was neutralized with 40 µL of 40 mM Tris-HCl, pH 5.0 which was diluted 1:2 in ddH_2O. 2 µL of the diluted lysate was used for subsequent PCR. For counter-selection, a single cointegrate colony was grown overnight in TYG, diluted 1:100 in 5 mL fresh TYG, and grown for ~6 h. 3 µL of the culture was spread on BHIS plates containing 100 ng/mL of anhydrotetracycline (Sigma) for counter-selection. After 3–4 days, eight colonies were inoculated into 200 µL TYG in a 96-well plate and grown overnight. 4 µL of the cultures was lysed as mentioned above and analyzed by PCR to screen for gene deletion.

Complementation of deletion strains and introduction of *BT1871* in other *Bacteroides* species was done using the pNBU2-ErmG vector (79). The gene to be introduced was cloned into the pNBU2 vector, was transformed into *E. coli* S-17 cells and positive clones were identified. The vector was conjugated into recipient *Bacteroides* strains as described above. After conjugation, cells were spread on BHIS with erythromycin and gentamycin plates. After 2–3 days, two colonies were restreaked on BHIS plus erythromycin plates for isolation. A single integrant was selected, grown overnight in TYG with erythromycin, and 4 µL of the culture was lysed as above and subjected to PCR to verify which *att* site the plasmid has integrated into. A further PCR verification was done using primers specific to pNBU2 to confirm integration into the genome.

Minimal media growth assays

Strains were cultured in triplicate in 5 mL TYG overnight to the stationary phase. 1 mL of overnight cultures was spun down at 5,000 × *g* for 10 min at room temperature.

Supernatants were removed, and the pellets were washed twice by resuspending in 1 mL of minimal medium without a carbon source. Next, 2 µL of resuspended cells was inoculated into 198 µL of minimal media containing carbon sources in flat-bottom, 96-well Corning Costar tissue culture-treated plates (catalog #3598). Plates were sealed with a Breathe-Easy gas permeable membrane (Sigma) to prevent evaporative loss and facilitate gas exchange and statically cultured anaerobically in the BioTek plate reader at 37°C for as long as needed (usually 72 h) with the optical density (OD_{630}) recorded every 30 min. Growth rates were calculated using the Growthcurver (80) package implemented in R (v4.3.2) by taking the portion of the growth curves till their maximal OD_{630}. The lag phase was calculated using the "Microbial Lag Calculator" web-based app *via* the "parameter fitting to a model method" (81).

Measurement of gene expression by RT-qPCR

Strains were grown in triplicate in 5 mL TYG to stationary phase overnight and culture volumes equivalent to 2 OD_{600} units were washed twice by centrifuging at 5,000 × *g* for 10 min and resuspending in 1 mL minimal medium without a carbon source. Washed cells were inoculated 1:100 in a 5-mL minimal medium containing a carbon source. All cultures for strains containing pNBU2_ermG vectors contained erythromycin. Cultures were grown to the mid-exponential phase at an OD_{600} ~0.8. Next, 5 mL of cells were pelleted at 4,000 × *g* for 10 min at 4°C, the supernatant was decanted, and then RNA was isolated with a Qiagen RNeasy mini kit according to the manufacturer's instructions. Residual DNA was degraded using 4–5 µL DNase-I (Thermo) and RNA was cleaned up using phenol-chloroform followed by overnight ethanol, sodium acetate precipitation, and quantified using a qubit assay kit (Thermo) or nanodrop. We used a probe-based approach for RT-qPCR to reduce off-target amplification and allow quantification of target genes and normalization (using 16S rRNA) in a single reaction. cDNA synthesis and PCR amplification were done with the Luna Probe One-Step RT-qPCR 4X Mix with UDG according to the manufacturer's instructions. 10 µL reactions were performed using 20–25 ng RNA as input. 400 nM of gene-specific primers and 200 nM of gene-specific probes were used. The probe for the 16S rRNA gene was labeled with 5′-FAM (IDT) while *BT1871* and *BT1875* specific probes were labeled with 5′-HEX (IDT). The primers and probes were designed using Primer3 (through Benchling) with an optimum Tm of 60°C and an optimum amplicon length of 150 bp. To calculate primer efficiency, five 5-fold dilutions of template RNA (RNA extracted from a Δ*BT1876* strain grown on raffinose) were made by serial dilution and qPCRs run in triplicate as described above. Primer efficiencies were calculated using the slope of the Ct versus the log of concentration curve which was obtained by linear regression (Prism). All three primer pairs had an efficiency of >80% and the values are given in Table S6. Using the ddCT method, raw values were normalized to 16S rRNA values and then minimal medium with melibiose or raffinose values were referenced to the values obtained in minimal medium with glucose to obtain a fold-change.

