A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes

Johan Larsbrink1,*, Theresa E. Rogers2,*, Glyn R. Hemsworth3,*, Lauren S. McKee1,4, Alexandra S. Tauzin5, Oliver Spadiut1,4, Stefan Klinter1, Nicholas A. Pudlo5, Karthik Urs2, Nicole M. Koropatkin2, A. Louise Creagh6, Charles A. Haynes6, Amelia G. Kelly2, Stefan Nilsson Cederholm1, Gideon J. Davies3, Eric C. Martens3 & Harry Brumer1,5

A well-balanced human diet includes a significant intake of non-starch polysaccharides, collectively termed ‘dietary fibre’, from the cell walls of diverse fruits and vegetables1. Owing to the paucity of alimentary enzymes encoded by the human genome2, our ability to derive energy from dietary fibre depends on the saccharification and fermentation of complex carbohydrates by the massive microbial community residing in our distal gut3–4. The xyloglucans (XyGs) are a ubiquitous family of highly branched plant cell wall polysaccharides5–6 whose mechanism(s) of degradation in the human gut and consequent importance in nutrition have been unclear7–9. Here we demonstrate that a single, complex genetic locus in Bacteroides ovatus confers XyG catabolism in this common colonic symbiont. Through targeted gene disruption, biochemical analysis of all predicted glycoside hydrolases and carbohydrate-binding proteins, and three-dimensional structural determination of the vanguard endo-xyloglucanase, we reveal the molecular mechanisms through which XyGs are hydrolysed to component monosaccharides for further metabolism. We also observe that orthologous XyG utilization loci (XyGULs) serve as genetic markers of XyG catabolism in Bacteroidetes, that XyGULs are restricted to a limited number of phylogenetically diverse strains, and that XyGULs are ubiquitous in surveyed human metagenomes. Our findings reveal that the metabolism of even highly abundant components of dietary fibre may be mediated by niche species, which has immediate fundamental and practical implications for gut symbiont population ecology in the context of human diet, nutrition and health10–12.

Despite our omnivory, a census of the glycoside hydrolases (GHs) encoded by the human genome indicates that our inherent ability to digest carbohydrates is restricted to starch and simple saccharides such as malto-oligosaccharides, sucrose and lactose2. Consequently, the human gut microbiota and its cohort of predicted carbohydrate-active enzymes are implicated in the conversion of otherwise indigestible plant polysaccharides to short-chain fatty acids7,13, which provide up to 10% of daily caloric intake in humans14,15, and are central to colonic health16–18. Despite an increasing body of (meta)genomic sequence data9–12, the enzymatic pathways by which the most common dietary polysaccharides are digested in the human gut have not been elucidated7,13.

XyGs are widespread in the vegetables we consume: dicot primary cell walls, for example those of lettuce, onions and tomatoes, may contain up to 25% XyG on a dry-weight basis19,20. The primary walls of commelinioid monocots, including the cereals, contain much lower (1–5%)—but still non-zero—amounts of XyGs21. Seed XyGs are also widely used as food-thickening agents and have been used as drug delivery matrices in the intestine3. This family of polysaccharides is typified by a β(1–4)-glucan main chain that is heavily substituted with pendant α(1–6)-linked xylosyl units. Depending on the species and tissue of origin, these branches may be further extended by additional monosaccharides, including galactose, fucose and/or arabinose5,22 (Fig. 1). As such, complete saccharification in the gut necessarily requires a cadre of enzymes to address the monosaccharide and linkage diversity of these complex polysaccharides. We recently identified a polysaccharide utilization locus (PUL) in the genome of a common human gut symbiont, B. ovatus—but not in the closely related model species B. thetaiotaomicron23—that was transcriptionally upregulated in response to growth on galactoxyloglucan. By homology with the archetypal starch utilization system (Sus) of B. thetaiotaomicron, this PUL was predicted to encode an outer membrane sugar-binding protein (SusD-like), a TonB-dependent sugar receptor/transporter (SusC-like), and an inner membrane hybrid two-component sensor. Further analysis revealed that this PUL was also predicted to encode eight GHs from six enzyme families (Fig. 2), which tantalizingly suggested a collective role in XyG utilization by B. ovatus. To establish a direct causal link for growth on XyG24 and outline a pathway for its degradation, we performed an in-depth molecular characterization of the PUL through reverse genetics, in vitro protein biochemistry and enzymology, and structural biology.

A mutant strain of B. ovatus harbouring a targeted deletion of the complete predicted XyGUL (Fig. 2) was indeed totally unable to grow in vitro on XyG.
Figure 2 | Structure of the *B. ovatus* XyGUL and evolution in the Bacteroidetes lineage. a, PULs with partial homology and synteny; homologous genes are connected by grey bars and flanking genes lacking synteny are shown as semi-transparent. b, PULs with partial homology, but lacking overall synteny. Extended Data Fig. 2 provides transcriptional evidence that each of these gene clusters is responsive to growth on XyG.

