Specificity of Polysaccharide Use in Intestinal Bacteroides Species Determines Diet-Induced Microbiota Alterations

Erica D. Sonnenburg,1,3 Hongjun Zheng,2,3 Payal Joglekar,1 Steven K. Higginbottom,1 Susan J. Firbank,2 David N. Bolam,2,4 and Justin L. Sonnenburg1,*

1Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA
2Institute for Cell and Molecular Biosciences, Newcastle University, Medical School, Newcastle upon Tyne, NE2 4HH, UK
3These authors contributed equally to this work
*Correspondence: d.n.bolam@ncl.ac.uk (D.N.B.), jsonnenburg@stanford.edu (J.L.S.)
DOI 10.1016/j.cell.2010.05.005

SUMMARY

The intestinal microbiota impacts many facets of human health and is associated with human diseases. Diet impacts microbiota composition, yet mechanisms that link dietary changes to microbiota alterations remain ill-defined. Here we elucidate the basis of Bacteroides proliferation in response to fructans, a class of fructose-based dietary polysaccharides. Structural and genetic analysis disclosed a fructose-binding, hybrid two-component signaling sensor that controls the fructan utilization locus in Bacteroides thetaiotaomicron. Gene content of this locus differs among Bacteroides species and dictates the specificity and breadth of utilizable fructans. BT1760, an extracellular β2-6 endo-fructanase, distinguishes B. thetaiotaomicron genetically and functionally, and enables the use of the β2-6-linked fructan levan. The genetic and functional differences between Bacteroides species are predictive of in vivo competitiveness in the presence of dietary fructans. Gene sequences that distinguish species’ metabolic capacity serve as potential biomarkers in microbiomic datasets to enable rational manipulation of the microbiota via diet.

INTRODUCTION

The trillions of microbial cells that reside within the intestine shape aspects of host metabolism and immune function and extend the physiological definition of humans (Backhed et al., 2005; Hooper, 2009; Louis et al., 2007). While the general composition of the intestinal microbiota is similar in most healthy people, with greater than 90% of the cells belonging to the Firmicutes or Bacteroidetes phyla (Dethlefsen et al., 2008), the species composition is highly personalized (Turnbaugh et al., 2009).

Community membership and function of the microbiota can change due to numerous variables including antibiotic treatment, inflammation, or changes in diet (Dethlefsen et al., 2008; Frank et al., 2007; Jernberg et al., 2007; Ley et al., 2006). Perturbed loss of the typical composition has been associated with several disorders including inflammatory bowel diseases (Frank et al., 2007). In addition, changes in composition have been associated with obesity and weight loss; however, factors that cause these changes are not well defined (Duncan et al., 2008; Ley et al., 2006). The alterations in community membership, whether chronic or short-term, are accompanied by changes in the microbiota’s collective genome, or microbiome, and the patterns and metabolic capabilities it specifies (Turnbaugh et al., 2009). Therefore, the mechanisms that link relevant variables, such as changes in diet, to changes in the microbiome, are integral to understanding how environmental factors and behavior influence human biology.

Many complex plant polysaccharides in the human diet are resistant to host-mediated degradation due to either insolubility or lack of human-encoded hydrolytic enzymes (Flint et al., 2008; Louis et al., 2007; Sonnenburg et al., 2005). These carbohydrates are not absorbed in the upper gastrointestinal tract and serve as a major source of carbon and energy for the distal gut microbial community. Polysaccharide degradation is one of the core functions encoded in the microbiome (Lozupone et al., 2008; Turnbaugh et al., 2007). Broad expansion of the genes and operons dedicated to degrading and consuming polysaccharides has occurred within the genomes of microbiota-resident species (Xu et al., 2003, 2007), a logical outcome of the intense competition for these resources. It is, therefore, expected that alterations in the type and quantity of polysaccharides consumed can result in changes in the microbiota community composition and function.

Inulin- and levan-type fructans (homopolymers of β2-1 or β2-6 fructose units, respectively) are common dietary plant polysaccharides that feed the intestinal microbiota (Roberfroid et al., 1993). Multiple bacterial taxa in the gut utilize fructans, including members of Firmicutes, Bacteroides, and Bifidobacterium, (Duncan et al., 2003; Rossi et al., 2005; Van der Meulen et al., 2006), and dietary fructan can result in expansion of Actinobacteria, Firmicutes, or Bacteroides (Kolida et al., 2007; Menne et al., 2000; Ramirez-Farias et al., 2008). Lack of predictability in how the microbiota responds to such dietary interventions reflects our limited understanding of nutrient sensing and utilization by members of the intestinal microbiota.
Bacteroides, a major genera in the human microbiota, have a widely expanded capacity to use diverse types of dietary polysaccharides (Xu et al., 2007). Much of the glycan degrading and import machinery within Bacteroides genomes are encoded within clusters of coregulated genes known as polysaccharide utilization loci (PULs). *B. thetaiotaomicron* (Bt), a prototypic member of the Bacteroides, possesses 88 PULs, which differ in polysaccharide specificity (Martens et al., 2008). The defining characteristic of a PUL is the presence of a pair of genes homologous to *Bt* susD and susC, which encode outer membrane proteins that bind and import starch oligosaccharides, respectively (Figure 1A) (Martens et al., 2009; Shipman et al., 2000). The pair of susC and susD homologs is usually associated with genes that encode the machinery necessary to convert extracellular polysaccharides into intracellular monosaccharides, such as glycoside hydrolases (susA, susB, and susG in Figure 1A).

In addition to machinery for polysaccharide acquisition, most Bt genomes encode eight open reading frames on the negative strand of the *Bt* genome, including one example of a well-defined ligand for a member of this class of sensor regulators.

The fructan PUL is conserved to varying extents among Bacteroides species, corresponding to a range of fructan utilization capability across the genus. Using model intestinal microorganisms living within gnotobiotic mice, we demonstrate that dietary fructan can have disparate effects on community composition, depending upon the fructan degrading capacity of members of the microbiota. These studies suggest that within personal microbiomic datasets, we will be able to identify genetic biomarkers of discrete functions. Inference of function from these biomarkers should provide predictive power in determining how an individual's microbiota will respond to changes in diet and other interventions.

**RESULTS**

**BT1757-BT1763 and BT1765 Form a Putative Polysaccharide Utilization Locus That Is Transcribed Early in Bt’s Growth in Rich Media**

*Bt*'s genome encodes eight open reading frames on the negative strand of the *Bt* genome, including one...
susC/susD homolog pair (BT1763 and BT1762), a putative outer membrane lipoprotein (BT1761), a putative inner membrane monosaccharide importer (BT1758), a putative fructokinase (BT1757), and three putative glycoside hydrolases (BT1759, BT1760, BT1765) (Figure 1A). These glycoside hydrolases are members of Glycoside Hydrolase Family 32 (GH32), a family of enzymes specific for fructans (Cantarel et al., 2009). One of these, BT1760, possesses a N-terminal lipidation motif and is predicted to reside on the cell surface; the other two, BT1759 and BT1765, are predicted to be periplasmic and intracellular, respectively (www.cbs.dtu.dk/services/LipoP/ and www.cbs.dtu.dk/services/SignalP/). Directly adjacent to the locus is a putative inner membrane-associated sensor regulator of the HTCS family, BT1754. These data suggest that this PUL encodes the proteins required for Bt’s use of fructans.

Expression profiling of Bt in rich medium has revealed the upregulation of several PULs, each of which is confined to a discrete phase of growth (Sonnenburg et al., 2006). Analysis of Bt transcriptional profiles at five time points that spanned from early log to stationary phase in vitro in rich medium, compared to basal expression in minimal medium (MM) containing glucose as the sole carbohydrate, revealed that 14 pairs of susC/susD homologs were induced greater than 20-fold at one or more time points during the growth (Figure 1B) (Gene Expression Omnibus database, www.ncbi.nlm.nih.gov/geo/; accession numbers, GSM 40897–40926). The putative fructan PUL showed upregulation early in Bt’s growth suggesting it is responsive to a high priority substrate accessible early in growth on rich medium (Figure 1B). Genes within this PUL are coexpressed both in vitro in rich medium and in vivo in Bt mono-associated gnotobiotic mice fed a polysaccharide-rich diet (Figure S1A available online), consistent with the functional relatedness of adjacent genes and operon predictions in Bt (Westover et al., 2005). Bt increases expression of this PUL in vivo while downregulating the vast majority of other PULs when bi-associated in the gnotobiotic mouse intestine with the methanogenic archean, Methanobrevibacter smithii (Samuel and Gordon, 2006). The upregulation of the putative fructan PUL is concomitant with increased densities of Bt in vivo, suggesting that expression of this locus is associated with growth potentiation of Bt.