5′ RACE

5′ RACE was done using a Template-Switching (TS) Reverse Transcriptase Enzyme mix (NEB) that takes advantage of a template-switching reverse transcriptase and a template-switching oligonucleotide (TSO). The protocol is identical to the one in (82). Briefly, 250 ng of total RNA (in 4 µL) was mixed with 1 µL random hexamers (50 µM), 1 µL dNTP (10 mM), and heated at 70°C for 5 min and kept on ice. 2.5 µL of TS buffer (4×), 0.5 µL of a template switching oligo (75 µM), and 1 µL of TS Enzyme mix (10×) were added to the RNA and incubated at 42°C for 90 min and then at 85°C for 5 min. The RT reaction was diluted twofold with nuclease-free water and 1 µL of the diluted mix was subjected to 5′ RACE PCR using Q5 polymerase (NEB) with a touchdown PCR protocol. A TSO-specific primer and a gene-specific primer were used for each PCR. The PCR products were separated on a 1% agarose gel and bands were excised using a QIAquick gel extraction

kit (Qiagen). The purified PCR bands were cloned using the NEB PCR cloning kit following the manufacturer's instructions and transformed colonies were screened using colony PCR. Positive colonies were grown overnight, plasmids were extracted, and sent for whole plasmid sequencing.

Whole-genome sequencing sample preparation, processing, and analysis

Overnight cultures were grown and pelleted, and total high molecular weight genomic DNA was purified using a DNEasy blood and tissue kit (Qiagen) according to the manufacturer's instructions. A total of 500 µL of overnight culture was used to extract genomic DNA. Purified DNA was submitted for whole-genome sequencing at SeqCenter (Pittsburgh, PA) using both Illumina and Oxford Nanopore sequencing. For the WT dupl− mutants capable of growth on AMG as the sole carbon source, sample libraries for Illumina sequencing were prepared using the Illumina DNA Prep kit and IDT 10 bp unique dual indices, and sequenced on an Illumina NextSeq 2000, producing 2 × 151 bp reads. Demultiplexing, quality control, and adapter trimming were performed with bcl-convert (v3.9.3) (83). Nanopore samples were prepared for sequencing using Oxford Nanopore's "Genomic DNA by Ligation" kit (SQK-LSK109) and protocol. All samples were run on Nanopore R9 flow cells (R9.4.1) on a MinION. Guppy (v5.0.16) (84) was used for high-accuracy basecalling, demultiplexing, and adapter removal. Porechop (v0.2.3_seqan2.1.1, default parameters) (85) was used to trim residual adapter sequences from Nanopore reads that may have been missed during basecalling and demultiplexing. Hybrid assembly with Illumina and Oxford Nanopore reads was performed with Unicycler (v0.4.8, default parameters) (86). Assembly statistics were recorded with QUAST (v5.0.2, default parameters) (87) and annotation was performed with Prokka (v1.14.5, default parameters + "--rfam") (88)p. For WT dupl+, WT dupl−, Δ*rbpB* dupl+, and Δ*rbpB* dupl− genomes, sample libraries for Illumina sequencing were prepared using the tagmentation-based and PCR-based Illumina DNA Prep kit and custom IDT 10 bp unique dual indices with a target insert size of 280 bp. Illumina sequencing was performed on an Illumina NovaSeq X Plus sequencer producing 2 × 151 bp paired-end reads. Demultiplexing, quality control, and adapter trimming were performed with bcl-convert (v4.2.4). For Nanopore sequencing, sample libraries were prepared using the PCR-free Oxford Nanopore Technologies (ONT) Ligation Sequencing Kit (SQK-NBD114.24) with the NEBNext Companion Module (E7180L) to the manufacturer's specifications. Nanopore sequencing was performed on an Oxford Nanopore a MinION Mk1B sequencer or a GridION sequencer using R10.4.1 flow cells in one or more multiplexed shared-flow-cell runs. Run design utilized the 400 bps sequencing mode with a minimum read length of 200 bp. Guppy (v6.4.6) (84) was used for super-accurate basecalling (SUP), demultiplexing, and adapter removal. Porechop (v0.2.4, default parameters) (85) was used to trim residual adapter sequences from Nanopore reads that may have been missed during basecalling and demultiplexing. *De novo* genome assemblies were generated from the Oxford Nanopore Technologies (ONT) read data with Flye2 (v2.9.2, --asm-coverage 50, --genome-size 6000000) (89) under the nano-hq (ONT high-quality reads) model. Subsequent polishing used the Illumina read data with Pilon (v.1.24, default parameters) (90). To reduce erroneous assembly artifacts caused by low-quality nanopore reads, long-read contigs with an average short-read coverage of 15× or less were removed from the assembly. Assembled contigs were evaluated for circularization via circulator (v1.5.5) (91) using the ONT long reads. Assembly annotation was then performed with Bakta (v1.8.1) (92) using the Bakta v5 database. Finally, assembly statistics were recorded with QUAST (v5.2.0) (87). For the AMG mutants, short-read mapping for mutation detection was completed with breseq (v0.33.2, default parameters) (93).