**Bt Upregulates Its Putative Fructan PUL When Grown on Fructose-Containing Carbohydrates**

We inoculated minimal medium containing specific fructose-based carbohydrates as the only carbon and energy source with Bt to test if the bacterium is competent to grow on fructans. Bt grew on a broad range of fructose-based glycans, including free fructose, sucrose, levan (high MW fructose polymer with predominantly β2-6-linkages), and fructo-oligosaccharides (FOS; short-chain β2-1 polymers of 2–10 fructose units) (Figure 1C; see Figure S2 for carbohydrate structures). However, Bt grew poorly on inulin (β2-1 fructose polymer with an average degree of polymerization of ~25), with growth only apparent three days after inoculation. Doubling times on simple monosaccharides and disaccharide were similar to one another (Table S1). In contrast, growth rates of Bt between the different fructans showed large linkage-dependent differences: β2-6 levan resulted in the fastest doubling time (2.7 hr), while β2-1 FOS and inulin were significantly slower (doubling times of 5.6 hr and 96.4 hr, respectively) (Table S1).

To determine whether these fructose-based substrates induced expression of genes associated with the putative fructan PUL, Bt was grown in either glucose or one of five fructose-containing substrates (fructose, sucrose, levan, FOS, or inulin) as the sole carbohydrate. Cells were harvested at mid-log phase for quantitative RT-PCR (qPCR) analysis, and RNA levels of the 3’ and the 5’ ends of the operon, BT1757 (encoding the fructokinase) and BT1763 (encoding the SusC-like protein), respectively, were used as an indicator of PUL expression (Figure 1D). Both BT1757 and BT1763 were dramatically upregulated in all media containing fructose, whether as a free monosaccharide or in glycosidic linkage. Across all conditions, expression of BT1757, BT1763, and BT1765 showed coordinated increases consistent with the predicted operon structure. However, BT1754 (the PUL-associated putative HTCS) showed no significant induction under all conditions tested. Therefore, the operon that encodes the structural genes of Bt’s putative fructan PUL is transcriptionally responsive to fructose-containing carbohydrates. Published surveys of Bt gene expression in numerous carbohydrates support that upregulation of the fructan PUL is specific to fructose-containing substrates (Martens et al., 2009; Sonnenburg et al., 2005).

Two genes within Bt’s genome that are not physically associated with the putative fructan PUL, a second putative periplasmic GH32 (BT3082) and a second putative fructokinase (BT3305), were likely candidates to be involved in fructan utilization. Analysis of BT3082 and BT3305 expression by qPCR revealed that BT3082 was induced in all fructose-containing media and showed a pattern of induction consistent with those seen for BT1757, BT1763, and BT1765 (Figure 1D); however, BT3305 showed no change in expression or a slightly reduced expression in all conditions (data not shown). These data suggest that the fructosidase, BT3082, but not the putative fructokinase, BT3305, is part of the regulon of the fructan PUL.

**The Hybrid Two-Component System BT1754 Is Required for Efficient Fructan Utilization by Bt**

We assessed the ability of an isogenic mutant of Bt lacking the BT1754 gene to grow in a panel of fructose-based minimal media to test if upregulation of the PUL was dependent upon the HTCS signaling sensor. An in-frame, unmarked deletion of BT1754 was constructed using a standard counter-selectable allele-exchange procedure. Bt-ΔBT1754 exhibited normal colony morphology on solid medium and grew with a similar doubling time to wild-type in MM-glucose (2.6 hr); however, Bt-ΔBT1754 failed to grow in any of the three fructans (FOS, inulin, and levan) and showed retarded growth in fructose and sucrose (Figure 2A and Table S1). Additionally, Bt-ΔBT1754 does not exhibit prioritized upregulation of the putative fructan PUL during growth in rich media (Figure S1B). Complementation of this mutant was achieved by introducing the genomic fragment containing BT1754 and its 5’ intergenic upstream promoter region in trans. Growth of the ΔBT1754::BT1754 complemented mutant restored growth in all fructose-based media to levels comparable to wild-type (Figure 2A and Table S1). These data demonstrate the HTCS encoded by BT1754 is required for Bt’s use of fructans.
The Periplasmic Domain of the Hybrid Two-Component System BT1754 Binds to Monomeric Fructose

One of the key unanswered questions concerning the HTCS family, and many other extracellular sensory systems, is the identity of the molecular triggers for signaling events. The predicted inner-membrane localization of Bt’s HTCS family members, including BT1754, suggests that the periplasmic region likely serves as the sensor/receptor, similar to classic two-component systems. Analysis of the sequence of BT1754 revealed a typical HTCS architecture with an N-terminal predicted periplasmic sensor domain flanked by two transmembrane regions and a C-terminal cytoplasmic histidine kinase domain, a phosphoacceptor domain and a response regulator (including a receiver and an HTH_AraC-type DNA binding domain) (Figure 2B and Figure S3). Uniquely within Bt’s HTCS, the sensor domain displays homology to Type I bacterial periplasmic binding proteins (PBPs) (Dwyer and Hellinga, 2004). As PBPs are known to bind small molecules such as sugars, we expressed the periplasmic domain of BT1754 (BT1754-PD; residues 29–343) in a recombinant form and tested for binding to a range of monosaccharides and fructan-derived oligosaccharides to see if direct interaction with a specific carbohydrate is the means of signal perception in BT1754. The isothermal calorimetry data reveal that BT1754-PD binds specifically to fructose, with a Kd of ~2 µM and a stoichiometry of 1:1, but does not interact with either β2-1- or β2-6-linked fructooligosaccharides or any other monosaccharides, including glucose and ribose (Figure 2C).

Structure of BT1754 Periplasmic Sensor Domain

To understand the mechanism of signal perception in more detail, we determined the structure of BT1754-PD in complex with fructose to 2.66Å. The closest homolog with known structure, a ribose-binding PBP from Thermoanaerobacter tengcongensis (TtRBP), PDB 2IOY, was used as a molecular replacement search model. Successful molecular replacement resulted in a dimer in the asymmetric unit. A least-squares alignment of the final model with TtRBP gave a root mean square deviation of 1.2 Å for 269 alpha carbons despite the relatively low sequence identity, indicative of the high structural conservation of this family. The BT1754-PD structure comprises a typical two-subdomain PBP-fold, with each subdomain consisting of a core of six α helices (Figure 3A). The polypeptide chain forms a hinge by crossing between the two subdomains three times along one side, the last of these exiting the PBP-fold and then forming a long α-helix, which extends back along the length of the protein to the N-terminal region (Figure 3A).

The C-terminal helix of BT1754-PD provides the predominant interface for homo-dimerization and is the main structural difference between classical soluble PBPs such as the TtRBP and BT1754-PD (Figure 3B). Though there are several hydrogen bonds to retain the turn between the PBP-fold and the helix, once the polypeptide has progressed beyond the first residue of the helix (Asn306), the remainder of the contacts, both inter- and intramolecular, are nonpolar. The dimer, generating a buried surface area of 2640 Å², appears to be biologically relevant as both the N- and C-termini of each molecule are oriented such that they face in the same direction and, therefore, both molecules are positioned correctly for insertion into the membrane (Figure 3A).
Crystals were grown in the presence of fructose, and electron density indicative of a fructose molecule in the β-furanose form was observed in the cleft between the two subdomains, the typical binding site of PBP family proteins (Dwyer and Hellinga, 2004)(Figure 3A). The sugar ring is sandwiched between two tryptophan residues, one from each subdomain (Trp45 and Trp196), with Tyr271 and Pro168 also forming hydrophobic interactions of BT1754-PD and fructose. Fo-Fc electron density prior to modeling the single molecule of fructose in the β-furanose form is shown (blue mesh contoured at 3σ). (B) Overlay of BT1754-PD (green) with TrRBP (blue); the extended C-terminal helix in BT1754-PD (bracket) is unique to BT1754. (C) Side view of the binding site illustrating hydrophobic interactions of BT1754-PD and fructose. Fo-Fc electron density prior to modeling the single molecule of fructose in the β-furanose form is shown (blue mesh contoured at 3σ). (D) Top view of the binding site of BT1754-PD illustrating the numerous H-bonds (dotted black lines) with fructose.

To test whether the only putative cell surface GH32 in Bt, BT1760, is required for levans utilization, an in-frame, unmarked deletion of BT1760 was constructed. Bt-ΔBT1760 exhibited normal colony morphology on solid medium and grew with a normal doubling time in MM-glucose. Bt-ΔBT1760 did not grow on levans, but showed normal growth on all other media tested including β2-1-linked FOS (Figure 4A), with doubling times comparable to wild-type in fructose, sucrose, and FOS (Table S1). Complementation of this mutant was achieved by fusing the upstream intergenic promoter region of BT1760 to the 5′ end of the genomic fragment containing BT1760. Levans growth was restored, albeit at a reduced rate, in the complemented Bt-ΔBT1760::BT1760 strain (Figure 4A), confirming the requirement of this glycoside hydrolase for utilization of the β-2 linked fructan.

We next assessed whether BT1760 is a β2-6-specific fructanase. Activity of a recombinant form of BT1760 was tested against a range of β2-6 and β2-1 fructan oligo- and polysaccharides. The data show that BT1760 is indeed a β2-6-fructan specific enzyme with no detectable activity against β2-1 fructans or fructooligosaccharides (Table S2). TLC analysis of levans digestion by BT1760 revealed that a mixture of different sized oligosaccharides was produced. Mono-, di-, tri-, and tetra-levanoligosaccharides accumulated as the main products as the reaction proceeded (Figure 4B). These data demonstrate that BT1760 is a β2-6-specific endo-acting fructanase.

To determine whether the β2-6 fructoside hydrolase activity of BT1760 could be detected on the cell surface, we measured the activity of washed whole Bt cells against levans and inulin. Fructose-grown wild-type cells could degrade the β2-6 polymer but had no detectable activity against inulin, mirroring the specificity of recombinant BT1760 (Figure 4C and data not shown). This levanase activity was completely lost in the Bt-ΔBT1760 strain and was largely restored in the complemented Bt-ΔBT1760::BT1760 strain (Figure 4C). Moreover, cells grown on glucose displayed ~100-fold lower levans activity, confirming that the levans-specific hydrolysis is inducible by fructose (data not shown).
saccharides (Koropatkin et al., 2008, 2009). We tested whether binding to oligosaccharides suggests that linkage is important.

We constructed a mutant in which levan is the sole carbon source. We tested the ability of this mutant to grow in minimal media and we tested the ability of this mutant to grow in minimal media in which levan is the sole carbon source. BT1762 was the sole specificity determinant in the hydrolase serving as a key step for converting long-chain levan into oligosaccharides for SusC/SusD-homolog-mediated import.

Structural insight into the nature of SusD and a SusD homolog binding to oligosaccharides suggests that linkage is an important determinant in cell surface structural recognition of oligosaccharides (Koropatkin et al., 2008, 2009). We tested whether BT1760 was the sole specificity determinant in BT’s efficient use of levan, or whether the SusD homolog within the fructan PUL, BT1762, also exhibited specificity for the β2-6 linkage. We constructed a BT mutant in which BT1762 was deleted, and we tested the ability of this mutant to grow in minimal media in which levan is the sole carbon source. BT-ΔBT1762 showed significantly retarded growth on levan compared with wild-type and growth of this mutant in levan was largely restored upon BT1762 complementation. (Figure 5A and Table S1). Absence of BT1762, however, did not affect extracellular levan degradation, supporting that BT1760 is responsible for cell surface levan degradation (Figure 4C). To determine the specificity of BT1762 directly, the protein was expressed in a recombinant form lacking its signal peptide and lipidation motif, and its interaction with levan and inulin was assessed by isothermal calorimetry (Figure 5B). The data show that BT1762 binds to the β2-6 fructose polymer but displays no affinity for the β2-1 equivalent. BT1762 displays a Kd of ~40 μM for levan, similar to the affinity of the prototypic SusD for cyclodextrins (Koropatkin et al., 2008).

Recent studies have indicated that within many Bacteroides PULs, the gene found downstream of the susD homolog also encodes a polysaccharide-binding lipoprotein (Martens et al., 2009). Although the products of these “susE-positioned” genes have no obvious sequence homology to one another they appear to be functionally conserved. To explore the role of the susE-positioned gene from the BT fructan PUL, BT1761, we assessed the ability of a recombinant form of the protein to interact with inulin and levan. The data reveal that BT1761 bound specifically to levan (Figure S4). Reducing sugar and TLC assays with BT1761 and BT1762 against inulin and levan revealed that neither protein had any detectable degradative capacity (data not shown). Together, these genetic and biochemical data show that the cell surface components of BT’s fructan PUL exhibit β2-6 linkage specificity.

**BT Has Three GH32 Enzymes that Are Not Linkage Specific**

To understand the pattern of fructan degradation in BT in more detail we biochemically characterized the three other GH32s expressed during growth on fructose-containing media, the predicted periplasmic BT1759 and BT3082 and the predicted intracellular BT1765. The data revealed that all three of these enzymes are exo-acting fructosidases that release fructose from both β2-1 and β2-6 fructans, although some differences in their kinetic characteristics were observed (Figure 4B and Table S2). BT1759 and BT3082 act equally well on inulin and levan, as well as oligosaccharides of these polymers, although BT3082 appears to be overall a more efficient enzyme with ~2- to 4-fold higher kcat/Km values than BT1759 for most substrates, driven mainly by its higher turnover number. Considering the β2-6 fructan preference of BT, it is interesting that both enzymes display lower Km values (~2- to 8-fold) for β2-1 oligosaccharides compared to their β2-6 equivalents (Table S2).

BT1759 and BT3082 also cleave sucrose, a trait shared with other bacterial fructosidases, although both have a higher Km for the disaccharide than for larger β2-1 kesto-oligosaccharides. By contrast, BT1765 much prefers sucrose over any of the other oligo- or polysaccharides tested, although the enzyme is also
able to efficiently hydrolyse levanbiose (Table S2). The predicted cytoplasmic location of BT1765 and substrate specificity suggest that some of the disaccharide products of levan (and possibly FOS) digestion are transported across the inner membrane before they are degraded by the periplasmic fructosidases (see Figure S5).

The Fructan PUL Is Varially Conserved in Sequenced Bacteroides, which Have Differing Capacity to Utilize Fructan

We performed a comparative genomic analysis focused on Bt’s fructan utilization locus between five sequenced species of Bacteroides to gain further insight into the mechanism of fructan use for this major group of gut resident microbes. Using the N-terminal fructose-binding domain of the HTCS BT1754 to query a BLAST database consisting of the Bacteroides species B. caccae, B. vulgatus, B. uniformis, B. fragilis, and B. ovatus, we have identified a single orthologous HTCS in each species, with the exception of B. fragilis, which harbors two BT1754-like genes. Sequence identity between the periplasmic sensor domains of the BT1754 orthologs was high for all but one, ranging from 93% for the B. ovatus protein to 58% for the B. vulgatus domain. Furthermore, the residues involved in fructose binding in BT1754 are almost completely conserved among orthologs, consistent with conservation of the ligand sensed by each HTCS (Figure S3). The periplasmic domain of one of the two B. fragilis orthologs (BF4326) displayed only 36% identity with BT1754-PD, and this domain was unique in its lack of fully conserved fructose binding residues (Figure S3). Regions adjacent to the HTCS in each genome were analyzed and found to display local synteny with the Bt locus (Figure 6, left panel), including the presence of open reading frames that are predicted to play a role in utilization of fructose-containing carbohydrates. In all six Bacteroides species, the HTCS is adjacent to a predicted fructokinase, a putative inner membrane monosaccharide importer, and GH32-family glycoside hydrolase. In each genome, except that of B. vulgatus, the syntetic regions also contain a susC/susD homologous pair.

The presence of an apparent fructan PUL in multiple Bacteroides species suggested that fructan utilization is shared between members of this genus. Testing for growth on fructose-based glycans revealed that all six species are competent for growth on fructose (Figure 6, right panel), sucrose and FOS (Table S1). All Bacteroides species tested, except B. vulgatus, were able to grow efficiently using one of the long-chain fructans, inulin or levan. The inability of B. vulgatus to grow on long-chain fructans is consistent with the absence of a susC/susD-like pair within its locus. B. caccae, B. ovatus, B. fragilis and B. uniformis can utilize inulin with efficiency similar to their use of glucose. This contrasts with Bt inulin use, which is only observed after three days (Figure 6).

Bt is the only species tested able to use levan, which was particularly striking when considering the overall similarity in PUL structure between Bt, B. caccae, and B. ovatus. However, examination of PUL gene content of the two inulin-utilizing species revealed genes encoding PL19 enzymes, a family that is known to include members capable of degrading the j2-1 fructan. Additionally, Bt’s extracellular j2-1-specific GH32, BT1760, does not possess an orthologous gene in the other species (Figure S6). Notably, two other sequenced Bt strains utilize levan more efficiently than inulin in vitro (data not shown), similar to the type strain. Both of these strains possess orthologs to the type strain’s BT1760 (Figure S6). Together these data demonstrate that differences in fructan specificity of Bacteroides species correspond to differences in the gene content of their respective fructan PULs.
Genomic Content of Bacteroides Species Predicts Changes in Microbiota Composition Induced by an Inulin-Based Diet

The differences in ability to utilize fructans between the Bacteroides species implies that the relative success of a species within a gut ecosystem may be determined, in part, by the abundance and type of fructan in the host diet. Furthermore, the comparison of genomic sequences and differences in fructan use between species suggests that personalized predictions of microbiota response to specific dietary polysaccharides may be made based on metagenomic microbiome sequence data.

We constructed defined two-member communities of Bacteroides species within the intestines of gnotobiotic mice to test how model microbiotas respond in vivo to dietary inulin, which, unlike levan, is available in pure form in quantities sufficient to conduct such a study. Our in vivo experiment aimed to test how differing functionalities embedded within the genomes of two different two-species model microbiotas influence inulin-induced changes in community composition.