DNA-seq processing statistics are summarized in Table S7.

RNA sequencing sample preparation, processing, and analysis

Strains were grown in triplicate in 5 mL TYG to stationary phase overnight and cultures equivalent to 2 OD_{600} units were washed twice by centrifuging at 5,000 × *g* for 10

min and resuspending in 1 mL minimal medium without a carbon source. Washed cells were inoculated 1:100 in a 5-mL minimal medium containing glucose or raffinose to a final concentration of 25 mM. The cells were grown to an OD_{600} of ~0.8 and total RNA was isolated from 5 mL cultures and quantified as described above. 1 µg of total RNA was then submitted to SeqCenter for rRNA depletion, library construction, and sequencing. Briefly, samples were DNAse treated with Invitrogen DNAse (RNAse free). Library preparation was performed using Illumina's Stranded Total RNA Prep Ligation with Ribo-Zero Plus kit and 10 bp unique dual indices. Sequencing was done on a NovaSeq 6000, producing paired-end 151 bp reads. Demultiplexing, quality control, and adapter trimming were performed with bcl-convert (v4.1.5). Fastq files were quality checked using FastQC (v0.11.9) (94) and aligned to the genome of *B. thetaiotaomicron* VPI-5482 type strain (ASM1106v1) using Bowtie2 (v2.5.0, default parameters) (95). Aligned reads were quantified using FeatureCounts (v2.0.3, paired-end, -s parameter set to 2) (96). Differential expression analysis was done in R (v4.3.2) using DESeq2 (v1.44.0) (97) with α (False Discovery Rate) set to 0.05 and cutoffs for differentially expressed genes set to the absolute value of log 2 fold change >1 and adjusted *P* value < 0.05.

RNA-seq processing statistics are summarized in Table S7.

ACKNOWLEDGMENTS

We thank the members of AB's thesis committee, James Imlay, Bill Metcalf, and Shannon Sirk, for their advice throughout the course of this project. We also thank the former and current Vanderpool and Slauch laboratory members and members of the Microbiome Metabolic Engineering theme for thought-provoking discussions. We are grateful to Dr. Sandy Westermann of Scigraphix for the graphic design. This work was funded by National Institutes of Health grant R01 R35 GM139557 to CKV. This research was also supported by the GEMS Biology Integration Institute, funded by the National Science Foundation DBI Biology Integration Institutes Program, Award #2022049 (CKV).

All growth curves were plotted, and statistical analyses were done using Prism (v10.0.3). Adobe Illustrator was used to make figures and models.

AUTHOR AFFILIATIONS

[1]Department of Microbiology, University of Illinois Urbana-Champaign, Urbana, Illinois, USA
[2]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[3]Department of Microbiology and Plant Pathology, University of California Riverside, Riverside, California, USA
[4]Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, Illinois, USA

AUTHOR ORCIDs

Anubhav Basu http://orcid.org/0009-0006-8701-2048
Amanda N. D. Adams http://orcid.org/0000-0003-2688-1849
Carin K. Vanderpool http://orcid.org/0000-0002-2850-5532

FUNDING

Funder	Grant(s)	Author(s)
HHS | NIH | National Institute of General Medical Sciences (NIGMS)	R35 GM139557	Anubhav Basu Amanda N. D. Adams Carin K. Vanderpool
National Science Foundation (NSF)	2022049	Carin K. Vanderpool

AUTHOR CONTRIBUTIONS

Anubhav Basu, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing | Amanda N. D. Adams, Conceptualization, Data curation, Investigation, Methodology, Writing – review and editing | Patrick H. Degnan, Formal analysis, Investigation, Writing – review and editing | Carin K. Vanderpool, Conceptualization, Data curation, Formal analysis, Funding acquisition, Supervision, Writing – review and editing

DATA AVAILABILITY

DNA and RNA sequencing data sets from this manuscript are publicly available from NCBI under BioProject accession number PRJNA1107238.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental figures and legends (JB00235-24-s0001.docx). Fig. S1-S8.
Supplemental tables (JB00235-24-s0002.xlsx). Tables S1-S7.