Due to B. caccae’s superior ability to use inulin compared to Bt, we tested whether B. caccae would become dominant over Bt within the intestines of gnotobiotic mice to test how model microbiotas respond in vivo to dietary inulin, which, unlike levan, is available in pure form in quantities sufficient to conduct such a study. Our in vivo experiment aimed to test how differing functionalities embedded within the genomes of two different two-species model microbiotas influence inulin-induced changes in community composition.

Each mouse was maintained on a standard polysaccharide-rich diet for the first 7 days of colonization and then switched to a diet in which the sole polysaccharide was inulin (10% w/w) for an additional 14 days (Figure 7A). Mice were individually housed throughout the experiment to ensure no cross inoculation could occur and bedding was changed every two days. Total bacterial colonization density was determined by assessing the CFUs in feces over 21 days. The change in each species’ relative abundance before and after dietary inulin supplementation was assessed using species-specific primers in a quantitative PCR assay.

Our results disclosed that total fecal bacterial densities over the course of the experiment did not differ significantly upon dietary shift (total densities ranged from $10^{10}$–$10^{11}$ bacteria/ml of fecal material). Relative densities were determined on days 4 and 6 (standard diet) and on days 13 and 21 (6 and 14 days after dietary switch). In the Bt/B. caccae, bi-associated mice, before the diet switch (day 6 postcolonization in mice fed a standard diet), Bt comprised 87 ± 3% of the community, indicating that Bt is better adapted than B. caccae to these in vivo conditions. Six days after a change to the inulin-based diet, Bt levels dropped to 80 ± 4%, and B. caccae increased to 20 ± 4%. After two weeks consuming the inulin diet, the relative proportion of the two species showed a more drastic shift in favor of B. caccae; Bt representation decreased to approximately 49 ± 6% versus 51 ± 6% B. caccae ($p = 8 \times 10^{-5}$, day 21 versus day 6; n = 7 mice; Student’s t test) (Figure 7B). In

![Figure 6. Comparative Genomic and Functional Analysis of Fructan Utilization among Bacteroides Species](image-url)

Fructan-utilization loci from Bacteroides species (left). Common predicted functions are color coded, intervening unrelated genes are white. PL19, polysaccharide lyase family 19; GH32, glycoside hydrolase family 32. Growth curves (right) of each Bacteroides species in fructose-based carbohydrates.
contrast, the Bt/B. vulgatus bi-associated mice did not exhibit any significant trend in changed community composition after 6 days on an inulin-based diet, but Bt increased in abundance from 74 ± 3% on day 6 to 84 ± 5% on day 21 (p = 0.1; n = 3 mice) on the inulin-enriched diet (Figure 7C). The delayed and modest effect of diet influencing the composition of the Bt/B. vulgatus bi-association is consistent with poor inulin use by Bt and no inulin use by B. vulgatus (Figure 6). Together, these data are consistent with dietary polysaccharide-induced changes in the microbiota composition that are predictable based on the resident species’ ability to use that polysaccharide.

In the previous experiment, inulin was the sole polysaccharide in the diet. We wondered whether we would observe the same inulin-induced increase in B. caccae relative to Bt if other polysaccharides were also present in the diet. To test this, gnotobiotic mice were co-colonized with Bt and B. caccae and maintained on the standard diet with inulin supplementation in the water (1% w/v). Over the 14 days the mice ingested an average of 117 ± 6 mg of inulin daily via the water (compared to 355 ± 7mg/day with the inulin diet). Fecal samples were tested by qPCR over the course of the 21-day experiment for relative levels of Bt or B. caccae. These data revealed no statistical difference in the change in relative colonization between mice fed inulin-supplemented water compared to controls that received the same standard diet for 21 days, but received no inulin (Figure 7D). These data suggest that when mice were fed a diet rich in carbohydrates, the presence of inulin did not provide enough of an advantage to B. caccae to allow it to out-compete Bt; however, the amount of inulin supplied in the water supplement inulin in the water to determine whether a lower dose of inulin in the absence of other polysaccharides was sufficient to provide B. caccae a competitive advantage over Bt in vivo. Under this experimental paradigm the mice consumed an average of 97mg of inulin per day. After 14 days on inulin-water supplementation, the proportion of B. caccae increased by 26 ± 8% (Figure 7D). While not as robust an increase as observed in the inulin-only diet experiment (which showed a 36 ± 7% increase in B. caccae), these data demonstrate that reduced inulin consumption in the absence of competing polysaccharides, offers a significant competitive advantage to inulin-utilizing B. caccae, consistent with the flexible nutrient foraging the Bacteroides species exhibit. The wide range of polysaccharides present in the standard diet allows Bt to compete effectively with B. caccae even in the presence of inulin.

We finally demonstrate the importance of inulin utilization for conferring a competitive advantage in hosts fed an inulin-rich diet using a genetic proof of this effect. The region of the B. caccae fructan-utilization locus from the susC-like gene through the GH32-encoding gene (BC02727-BC02731) was cloned and expressed in a strain of Bt that is compromised in its ability to utilize levans (Bt-ΔBT1763) under the control of the BT1763 promoter (data not shown). The resulting strain, Bt(In+) exhibits efficient growth in minimal medium containing inulin, similar to B. caccae (Figure S7). Repeating our original in vivo competition experiment with Bt(In+) revealed that conferring inulin use ability upon Bt eliminates the ability of B. caccae to become dominant in the presence of an inulin-based diet (Figure 7E). This result confirms that the specificity of dietary polysaccharide use is
the key functionality that dictates the alterations in the model microbiota that we observe. These results support our hypothesis that changes in microbiota community membership brought on by dietary change can be inferred based on genomic and functional knowledge of resident microbial populations. They also suggest that diet can be a dominant determinant in dictating changes in microbiota composition.

**DISCUSSION**

Inulin (\(\beta_2-1\) fructan) and levan (\(\beta_2-6\) fructan) are polysaccharides that are abundant in the human diet, but are resistant to host-mediated digestion in the upper gastrointestinal tract. These glycans instead serve as a carbon and energy source for the bacteria that reside in the distal intestine. *Bacteroides thetaiotaomicron*, a resident of the human GI tract, encodes a fructan utilization locus, BT1757-63 and BT1765, the gene products of which enable efficient acquisition and use of levan-type carbohydrates.

The fructan PUL is adjacent to a hybrid two-component system sensor-regulator, BT1754, which binds only to monomeric fructose, a signal sufficient to induce transcription of the locus. While the upregulation of polysaccharide utilization machinery in response to a monosaccharide may seem unexpected, this signal is a likely consequence of the environment in which *Bt* resides. Within the natural habitat of the large bowel, free fructose and simple disaccharides, such as sucrose, do not occur at appreciable levels as the host absorbs such sugars within the small intestine. Therefore, the regulation of this locus evolved in the absence of selective pressure to discriminate free monosaccharide from polysaccharides. In addition, unlike many other monosaccharides, fructose is found in only a single class of polysaccharide, namely homopolymeric fructans. *Bt* appears to use the liberated fructose as a proxy (i.e., indicator) for fructan, which results in upregulation of the machinery to utilize the polysaccharide. This is consistent with previous data that demonstrate *Bt*’s constitutive, low-level expression of PULs in conditions lacking the relevant substrates (Martens et al., 2009; Sonnenburg et al., 2005), as well as the low-level cell surface levanase activity we observe with whole cells grown in glucose. The constitutive expression suggests that *Bt* employs a strategy of being prepared to degrade multiple polysaccharides immediately upon their arrival into the distal gut environment. Specific liberated carbohydrates that result from the degradation serve as signals that augment expression of the appropriate PUL via a specific sensor-regulator such as a HTCS.

The binding of BT1754 to monomeric fructose also results in a failure of the sensor to differentiate \(\beta_2-1\) and \(\beta_2-6\) linkages despite *Bt* being much more efficient in use of the levan-type fructans. Specificity of signal is instead derived from the cell surface structural components of the PUL, which serve as the “gateway” for substrates crossing the outer membrane. The cell surface SusD homolog, BT1762, the *susE*-positioned gene product, BT1761, and the endo-levanase, BT1760, all contribute to the specific import of \(\beta_2-6\) fructans into *Bt*’s periplasm. BT1754 relies upon the specificity of the cell surface polysaccharide degradation and binding machinery to provide fructose derived from \(\beta_2-6\) fructan to the periplasm where the sensor is sequestered.

Despite *Bt*’s inability to utilize inulin efficiently it is able to grow well on FOS, a short chain \(\beta_2-1\) fructan. Notably, the fructan PUL of *Bt* is upregulated during growth in vitro in minimal medium containing FOS or inulin. *Bt*’s ability to grow in FOS at a rate that is significantly faster than inulin is likely due to the difference in degree of polymerization between the two substrates. Whether small oligosaccharides from these substrates undergo passive diffusion into the periplasm or are accessed via another mechanism requires further investigation.

Among the Bacteroides species tested, *Bt* appears to be unique in its ability to utilize levan, whereas other species are adept at utilizing polymeric \(\beta_2-1\) fructans. Such phenotypic differences, combined with dietary variation between individuals, could provide the basis for the striking person-to-person variability observed for Bacteroidetes in human microbiota enumeration studies (Eckburg et al., 2005). Our in vivo studies illustrate that species well-adapted to use inulin gain a competitive advantage when hosts are fed an inulin-based diet. Although a genetic loss-of function experiment, in which inulin use is compromised, could be used to test whether the observed changes in species abundance are due to inulin use, we have used a gain-of-function experiment, in which inulin use is conferred upon *Bt*, to illustrate this point unequivocally. These results suggest that some aspects of diet-induced changes in microbiota composition may be predetermined based on the intrinsic capacity of an individual species to use the substrates that are being consumed by the host. We speculate that diets enriched in different polysaccharides, or polysaccharide-deficient diets, could result in microbiotas of very different species composition. Future studies that follow species and gene composition of the human intestinal microbiota during consumption of levan- or inulin-based diets will provide insight into the rapidity with which members of a complex community adapt at a functional, compositional, and genetic level. How such niche specialization occurs over the course of evolution and the role that diet plays in determining a species’ glycan utilization repertoire remain important yet difficult questions to address.

**Perspective**

As the age of personal genomes approaches, some aspects of diet and medical therapies will be customized based on genotype. Diet can also be personalized to optimize microbiota function and interaction with the host based on the metagenomic analysis of an individual’s microbiota. A prerequisite for incorporating vast amounts of microbial genomic data into personalized, preventative analysis is to attain a mechanistic understanding of the most dominant aspects of microbiota function. Here we present a case study of how understanding the mechanisms that link the microbiome to microbiota function may enable individualized predictions of microbiota response to perturbations. We have taken two-species model microbiotas that collectively possess close to 10,000 genes and predicted how they will respond to a specific dietary cue based on a functional understanding of the \(~20\) relevant genes. A similar distillation of full microbiomic datasets that contain > \(10^6\) genes, to a relevant subset, will be required to make microbiota management
tractable. With an ever-increasing understanding of how the biology of host and microbiota integrate, we may soon be able to use genomic and microbiomic sequence data to intentionally program or reprogram the emergent properties of the host-microbial superorganism.

EXPERIMENTAL PROCEDURES

Culturing Bacteria
Bacteria were cultured in TYG and MM as described previously (Martens et al., 2008; Sonnenburg et al., 2005). The following bacteria were used: Bt (VPI-5482), B. caccae (ATCC-43185), B. ovatus (ATCC-8483), B. fragilis (NCTC-9343), B. uniformis (ATCC-8492), and B. vulgatus (ATCC-8482). Growth curves in MM were obtained using a Powerwave (Biotek) reading OD600 every 30 min from anaerobic cultures at 37°C.

Quantitative RT-PCR Analysis
Quantitative RT-PCR was performed using gene-specific primers as described previously (Table S3) with SYBR Green (ABgene) in a MX3000P thermocycler (Strategene) (Martens et al., 2008).

Gene Deletion and Complementation in Bt
In-frame (nonpolar) gene deletions for mutants were generated using counter-selectable allele exchange (Martens et al., 2008). PCR amplified genes for complementation were ligated into the pNBU2-tetQb vector and conjugated into Bt via E. coli S17.1 λ-pir (Martens et al., 2008). Resulting clones were screened by PCR and sequenced to confirm isolates.

Gene Cloning
Genes for expression were amplified from Bt genomic DNA using the primers stated in Table S3 and cloned into pRSETA (Invitrogen) or pET22b (Novagen).

Protein Expression and Purification
Recombinant proteins were expressed in E. coli C41 or BL21 cells and purified in a single step using metal affinity chromatography as described previously (Bolam et al., 2004).

Sources and Preparation of Carbohydrates
Monosaccharides, sucrose, and chicory inulin for enzymatic and binding assays were obtained from Sigma. Growth of Bacteroides strains, qRT-PCR, and mouse experiments used inulin (Beneo-Orafti group; OraftiHP, OraftiLP, OraftiS), levan (Montana Polysaccharides), and inulin, FOS (Beneo-Orafti group; OraftiHP, OraftiLP, OraftiS) as described previously (Bolam et al., 2004). Enzyme assays were obtained from Sigma. Growth of Bacteroides strains, qRT-PCR, and mouse experiments used inulin, FOS (Beneo-Orafti group; OraftiHP, OraftiLP, OraftiS) as described previously (Bolam et al., 2004).

Isothermal Titration Calorimetry
Measurements were carried out essentially as described previously (Bolam et al., 2004), except that a Microcal VP-ITC machine was used, and proteins were dialyzed into 20 mM Tris-HCl (pH 8.0). The assumption that n = 1 for BT1762 binding to levans was based on the structure of the starch binding site of SusD (Koropatkin et al., 2008).

Thin-Layer Chromatography
Samples were spotted onto foil backed silica plates and placed in a glass tank equilibrated with butanol/acetic acid:water (2:2:1). Sugars were visualized using orcinol-sulphuric acid (sulphuric acid:ethanol:H2O 3:70:20 v/v, orcinol 1% w/v), 90°C for 5–10 min.

Enzyme Assays
All assays were carried out at 37°C in 20 mM Tris-HCl (pH 8.0). Activity of BT1760 was determined by quantifying the amount of reducing sugar released using the DNSA assay (Miller, 1959). Free fructose was determined using a modified fructose detection kit (Megazyme International). Kinetic parameters were determined by fitting initial rates versus substrate concentration (measured at six substrate concentrations that spanned the KM) to the Michaelis-Menten equation using nonlinear regression (Graphpad Prism, v5.0).

Enzyme Localization Studies
Cultures grown on 0.5% (w/v) fructose or glucose were harvested by centrifugation (OD600 ~1.0). PBS-washed cells and 0.5% levans or inulin in 20 mM Tris-HCl, pH8.0, were incubated at 37°C. Reducing sugar present was quantified using DNSA reagent (Miller, 1959). Activities of the periplasmic marker alkaline phosphatase and cytoplasmic marker glucose-6-phosphate dehydrogenase were compared to lysed cells to ensure no cell lysis/leakage occurred.

Bacterial Colonization and Density Determination of Germ-Free Mice
Germ-free Swiss-Webster mice were maintained in gnotobiotic isolators and fed an autoclaved standard diet (Purina LabDiet 5K67) or custom diet (Bio-Serv, http://bio-serv.com/), in accordance with A-PLAC, the Stanford IACUC. Mice were bi-associated using oral gavage (10⁹ CFU of each bacterial species). Relative densities of bacteria were determined by qPCR using strain-specific primers (Table S3) (Martens et al., 2008).

Crystallization, Structure Determination, and Refinement of BT1754-PD
Crystals formed in 0.7 M K/Na phosphate, 0.1 M HEPES (pH 8.0) (protein at 8 mg/ml with 5 mM fructose). Diffraction data, collected at Diamond Light Source (Oxford, UK) on a tiled ADSC Q315 CCD detector were processed with MOSFLM (Leslie, 1992). Scaling of data, search model generation, molecular replacement and structure refinement were carried out using SCALA, CHAINSAW, MOLREP and REFMAC (Collaborative Computational Project, 1994), respectively, with model rebuilding in COOT (Emsley and Cowtan, 2004).

ACCESSION NUMBERS
Protein Data Bank coordinates have been deposited under the accession code 2X7X.

SUPPLEMENTAL INFORMATION
Supplemental Information includes four tables, seven figures, and Supplemental References and can be found with this article online at doi:10.1016/j.cell.2010.05.005.

ACKNOWLEDGMENTS
We thank Karla Kirkegaard and Stanley Falkow for valuable comments and Sara Fisher for editing the manuscript. Inulin and FOS for mouse experiments were a kind gift from Beneo-Orafti. Levan was a kind gift from Montana Polysaccharides. We thank Jeffrey Gordon and members of the Gordon Lab for valuable advice; Carl Morland for excellent technical assistance; and Eric Martens and Andrew Goodman for development of genetic tools used in this paper. Some Bacteroides genomic data were produced by The Genome Institute at Washington University School of Medicine in St. Louis (genome. wustl.edu). This work was funded in part by grants from National Institutes of Health through the NIH Director’s New Innovator Award Program (DP2-OD006515) the NIDDK (K01-DK077053), the Stanford Digestive Disease Center (PO3-DK56339) and the BBSRC (BB/F014163/1).