REFERENCES

1. McNeil NI. 1984. The contribution of the large intestine to energy supplies in man. Am J Clin Nutr 39:338–342. https://doi.org/10.1093/ajcn/39.2.338
2. LeBlanc JG, Milani C, de Giori GS, Sesma F, van Sinderen D, Ventura M. 2013. Bacteria as vitamin suppliers to their host: a gut microbiota perspective. Curr Opin Biotechnol 24:160–168. https://doi.org/10.1016/j.copbio.2012.08.005
3. Degnan PH, Taga ME, Goodman AL. 2014. Vitamin B12 as a modulator of gut microbial ecology. Cell Metab 20:769–778. https://doi.org/10.1016/j.cmet.2014.10.002
4. Caballero-Flores G, Pickard JM, Núñez G. 2023. Microbiota-mediated colonization resistance: mechanisms and regulation. Nat Rev Microbiol 21:347–360. https://doi.org/10.1038/s41579-022-00833-7
5. Karczewski J, Troost FJ, Konings I, Dekker J, Kleerebezem M, Brummer R-J, Wells JM. 2010. Regulation of human epithelial tight junction proteins by *Lactobacillus plantarum in vivo* and protective effects on the epithelial barrier. Am J Physiol Gastrointest Liver Physiol 298:G851–G859. https://doi.org/10.1152/ajpgi.00327.2009
6. Zheng D, Liwinski T, Elinav E. 2020. Interaction between microbiota and immunity in health and disease. Cell Res 30:492–506. https://doi.org/10.1038/s41422-020-0332-7
7. Round JL, Mazmanian SK. 2009. The gut microbiota shapes intestinal immune responses during health and disease. Nat Rev Immunol 9:313–323. https://doi.org/10.1038/nri2515
8. Morais LH, Schreiber HL, Mazmanian SK. 2021. The gut microbiota-brain axis in behaviour and brain disorders. Nat Rev Microbiol 19:241–255. https://doi.org/10.1038/s41579-020-00460-0
9. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, Creasy HH, Earl AM, FitzGerald MG, Fulton RS, et al. 2012. Structure, function and diversity of the healthy human microbiome. Nat New Biol 486:207–214. https://doi.org/10.1038/nature11234
10. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, et al. 2011. Enterotypes of the human gut microbiome. Nat New Biol 473:174–180. https://doi.org/10.1038/nature09944
11. Salyers AA, Vercellotti JR, West SE, Wilkins TD. 1977. Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. Appl Environ Microbiol 33:319–322. https://doi.org/10.1128/aem.33.2.319-322.1977
12. Wexler AG, Goodman AL. 2017. An insider's perspective: *Bacteroides* as a window into the microbiome. Nat Microbiol 2:17026. https://doi.org/10.1038/nmicrobiol.2017.26
13. Glowacki RWP, Martens EC. 2021. If you eat it or secrete it, they will grow: the expanding list of nutrients utilized by human gut bacteria. J Bacteriol 203:e00481–20. https://doi.org/10.1128/JB.00481-20
14. Schwalm ND, Groisman EA. 2017. Navigating the gut buffet: control of polysaccharide utilization in *Bacteroides* spp. Trends Microbiol 25:1005–1015. https://doi.org/10.1016/j.tim.2017.06.009
15. Hao Z, Wang X, Yang H, Tu T, Zhang J, Luo H, Huang H, Su X. 2021. PUL-mediated plant cell wall polysaccharide utilization in the gut *Bacteroidetes*. IJMS 22:3077. https://doi.org/10.3390/ijms22063077
16. Anderson KL, Salyers AA. 1989. Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. J Bacteriol 171:3192–3198. https://doi.org/10.1128/jb.171.6.3192-3198.1989
17. Anderson KL, Salyers AA. 1989. Genetic evidence that outer membrane binding of starch is required for starch utilization by *Bacteroides thetaiotaomicron*. J Bacteriol 171:3199–3204. https://doi.org/10.1128/jb.171.6.3199-3204.1989
18. Terrapon N, Lombard V, Gilbert HJ, Henrissat B. 2015. Automatic prediction of polysaccharide utilization loci in *Bacteroidetes species*. Bioinformatics 31:647–655. https://doi.org/10.1093/bioinformatics/btu716
19. Koropatkin NM, Cameron EA, Martens EC. 2012. How glycan metabolism shapes the human gut microbiota. Nat Rev Microbiol 10:323–335. https://doi.org/10.1038/nrmicro2746
20. Foley MH, Cockburn DW, Koropatkin NM. 2016. The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. Cell Mol Life Sci 73:2603–2617. https://doi.org/10.1007/s00018-016-2242-x
21. Schwalm N 3rd, Groisman EA. 2017. Navigating the gut buffet: control of polysaccharide utilization in *Bacteroides* spp. Trends Microbiol 25:1005–1015. https://doi.org/10.1016/j.tim.2017.06.009
22. Ravcheev DA, Godzik A, Osterman AL, Rodionov DA. 2013. Polysaccharides utilization in human gut bacterium *Bacteroides thetaiotaomicron*: comparative genomics reconstruction of metabolic and regulatory networks. BMC Genomics 14:873. https://doi.org/10.1186/1471-2164-14-873
23. Lowe EC, Baslé A, Czjzek M, Firbank SJ, Bolam DN. 2012. A scissor blade-like closing mechanism implicated in transmembrane signaling in a *Bacteroides* hybrid two-component system. Proc Natl Acad Sci U S A 109:7298–7303. https://doi.org/10.1073/pnas.1200479109
24. Townsend GE, Raghavan V, Zwir I, Groisman EA. 2013. Intramolecular arrangement of sensor and regulator overcomes relaxed specificity in

hybrid two-component systems. Proc Natl Acad Sci U S A 110:E161–E169. https://doi.org/10.1073/pnas.1212102110
25. Lynch JB, Sonnenburg JL. 2012. Prioritization of a plant polysaccharide over a mucus carbohydrate is enforced by a *Bacteroides* hybrid two-component system. Mol Microbiol 85:478–491. https://doi.org/10.1111/j.1365-2958.2012.08123.x
26. Schwalm ND, Townsend GE, Groisman EA. 2017. Prioritization of polysaccharide utilization and control of regulator activation in *Bacteroides thetaiotaomicron*. Mol Microbiol 104:32–45. https://doi.org/10.1111/mmi.13609
27. Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW, Henrissat B, Gilbert HJ, Bolam DN, Gordon JI. 2011. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol 9:e1001221. https://doi.org/10.1371/journal.pbio.1001221
28. Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, Sonnenburg JL. 2010. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. Cell 141:1241–1252. https://doi.org/10.1016/j.cell.2010.05.005
29. Martens EC, Chiang HC, Gordon JI. 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe 4:447–457. https://doi.org/10.1016/j.chom.2008.09.007
30. Pollet RM, Martin LM, Koropatkin NM. 2021. TonB-dependent transporters in the *Bacteroidetes*: unique domain structures and potential functions. Mol Microbiol 115:490–501. https://doi.org/10.1111/mmi.14683
31. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI. 2003. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. Science 299:2074–2076. https://doi.org/10.1126/science.1080029
32. Terrapon N, Lombard V, Drula É, Lapébie P, Al-Masaudi S, Gilbert HJ, Henrissat B. 2018. PULDB: the expanded database of polysaccharide utilization loci. Nucleic Acids Res 46:D677–D683. https://doi.org/10.1093/nar/gkx1022
33. 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 New Biol 517:165–169. https://doi.org/10.1038/nature13995
34. Schwalm ND, Townsend GE, Groisman EA. 2016. Multiple signals govern utilization of a polysaccharide in the gut bacterium *Bacteroides thetaiotaomicron*. MBio 7:e01342-16. https://doi.org/10.1128/mBio.01342-16
35. Ndeh D, Baslé A, Strahl H, Yates EA, McClurgg UL, Henrissat B, Terrapon N, Cartmell A. 2020. Metabolism of multiple glycosaminoglycans by *Bacteroides thetaiotaomicron* is orchestrated by a versatile core genetic locus. Nat Commun 11:646. https://doi.org/10.1038/s41467-020-14509-4
36. Ndeh D, Rogowski A, Cartmell A, Luis AS, Baslé A, Gray J, Venditto I, Briggs J, Zhang X, Labourel A, Terrapon N, Buffetto F, Nepogodiev S, Xiao Y, Field RA, Zhu Y, O'Neill MA, Urbanowicz BR, York WS, Davies GJ, Abbott DW, Ralet M-C, Martens EC, Henrissat B, Gilbert HJ. 2017. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. Nature New Biol 544:65–70. https://doi.org/10.1038/nature21725
37. Elango D, Rajendran K, Van der Laan L, Sebastiar S, Raigne J, Thaiparambil NA, El Haddad N, Raja B, Wang W, Ferela A, Chiteri KO, Thudi M, Varshney RK, Chopra S, Singh A, Singh AK. 2022. Raffinose family oligosaccharides: friend or foe for human and plant health? Front Plant Sci 13:829118. https://doi.org/10.3389/fpls.2022.829118
38. Jukanti AK, Gaur PM, Gowda CLL, Chibbar RN. 2012. Nutritional quality and health benefits of chickpea (*Cicer arietinum* L.): a review. Br J Nutr 108 Suppl 1:S11–26. https://doi.org/10.1017/S0007114512000797
39. Andersen KE, Bjergegaard C, Møller P, Sørensen JC, Sørensen H. 2005. Compositional variations for alpha-galactosides in different species of *Leguminosae*, *Brassicaceae*, and barley: a chemotaxonomic study based on chemometrics and high-performance capillary electrophoresis. J Agric Food Chem 53:5809–5817. https://doi.org/10.1021/jf040471v
40. Beebe DU, Turgeon R. 1992. Localization of galactinol, raffinose, and stachyose synthesis in *Cucurbita pepo* leaves. Planta 188:354–361. https://doi.org/10.1007/BF00192802
41. Van Bel AJE. 1993. Strategies of phloem loading. Annu Rev Plant Physiol Plant Mol Biol 44:253–281. https://doi.org/10.1146/annurev.pp.44.060193.001345
42. Turgeon R. 1996. Phloem loading and plasmodesmata. Trends Plant Sci 1:418–423. https://doi.org/10.1016/S1360-1385(96)10045-5
43. Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi‐Shinozaki K, Shinozaki K. 2002. Important roles of drought‐ and cold‐inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana* . Plant J 29:417–426. https://doi.org/10.1046/j.0960-7412.2001.01227.x
44. Panikulangara TJ, Eggers-Schumacher G, Wunderlich M, Stransky H, Schöffl F. 2004. Galactinol synthase1. A novel heat shock factor target gene responsible for heat-induced synthesis of raffinose family oligosaccharides in *Arabidopsis*. Plant Physiol 136:3148–3158. https://doi.org/10.1104/pp.104.042606
45. Zhuo C, Wang T, Lu S, Zhao Y, Li X, Guo Z. 2013. A cold responsive galactinol synthase gene from *Medicago falcata* (MfGolS1) is induced by myo-inositol and confers multiple tolerances to abiotic stresses. Physiol Plant 149:67–78. https://doi.org/10.1111/ppl.12019
46. Zhao T-Y, Corum III JW, Mullen J, Meeley RB, Helentjaris T, Martin D, Downie B. 2006. An alkaline α-galactosidase transcript is present in maize seeds and cultured embryo cells, and accumulates during stress. Seed Sci Res 16:107–121. https://doi.org/10.1079/SSR2006243
47. Blackman SA, Obendorf RL, Leopold AC. 1992. Maturation proteins and sugars in desiccation tolerance of developing soybean seeds. Plant Physiol 100:225–230. https://doi.org/10.1104/pp.100.1.225
48. Corbineau F, Picard MA, Fougereux J-A, Ladonne F, Côme D. 2000. Effects of dehydration conditions on desiccation tolerance of developing pea seeds as related to oligosaccharide content and cell membrane properties. Seed Sci Res 10:329–339. https://doi.org/10.1017/S0960258500000374
49. Angelovici R, Galili G, Fernie AR, Fait A. 2010. Seed desiccation: a bridge between maturation and germination. Trends Plant Sci 15:211–218. https://doi.org/10.1016/j.tplants.2010.01.003
50. Han IH, Baik B-K. 2006. Oligosaccharide content and composition of legumes and their reduction by soaking, cooking, ultrasound, and high hydrostatic pressure. Cereal Chem 83:428–433. https://doi.org/10.1094/CC-83-0428
51. Kanwal F, Ren D, Kanwal W, Ding M, Su J, Shang X. 2023. The potential role of nondigestible Raffinose family oligosaccharides as prebiotics. Glycobiology 33:274–288. https://doi.org/10.1093/glycob/cwad015
52. Zartl B, Silberbauer K, Loeppert R, Viernstein H, Praznik W, Mueller M. 2018. Fermentation of non-digestible raffinose family oligosaccharides and galactomannans by probiotics. Food Funct 9:1638–1646. https://doi.org/10.1039/c7fo01887h
53. Andersen JM, Barrangou R, Abou Hachem M, Lahtinen SJ, Goh YJ, Svensson B, Klaenhammer TR. 2013. Transcriptional analysis of oligosaccharide utilization by *Bifidobacterium lactis* Bl-04. BMC Genomics 14:312. https://doi.org/10.1186/1471-2164-14-312
54. Amorim C, Silvério SC, Cardoso BB, Alves JI, Pereira MA, Rodrigues LR. 2020. *In vitro* fermentation of raffinose to unravel its potential as prebiotic ingredient. LWT 126:109322. https://doi.org/10.1016/j.lwt.2020.109322
55. Morabbi Heravi K, Watzlawick H, Altenbuchner J. 2019. The *melREDCA* operon encodes a utilization system for the Raffinose family of oligosaccharides in *Bacillus subtilis*. J Bacteriol 201:e00109-19. https://doi.org/10.1128/JB.00109-19
56. Hobbs JK, Meier EPW, Pluvinage B, Mey MA, Boraston AB. 2019. Molecular analysis of an enigmatic *Streptococcus pneumoniae* virulence factor: the raffinose-family oligosaccharide utilization system. J Biol Chem 294:17197–17208. https://doi.org/10.1074/jbc.RA119.010280
57. Ejby M, Fredslund F, Andersen JM, Vujičić Žagar A, Henriksen JR, Andersen TL, Svensson B, Slotboom DJ, Abou Hachem M. 2016. An ATP binding cassette transporter mediates the uptake of α-(1,6)-linked dietary oligosaccharides in *Bifidobacterium* and correlates with competitive growth on these substrates. J Biol Chem 291:20220–20231. https://doi.org/10.1074/jbc.M116.746529
58. Muiznieks I, Schmitt R. 1994. Role of two operators in regulating the plasmid-borne raf operon of *Escherichia coli*. Mol Gen Genet 242:90–99. https://doi.org/10.1007/BF00277352