Received: October 15, 2009
Revised: January 20, 2010
Accepted: April 27, 2010
Published: June 24, 2010
REFERENCES


**Note Added in Proof**

PL19 enzymes have recently been reclassified into glycoside hydrolase family 91 (see www.cazy.org/GH91.html).
Supplemental Information

SUPPLEMENTAL REFERENCES


Figure S1. Related to Figure 1. Coordinate Expression of Genes within the *B. thetaiotaomicron* Fructan PUL. BT1754 Is Required for Prioritized Fructan PUL Expression in Rich Media

(A) Expression of genes within the *B. thetaiotaomicron* fructan PUL when grown in rich media (TYG) from early log to stationary phase (five time points profiled in duplicate at 3.5, 4.5, 5.5, 6.5, and 8.8 hr after the initiation of growth; top panel) or in six mouse cecal (in vivo) samples after a ten day mono-association relative to expression in minimal medium plus 0.5% glucose (MMG). Colors indicate standard deviations above (red) and below (green) the gene’s mean (black) expression.

(B) Top panel shows growth of wild-type *B. thetaiotaomicron* (black circles) and *Bt*-ΔBT1754 (red circles) in TYG. Blue circles denote time points at which gene expression levels of BT1763 were measured. Bottom panel shows fold induction of BT1763 in wild-type *Bt* (black) and *Bt*-ΔBT1754 (red) at three time points from early log to stationary phase.
Figure S2. Related to Figure 2, Structures of the Main Carbohydrates Used in This Study

Levan and inulin chains have a terminal glucose residue as they are synthesized from sucrose. FOS (not shown) are low molecular weight β2-1 fructooligosaccharides derived from inulin, some of which are terminated by a glucose.
Figure S3. (continued).
Figure S3. Conservation of Fructose-Coordinating Residues in the Periplasmic Domain of Bacteroides Species' BT1754 Orthologs, Related to Figure 3
Amino acids in BT1754-PD that interact with ligand via H-bonds (blue arrows) or hydrophobic interactions (green arrows) are indicated. The position of the non-conserved residue in BF4326, the most divergent ortholog, is shown (blue circle). The presence of a completely conserved cysteine (red circle) just after the N-terminal transmembrane (TM) domain, and preceding three residues displaying similarity to the consensus lipobox sequence, suggests the HTCS are anchored to the inner membrane via a lipid linkage, despite lacking the aspartate at position +2 from the linked cysteine normally required for inner membrane localization (Juncker et al., 2003). The SMART predicted BT1754 domain boundaries are shown below the sequence (http://smart.embl-heidelberg.de/)(Letunic et al., 2009). The location of the invariant histidine and aspartate in the cytoplasmic phosphoacceptor and receiver domains, respectively, that are phosphorylated on activation of the HTCS are indicated (purple stars). The alignment was generated using ClustalW2 and ESPript (http://www.ebi.ac.uk/Tools/clustalw2/index.html and http://esprit.ibcp.fr/ESPript/ESPript/).
Figure S4. Related to Figure 5, Affinity Gels Showing the β2-6 Fructan Specificity of the susE-Positioned Gene Product, BT1761

Native gels were run in parallel in the presence (+) or absence (-) of polysaccharide (0.1% w/v final). BSA was used as a noninteracting negative control. The upper band in the BSA lanes is the dimeric form of the protein. Affinity gels were run as described previously (Bolam et al., 2004).
Figure S5. Related to Figure 4, Model of Levan Utilization by B. thetaiotaomicron

Based on our studies, protein localization predictions, and previous genetic and biochemical studies focused on other PULs of Bt (Cho and Salyers, 2001; Martens et al., 2008), we have developed a working model of β2-6 fructan utilization in Bt. In this model, levan is bound by the SusD homolog (D), BT1762, and the susE-positioned gene product, BT1761, on the surface of the bacterium. The polysaccharide is then cleaved into oligosaccharides by the extracellular endo-acting levanase, BT1760. Oligosaccharides are actively imported by the SusC homolog (C), BT1763 (a TonB dependent porin). In the periplasm, exo-acting glycoside hydrolases, BT1759 and BT3082, liberate monosaccharide fructose (green pentagons) from the oligosaccharides. Free fructose binds to the periplasmic sensor domain of the homodimeric HTCS, BT1754, activating the regulatory protein and resulting in upregulation of the PUL. The inner membrane monosaccharide importer, BT1758, transports periplasmic fructose into the cell where it is shunted into the glycolytic pathway upon phosphorylation by the PUL-encoded fructokinase, BT1757. Some transport across the inner membrane of the disaccharide products of fructan degradation (sucrose and levanbiose) also occurs and these are broken down to their constituent monosaccharides by the putative intracellular fructosidase, BT1765 (not shown). Prior to activation, low levels of all components of the PUL are expressed constitutively such that Bt is always in a prepared state.
Figure S6. Related to Figure 6, Three *B. thetaiotaomicron* Strains Possess a fructan PUL-Associated GH32 Family Member that Lacks Orthologs in Other Bacteroides Species

Phylogram illustrating relationship between genes of the GH32 enzymes from different Bacteroides species. Predicted amino acid sequences from Bacteroides GH32-family glycoside hydrolase genes were aligned using Clustal-W2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html). Alignments were used to generate a neighbor-joining phylogenetic tree. BT1760, *B. thetaiotaomicron*’s predicted extracellular GH32 has diverged from the GH32 genes in the other Bacteroides species, but has orthologs in two other *B. thetaiotaomicron* strains (7330, 3731), which also utilize levan (red box). BT, *B. thetaiotaomicron*; BF, *B. fragilis*; BACC, *B. caccae*; BACOVA, *B. ovatus*; BVU, *B. vulgatus*; BACUNI, *B. uniformis*. 
Figure S7. Related to Figure 7, Efficient Growth of Bt(In+) on Inulin

Growth curves of wild-type Bt and Bt-ΔBT1763, Bt(In+) and B. cacaoae in glucose (left) and inulin (right). Transfer of the region of the B. cacaoae fructan PUL consisting of the SusC gene through the GH32-encoding gene (BC02727 - BC02731) containing the BT1763 promoter region (intergenic region between BT1763 and BT1764 spliced immediately upstream of BC02727) into Bt-ΔBT1763 results in an inulin utilizing strain of Bt, Bt(In+).
Table S1. Doubling times for bacterial strains used in the study, related to Figure 2, Figure 4, Figure 5, and Figure 6

<table>
<thead>
<tr>
<th></th>
<th>B. theta</th>
<th>B. caccae</th>
<th>B. vulgatus</th>
<th>B. uniformis</th>
<th>B. fragilis</th>
<th>B. ovatus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>1.3</td>
<td>4.8</td>
<td>2.8</td>
<td>2.0</td>
<td>1.5</td>
<td>1.4</td>
</tr>
<tr>
<td>Fructose</td>
<td>1.2</td>
<td>13.5</td>
<td>3.2</td>
<td>2.1</td>
<td>1.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Sucrose</td>
<td>1.6</td>
<td>1.4</td>
<td>2.4</td>
<td>2.1</td>
<td>2.1</td>
<td>1.3</td>
</tr>
<tr>
<td>Levan</td>
<td>2.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FOS</td>
<td>5.6</td>
<td>2.1</td>
<td>3.0</td>
<td>2.8</td>
<td>2.0</td>
<td>1.4</td>
</tr>
<tr>
<td>Inulin</td>
<td>96</td>
<td>2.7</td>
<td>-</td>
<td>3.9</td>
<td>9.9</td>
<td>2.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>BTΔ1754</th>
<th>BTΔ1754: BT1754</th>
<th>BTΔ1760</th>
<th>BTΔ1760: BT1760</th>
<th>BTΔ1762</th>
<th>BTΔ1762: BT1762</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>2.6</td>
<td>2.6</td>
<td>2.9</td>
<td>3.4</td>
<td>2.1</td>
<td>ND</td>
</tr>
<tr>
<td>Fructose</td>
<td>11.8</td>
<td>2.5</td>
<td>2.6</td>
<td>2.1</td>
<td>1.9</td>
<td>ND</td>
</tr>
<tr>
<td>Sucrose</td>
<td>8.3</td>
<td>3.1</td>
<td>3.2</td>
<td>ND</td>
<td>2.5</td>
<td>ND</td>
</tr>
<tr>
<td>Levan</td>
<td>-</td>
<td>2.1</td>
<td>-</td>
<td>3.6</td>
<td>19.2</td>
<td>8.9</td>
</tr>
<tr>
<td>FOS</td>
<td>-</td>
<td>3.0</td>
<td>5.8</td>
<td>ND</td>
<td>3.8</td>
<td>3.5</td>
</tr>
<tr>
<td>Inulin</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

1Doubling times in hours of the Bacteroides species (top table) and the B. thetaiotaomicron mutants (bottom table) in minimal media containing a single carbon source. Doubling times were calculated using the KinetiCalc software (F. Breidt, 1994). ND : not determined, - : no growth.
Table S2. Related to Figure 4. Kinetic parameters of *Bacteroides thetaiotaomicron* GH32 enzymes against β2-6 and β2-1 fructans and fructooligosaccharides