59. Zhang X, Vrijenhoek JEP, Bonten MJM, Willems RJL, van Schaik W. 2011. A genetic element present on megaplasmids allows *Enterococcus faecium* to use raffinose as carbon source. Environ Microbiol 13:518–528. https://doi.org/10.1111/j.1462-2920.2010.02355.x
60. Adams AND, Azam MS, Costliow ZA, Ma X, Degnan PH, Vanderpool CK. 2021. A novel family of RNA-binding proteins regulate polysaccharide metabolism in *Bacteroides thetaiotaomicron*. J Bacteriol 203:e0021721. https://doi.org/10.1128/JB.00217-21
61. Gloster TM, Turkenburg JP, Potts JR, Henrissat B, Davies GJ. 2008. Divergence of catalytic mechanism within a glycosidase family provides insight into evolution of carbohydrate metabolism by human gut flora. Chem Biol 15:1058–1067. https://doi.org/10.1016/j.chembiol.2008.09.005
62. Okuyama M, Kitamura M, Hondoh H, Kang M-S, Mori H, Kimura A, Tanaka I, Yao M. 2009. Catalytic mechanism of retaining alpha-galactosidase belonging to glycoside hydrolase family 97. J Mol Biol 392:1232–1241. https://doi.org/10.1016/j.jmb.2009.07.068
63. Ryan D, Jenniches L, Reichardt S, Barquist L, Westermann AJ. 2020. A high-resolution transcriptome map identifies small RNA regulation of metabolism in the gut microbe *Bacteroides thetaiotaomicron*. Nat Commun 11:3557. https://doi.org/10.1038/s41467-020-17348-5
64. Kennedy MS, Zhang M, DeLeon O, Bissell J, Trigodet F, Lolans K, Temelkova S, Carroll KT, Fiebig A, Deutschbauer A, Sidebottom AM, Lake J, Henry C, Rice PA, Bergelson J, Chang EB. 2023. Dynamic genetic adaptation of *Bacteroides thetaiotaomicron* during murine gut colonization. Cell Rep 42:113009. https://doi.org/10.1016/j.celrep.2023.113009
65. Liu H, Shiver AL, Price MN, Carlson HK, Trotter VV, Chen Y, Escalante V, Ray J, Hern KE, Petzold CJ, Turnbaugh PJ, Huang KC, Arkin AP, Deutschbauer AM. 2021. Functional genetics of human gut commensal *Bacteroides thetaiotaomicron* reveals metabolic requirements for growth across environments. Cell Rep 34:108789. https://doi.org/10.1016/j.celrep.2021.108789
66. Pearce VH, Groisman EA, Townsend GE II. 2023. Dietary sugars silence the master regulator of carbohydrate utilization in human gut *Bacteroides* species. Gut Microbes 15:2221484. https://doi.org/10.1080/19490976.2023.2221484
67. Townsend GE, Han W, Schwalm ND, Hong X, Bencivenga-Barry NA, Goodman AL, Groisman EA. 2020. A master regulator of *Bacteroides thetaiotaomicron* gut colonization controls carbohydrate utilization and an alternative protein synthesis factor. MBio 11:e03221-19. https://doi.org/10.1128/mBio.03221-19
68. Belorkar SA, Gupta AK. 2016. Oligosaccharides: a boon from nature's desk. AMB Express 6:82. https://doi.org/10.1186/s13568-016-0253-5
69. Lee DH, Seong H, Chang D, Gupta VK, Kim J, Cheon S, Kim G, Sung J, Han NS. 2023. Evaluating the prebiotic effect of oligosaccharides on gut microbiome wellness using *in vitro* fecal fermentation. NPJ Sci Food 7:18. https://doi.org/10.1038/s41538-023-00195-1
70. Cheong K-L, Chen S, Teng B, Veeraperumal S, Zhong S, Tan K. 2023. Oligosaccharides as potential regulators of gut microbiota and intestinal health in post-COVID-19 management. Pharmaceuticals (Basel) 16:860. https://doi.org/10.3390/ph16060860
71. Marcobal A, Barboza M, Sonnenburg ED, Pudlo N, Martens EC, Desai P, Lebrilla CB, Weimer BC, Mills DA, German JB, Sonnenburg JL. 2011. Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. Cell Host Microbe 10:507–514. https://doi.org/10.1016/j.chom.2011.10.007
72. Lammerts van Bueren A, Mulder M, Leeuwen S van, Dijkhuizen L. 2017. Prebiotic galactooligosaccharides activate mucin and pectic galactan utilization pathways in the human gut symbiont *Bacteroides thetaiotaomicron*. Sci Rep 7:40478. https://doi.org/10.1038/srep40478