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>1759 kcat (min⁻¹)</th>
<th>1759 K_M (mM)</th>
<th>1759 kcat/K_M (min⁻¹ mM⁻¹)</th>
<th>1760 kcat (min⁻¹)</th>
<th>1760 K_M (mM)</th>
<th>1760 kcat/K_M (min⁻¹ mM⁻¹)</th>
<th>1765 kcat (min⁻¹)</th>
<th>1765 K_M (mM)</th>
<th>1765 kcat/K_M (min⁻¹ mM⁻¹)</th>
<th>3082 kcat (min⁻¹)</th>
<th>3082 K_M (mM)</th>
<th>3082 kcat/K_M (min⁻¹ mM⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sucrose</td>
<td>816 ±136</td>
<td>7.8 ±1.2</td>
<td>104 ±113</td>
<td>NA</td>
<td>27087 ±3782</td>
<td>2.2 ±0.7</td>
<td>11286 ±2004</td>
<td>2.4 ±0.2</td>
<td>1421 ±35</td>
<td>1950 ±0.7</td>
<td>1910 ±1.0</td>
<td>1910 ±0.7</td>
</tr>
<tr>
<td>Kestose</td>
<td>379 ±75</td>
<td>2.7 ±0.8</td>
<td>140 ±25</td>
<td>NA</td>
<td>3329 ±922</td>
<td>11.8 ±3.0</td>
<td>282 ±204</td>
<td>0.9 ±0.0</td>
<td>2227 ±204</td>
<td>1950 ±0.1</td>
<td>1910 ±1.0</td>
<td>1910 ±0.1</td>
</tr>
<tr>
<td>Kestotetraose</td>
<td>706 ±125</td>
<td>0.9 ±0.1</td>
<td>784 ±30</td>
<td>NA</td>
<td>3829 ±684</td>
<td>5.1 ±2.1</td>
<td>751 ±1365</td>
<td>0.7 ±0.1</td>
<td>1950 ±1.5</td>
<td>1950 ±0.1</td>
<td>1910 ±1.0</td>
<td>1910 ±0.1</td>
</tr>
<tr>
<td>Kestopentaose</td>
<td>545 ±45</td>
<td>1.2 ±0.1</td>
<td>454 ±20</td>
<td>NA</td>
<td>4437 ±263</td>
<td>9.0 ±0.0</td>
<td>456 ±412</td>
<td>0.5 ±0.1</td>
<td>1910 ±1.0</td>
<td>1910 ±0.1</td>
<td>1910 ±1.0</td>
<td>1910 ±0.1</td>
</tr>
<tr>
<td>Inulin</td>
<td>156 ±63</td>
<td>0.3b ±0.1</td>
<td>520 ±15</td>
<td>NA</td>
<td>~1235 ±119</td>
<td>&gt;0.3b</td>
<td>609 ±119</td>
<td>0.6b ±0.1</td>
<td>1688 ±1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levanbiose</td>
<td>1052 ±209</td>
<td>6.0 ±0.3</td>
<td>175 ±47</td>
<td>ND</td>
<td>9330 ±1509</td>
<td>8.8 ±0.6</td>
<td>1060 ±250</td>
<td>5.3 ±0.5</td>
<td>412 ±218</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levantriose</td>
<td>594 ±129</td>
<td>7.3 ±0.6</td>
<td>81 ±27</td>
<td>ND</td>
<td>1857 ±67</td>
<td>9.3 ±0.3</td>
<td>200 ±1600</td>
<td>5.6 ±0.5</td>
<td>286 ±184</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levanetetraose</td>
<td>762 ±175</td>
<td>3.0 ±1.7</td>
<td>254 ±28</td>
<td>ND</td>
<td>2696 ±181</td>
<td>18.8 ±1.5</td>
<td>143 ±1841</td>
<td>5.6 ±0.4</td>
<td>329 ±192</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levanpentaose</td>
<td>737 ±74</td>
<td>3.0 ±1.6</td>
<td>246 ±24</td>
<td>ND</td>
<td>4043 ±272</td>
<td>34.7 ±0.4</td>
<td>116 ±1881</td>
<td>5.4 ±0.4</td>
<td>348 ±110</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levan</td>
<td>205 ±47</td>
<td>1.0mg/ml ±0.3</td>
<td>205 ±115</td>
<td>ND</td>
<td>~288 ±32987</td>
<td>&gt;100mg/ml</td>
<td>~3112 ±9383</td>
<td>1.6 ±0.1</td>
<td>596 ±95</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Values shown are the mean and SD of at least 3 independent assays for each enzyme/substrate pair. Kesto-oligosaccharides are β2-1 linked fructose with a terminal non-reducing glucose. Levanoligosaccharides are β2-6 linked fructose with no terminal glucose moiety.

aNA - No activity detected.
bmM K_M for inulin based on average MW of ~4000 (Sigma; Chicory, Cat. no. I2255).
cND - Not determined.
Table S3. List of primers used in this study

**Quantitative RT-PCR**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT1754_forward</td>
<td>CGCAATCTGATCGATTCTCA</td>
</tr>
<tr>
<td>BT1754_reverse</td>
<td>ACACGGCTTACGCTCATGTC</td>
</tr>
<tr>
<td>BT1757_forward</td>
<td>GTCTTTTAAACGCTGCAACA</td>
</tr>
<tr>
<td>BT1757_reverse</td>
<td>GAGCTTCCGGAAACAGACTTTG</td>
</tr>
<tr>
<td>BT1763_forward</td>
<td>ATGCCTGGTCACCTACGAAC</td>
</tr>
<tr>
<td>BT1763_reverse</td>
<td>CAAGCGGTCCATTCTCATTTT</td>
</tr>
<tr>
<td>BT1765_forward</td>
<td>AGACCTGATGCATGGGGAAC</td>
</tr>
<tr>
<td>BT1765_reverse</td>
<td>CCATTACCTCCGGTGTCAT</td>
</tr>
<tr>
<td>BT3082_forward</td>
<td>CACCGGAACCTTCCGGCTTA</td>
</tr>
<tr>
<td>BT3082_reverse</td>
<td>CGTGGCATGAGAGAGGTTGA</td>
</tr>
<tr>
<td>BT3305_forward</td>
<td>CGGTGCGGACAAAGTATCAGC</td>
</tr>
<tr>
<td>BT3305_reverse</td>
<td>AAAGTCCTTTCCCACACTGAA</td>
</tr>
</tbody>
</table>

**Quantitative PCR**

<table>
<thead>
<tr>
<th>B. theta_forward</th>
<th>GGGGGTATCTTACCTACCTTCGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. theta_reverse</td>
<td>ATTCGGTTGAACGCTTGTCT</td>
</tr>
<tr>
<td>B. caccae_forward</td>
<td>CAGCCCGCTACTTGAAGCTC</td>
</tr>
<tr>
<td>B. caccae_reverse</td>
<td>TTGACGGAGGCAAAATAGG</td>
</tr>
<tr>
<td>B. vulgatus_forward</td>
<td>TAGAGATCCGCTCGTGTCG</td>
</tr>
<tr>
<td>B. vulgatus_reverse</td>
<td>TCCAAACGAGGAAGCCCATC</td>
</tr>
</tbody>
</table>

**B. theta mutant generation**

<table>
<thead>
<tr>
<th>BTΔ1754</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTΔ1754_1000up</td>
<td>AAAAGGATCCACCGGAAAAGTGGGAAATGTA</td>
</tr>
<tr>
<td>BTΔ1754_750up</td>
<td>AAAAGGATCCACCTTCTGGGAAAGCGGAGA</td>
</tr>
<tr>
<td>BTΔ1754_700up</td>
<td>AAAAGGATCCCTCCATGGTTCATTTGCAAAC</td>
</tr>
<tr>
<td>BTΔ1754_sewing_forward</td>
<td>CCTCTTTACGATATCAATGGAAATTACATTTTCATAGTTTTCTGTTGTCATCC</td>
</tr>
<tr>
<td>BTΔ1754_1000dwn</td>
<td>AAAATCTAGATGGCATTTGCTGCTGCTAT</td>
</tr>
</tbody>
</table>
BT\textsuperscript{\Delta}1754_750dwn AAAAAATCTAGAACAAGCTCCGTGCACCTCAGA
BT\textsuperscript{\Delta}1754_700dwn AAAAAATCTAGATCTGGCTCTTTTTGGGGAATT
BT\textsuperscript{\Delta}1754\_sewing\_reverse GGATTACAGAAAGAACTATGAAAAATGTAATTTCATGGATACGTAAAGAGG

\textbf{BT\textsuperscript{\Delta}1754:BT1754}

BT\textsuperscript{\Delta}1754:BT1754\_forward AAAAAATCTAGAATCTCCCTCGCTTCTCTTCTGTTGTTACTT
BT\textsuperscript{\Delta}1754:BT1754\_reverse AAAAAAGATCCTTTAAAAATCCGATGTTAAGTCCGAA