73. Brown HA, Koropatkin NM. 2021. Host glycan utilization within the *Bacteroidetes* Sus-like paradigm . Glycobiology 31:697–706. https://doi.org/10.1093/glycob/cwaa054
74. Singh RP. 2019. Glycan utilisation system in *Bacteroides* and *Bifidobacteria* and their roles in gut stability and health. Appl Microbiol Biotechnol 103:7287–7315. https://doi.org/10.1007/s00253-019-10012-z
75. Ryan D, Bornet E, Prezza G, Alampalli SV, Franco de Carvalho T, Felchle H, Ebbecke T, Hayward RJ, Deutschbauer AM, Barquist L, Westermann AJ. 2024. An expanded transcriptome atlas for *Bacteroides thetaiotaomicron* reveals a small RNA that modulates tetracycline sensitivity. Nat Microbiol 9:1130–1144. https://doi.org/10.1038/s41564-024-01642-9
76. Holdeman LV. 1973. Anaerobe laboratory manual. Virginia Polytechnic Institute and State University. Anaerobe Laboratory.
77. Degnan PH, Barry NA, Mok KC, Taga ME, Goodman AL. 2014. Human gut microbes use multiple transporters to distinguish vitamin B_{12} analogs and compete in the gut. Cell Host Microbe 15:47–57. https://doi.org/10.1016/j.chom.2013.12.007
78. García-Bayona L, Comstock LE. 2019. Streamlined genetic manipulation of diverse *Bacteroides* and *Parabacteroides* isolates from the human gut microbiota. MBio 10:e01762-19. https://doi.org/10.1128/mBio.01762-19
79. Mimee M, Tucker AC, Voigt CA, Lu TK. 2015. Programming a human commensal bacterium, *Bacteroides thetaiotaomicron*, to sense and respond to stimuli in the murine gut microbiota. Cell Syst 1:62–71. https://doi.org/10.1016/j.cels.2015.06.001
80. Sprouffske K, Wagner A. 2016. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. BMC Bioinformatics 17:172. https://doi.org/10.1186/s12859-016-1016-7
81. Smug BJ, Opalek M, Necki M, Wloch - Salamon D. 2024. Microbial lag calculator: a shiny - based application and an R package for calculating the duration of microbial lag phase. Methods Ecol Evol 15:301–307. https://doi.org/10.1111/2041-210X.14269
82. 5′ RACE protocol using the template switching RT enzyme mix (NEB #M0466). 2024. NEB. Available from: https://www.neb.com/en-us/protocols/2019/05/09/5-race-protocol-using-the-template-switching-rt-enzyme-mix. Retrieved 6 May 2024.
83. BCL convert. 2024. Available from: https://support-docs.illumina.com/SW/BCL_Convert/Content/SW/FrontPages/BCL_Convert.htm. Retrieved 11 May 2024.
84. Guppy protocol - Guppy software overview. 2023. Oxf Nanopore Technol. Available from: https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revax_14dec2018. Retrieved 11 May 2024.
85. Wick R. 2024. rrwick/Porechop. C++.
86. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLOS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595
87. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086
88. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
89. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8
90. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963
91. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0
92. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. 2021. Bakta: rapid & standardized annotation of bacterial genomes via alignment-free sequence identification. Bioinformatics. https://doi.org/10.1101/2021.09.02.458689
93. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol 1151:165–188. https://doi.org/10.1007/978-1-4939-0554-6_12
94. Babraham bioinformatics - FastQC a quality control tool for high throughput sequence data. 2023. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc. Retrieved 11 May 2024.
95. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923
96. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.

Bioinformatics 30:923–930. https://doi.org/10.1093/bioinformatics/btt656
97. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8