\textbf{BT1760}

BT\textsuperscript{\Delta}1760\_1000up AAAAAAGATCCCTCGTAGTGGCTCCGGCTGACA
BT\textsuperscript{\Delta}1760\_750up AAAAAAGATCCCTTTTGGTACGTTCTTCTTC
BT\textsuperscript{\Delta}1760\_700up AAAAAAGATCCGGATAATGTAATCGATGTCCGAAT
BT\textsuperscript{\Delta}1760\_sewing\_forward CTTGCCGGTGTAGTTTTCATCACATTGTTGCTTTATTCTTTATA
BT\textsuperscript{\Delta}1760\_1000dwn AAAAAATCTAGACAGCGTTTTAGTTGCTTCGTA
BT\textsuperscript{\Delta}1760\_750dwn AAAAAATCTAGACGGTGTTCCGAGTTTACGAG
BT\textsuperscript{\Delta}1760\_700dwn AAAAAATCTAGACGGTGTTCCGAGTTTACGAG
BT\textsuperscript{\Delta}1760\_sewing\_reverse TGGTTAAATAAAAAGAATAGAAAGCTGATGAAAACCTCACCCGACAAG

\textbf{BT\textsuperscript{\Delta}1760:BT1760}

BT\textsuperscript{\Delta}1760:BT1760\_forward AAAAAATCTAGACCCGAGTTTCTCGGATTCC
BT\textsuperscript{\Delta}1760:BT1760\_sewing\_reverse CATATAGGTAAGATCATATTATTCTCATAGGATATTCGAGTTTACTATG
BT\textsuperscript{\Delta}1760:BT1760\_sewing\_forward CATAGTAGAAACTGTAATCCCTGATGAAAATGATCTTTACCTATAG
BT\textsuperscript{\Delta}1760:BT1760\_reverse AAAAAAGATCCCTGATAAGTGCTTACCTGACGT

\textbf{BT1762}

BT\textsuperscript{\Delta}1762\_1000up AAAAAAGATCCCGGTATCGACTTCCGCTGGT
BT\textsuperscript{\Delta}1762\_750up AAAAAAGATCCGAGTTTCCGAAACGCTAGGC
BT\textsuperscript{\Delta}1762\_700up AAAAAAGATCCGGAAGTTGTAATCGGACATAAC
BT\textsuperscript{\Delta}1762\_sewing\_forward CTTCATCAAATAAACGAAATGATCATGTAACGTCACATCTTTAAAAAAGAAAATA
BT\textsuperscript{\Delta}1762\_1000dwn TATTCTTTTTAAATAGTGTTAGTTACATGATCTTTACCTTTATTGATGAAG
BT\textsuperscript{\Delta}1762\_750dwn AAAAAATCTAGAAGATATCCGACATCGGATACATT
BT\textsuperscript{\Delta}1762\_700dwn AAAAAATCTAGAAGATATCCGACATCGGATACATT
BT\textsuperscript{\Delta}1762\_sewing\_reverse TATTCTTTTTAAAGATGTTAAGTACATGATCATTCCGTTTTATTGATGAAG

\textbf{BT\textsuperscript{\Delta}1762:BT1762}

BT\textsuperscript{\Delta}1762:BT1762\_forward AAAAAATCTAGAATCTCCCTCGCTTCTCTTCTGTTGTTACTT
BTΔ1762:BT1762_sewing_reverse: CGATTGTTGCTATATATATATTATTTTTTCATTAGTTTAATGTTAATTAAAAAGTACG
BTΔ1762:BT1762_sewing_forward: CGTACTTTTAAAAATTAACATTTAACATTTAAATTGAAAAAGATAATATATATAGCAACAATC
BTΔ1762:BT1762_reverse: AAAAGGATCCCTACCAACCGAAATTCTGTGTAT

BTΔ1763
BTΔ1763_1000up: AAAAGGATCCATAGACCAGCTACCGGAGCA
BTΔ1763_750up: AAAAGGATCCGACATTTAAAATCCGGAATAA
BTΔ1763_700up: AAAAGGATCCGACACCGATTTTGAAATACA
BTΔ1763_sewing_forward: CATTCTGTTTTATTTGTAAGTAAATGTAATTATAAGTTATATAAGTTAAAAGTA
BTΔ1763_1000dwn: AAATCTAGACAGTATCTGCGGAGTGCTCA
BTΔ1763_750dwn: AAATCTAGACCTCGTTGATTCCGGTAAGG
BTΔ1763_700dwn: AAATCTAGATTTTTGAAGGTAAGCAGCA
BTΔ1763_sewing_reverse: TACTTTTTAAATTAACATTAATAACTAATGTAACCTACTTACTTACATCAATAAAAACGAATG

BT(In+)
BCexch_BT1763_forward: AAAAGCGGCCG CATCATTCAGTTTTCTGTTGGTTACTT
BC02727_reverse: AAAATCTAGAAATGTTCTGTATTTAGGATAAAAGATTAA
BC02728-BC02731_forward: AAAATCTAGAATGAAATTGAAATATATCCTTGG
BC02728-BC02731_reverse: AAAAGGATCCATAGAACATCCGGGTAGACG
BC_BT1763_sewing_reverse: TTTTGTTCTTCATAATACGTAGCATTAGTTAATTTAATTTAAAG
BC_BT1763_sewing_forward: CTTTTAAATTAACATTAATAACTAATGCTCAGTATTATGAAGAACAAAA

Amplification of fructan PUL genes for expression

BT1754-PD_F: CTCCCATAGGATGACACACTCCCCATTTCGTATTT
BT1754-PD_R: CCGCTCGAGACCTGTGGTTTAGCTAC

BT1759_F: CTCCCATAGGATGCGGGATTTCCTCTTTGC
BT1759_R: CCGCTCGAGTTTCTTACGTTTGTAAACC

BT1760_F: CTCCCATAGGATAGTGACGAGACTGAC
BT1760_R: CCGCTCGAGTTTCTTACGTTTGTAAACC

BT1765_F: CTCCCATAGGATAAAAAACACTCTTGGATAAACC
BT1765_R: CCGCTCGAGTAACCTAATCTTACACAC

BT3082_F: CCGCATCTCGGGAGAAGTATCTTAAAAATAAACGAC
BT3082_R  CTCGAATTCTACCAAATGGATTCTACGGAAAGAC

BT1761_F  CTCGGATCC TAGTGATGACTTCAAATCCGGCC
BT1761_R  CTCGAATTC TATTTACACAAAGTAGTTGATTGCATTGAGAG

BT1762_F  CTCGGATCC GACGATTTTTTGGACCGTCAGGTTCC
BT1762_R  CTCGAATTC TTTACAAACCGAAATTCTGTGTATAATTTCC

1 All GH32 enzymes (except BT1765), BT1754-PD, BT1761 and BT1762 were expressed without their predicted endogenous signal peptides (http://www.cbs.dtu.dk/services/SignalP/ and http://www.cbs.dtu.dk/services/LipoP/). BT1754-PD, BT1759, BT1760 and BT1765 were cloned into pET22 or pET28 (Novagen) such that the recombinant protein contained a C-terminal His<sub>6</sub> tag. BT1761, BT1762 and BT3082 were cloned into pRSETA (Invitrogen) encoding an N-terminal His<sub>6</sub> tag. Restriction sites introduced for cloning are highlighted.
Table S4. X-ray diffraction data collection and refinement statistics for BT1754-PD, related to Figure 3

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P 65</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>a=b=111.84, c=115.171</td>
</tr>
<tr>
<td>Resolution Range (Å)</td>
<td>37-2.64 (2.78-2.64)*</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>23919 (3512)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.7 (99.7)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.9 (6.0)</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.10 (0.35)</td>
</tr>
<tr>
<td>Mean I/(s)I</td>
<td>12.4 (5.6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>35.7-2.64 (2.70-2.64)</td>
</tr>
<tr>
<td>Rwork</td>
<td>18.6 (22.8)</td>
</tr>
<tr>
<td>Rfree</td>
<td>23.1 (28.6)</td>
</tr>
<tr>
<td>No. non-H atoms</td>
<td>4741</td>
</tr>
<tr>
<td>Protein</td>
<td>4596</td>
</tr>
<tr>
<td>Ligand</td>
<td>24</td>
</tr>
<tr>
<td>Water/ion</td>
<td>121</td>
</tr>
<tr>
<td>Mean B, all atoms (Å²)</td>
<td>36.2</td>
</tr>
<tr>
<td>Protein</td>
<td>36.1</td>
</tr>
<tr>
<td>Ligand</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Water/ion</td>
<td>39.5</td>
</tr>
<tr>
<td>Rmsd B values (Å(^2)) (m.c./s.c.)</td>
<td>0.5/1.1</td>
</tr>
<tr>
<td>No. Ramachandran outliers</td>
<td>2</td>
</tr>
<tr>
<td>Rmsd bond lengths (Å)</td>
<td>0.010</td>
</tr>
<tr>
<td>Rmsd bond angles (°)</td>
<td>1.3</td>
</tr>
</tbody>
</table>

\(^a\) Values in parentheses refer to values in the highest resolution shell