Figure 3 | The concerted action of XyGUL gene products in the degradation of XyGs. a, b, Most probable sequential pathways for the hydrolysis of galactoxyloglucan (a) and arabinogalactoxyloglucan (b) on the basis of enzyme kinetic data, product analysis and selected gene knockout studies (see Fig. 1 for XyG motif abbreviations). Enzymes are represented as circles, colour-coded as in panel c: rainbow, *endo*-xylloglucanase BoGH5A; tan, *endo*-xylloglucanase BoGH5A; orange, α-xyllosidase BoGH31A; turquoise, α-L-arabinofuranosidase BoGH9A; blue, β-glucosidase; BoGH3A and BoGH3B, yellow, β-galactosidase; BoGH43A and BoGH43B, dark blue, β-glucosidases BoGH3A and/or BoGH43B; yellow, β-galactosidase; BoGH3 and/or BoGH43. c, Model of enzyme localization by analogy with the archetypal sus locus and based on inference of N-terminal lipoprotein modification from protein sequence data. Reducing-sugar assays and mass spectrometry demonstrated that the recombinant *B. ovatus* (Bo)GH5A and BoGH9A enzymes were versatile *endo*-xylloglucanases, which cleaved the backbone of the three major types of natural XyGs, that is, seed galactoxyloglucan from tamarind kernel, dicot fucogalactoxyloglucan from lettuce leaves, and solanaceous arabinogalactoxyloglucan from tomato fruit, to produce XyG oligosaccharides (XyGOs) based on a Glc₄ backbone (Extended Data Table 1 and Supplementary Figs 7–10, compare with Fig. 1). Assays on chromogenic aryl β-glycosides and natural XyGOs, together with product analyses, revealed the following *exo*-specificities for the remaining XyGUL enzymes: BoGH2A, β-galactosidase; BoGH3A and BoGH3B, β-glucosidase; BoGH31A α-xyllosidase; BoGH43A and BoGH43B, α-L-arabinofuranosidase (Extended Data Table 1 and Supplementary Figs 11–20). This profile of activities, together with phenotypic data from additional gene-specific knockout strains (Extended Data Table 2), allowed us to outline a complete model of XyG degradation by *B. ovatus* (Fig. 3).

Analogous to the *endo*-amylase SusG of the *B. thetaiotaomicron* starch utilization system, the vanguard role in XyG utilization by *B. ovatus* is performed by the versatile *endo*-xylloglucanase BoGH5A, which generates short XyGOs for uptake (Fig. 3). Indeed, a gene-specific knockout of *BoGH5A* rendered *B. ovatus* incapable of growth on XyG polysaccharide, but this phenotype could be directly rescued by the addition of XyGOs produced exogenously by recombinant BoGH5A (Extended Data Table 1 and Supplementary Figs 11–20). This profile of activities, together with phenotypic data from additional gene-specific knockout strains (Extended Data Table 2), allowed us to outline a complete model of XyG degradation by *B. ovatus* (Fig. 3).

Analogous to the *endo*-amylase SusG of the *B. thetaiotaomicron* starch utilization system, the vanguard role in XyG utilization by *B. ovatus* is performed by the versatile *endo*-xylloglucanase BoGH5A, which generates short XyGOs for uptake (Fig. 3). Indeed, a gene-specific knockout of *BoGH5A* rendered *B. ovatus* incapable of growth on XyG polysaccharide, but this phenotype could be directly rescued by the addition of XyGOs produced exogenously by recombinant BoGH5A (Extended Data Table 1 and Supplementary Figs 11–20). This profile of activities, together with phenotypic data from additional gene-specific knockout strains (Extended Data Table 2), allowed us to outline a complete model of XyG degradation by *B. ovatus* (Fig. 3).
toxyloglucan (Fig. 1). This may reflect compensation by exogenous or as in the archetypal Sus system 24, polysaccharide binding is mediated locus tags Bacova_02651 and Bacova_02650, respectively) indicates that, By contrast, the observation of strong XyG binding (Extended Data Table 1) may suggest simple loss-of-function or evolution of a cur- rently unresolved, orthogonal activity contrasting with that of BoGH43A. Regardless, it is clear that a broad active-site cleft engendering binding plasticity is the key feature allowing BoGH5A to accommodate a wide range of natural XyGs.

Figure 4 | Structural biology of BoGH5A. a, Tertiary structure; the two conformations observed in crystallo have been oriented relative to the N-terminal, membrane-anchored BACON domain (see also Supplementary Video 1). b, Divergent stereo view of the binding of XXXG in the −4 to −1 subsites (see also Supplementary Video 2). The wire frame represents an unbiased 2Fo − Fc map (contoured at 0.3 electrons per Å^3) obtained using phases calculated from the best model prior to the incorporation of any ligand in refinement.

The catalytic domain has the prototypical (α/β)_N fold typical of other GH5 members and is most similar to the Paenibacillus pabuli endo-xylloglucanase PpXG5 (Protein Data Bank (PDB) accession 2JEP; root mean squared deviation (r.m.s.d.) 1.47 Å, 330 Cα atoms) and the Clos- tridium cellulovorans endo-glucanase EngD (PDB accession 3ICG; r.m.s.d. 1.45 Å, 345 Cα atoms) of subfamily 4 (ref. 26). The heptasaccharide XXXG was observed to extend from subsites −4 to −1 in the active-site cleft in both molecules of the asymmetric unit, with binding mediated through sugar–aromatic-ring interactions and hydrogen bonds to both backbone and branch- ing substrate moieties (Fig. 4b and Supplementary Video 2). The structure of this complex is well correlated with activity on chromogenic XyGO aryl-β-galactosidase converting ‘L’ units (a β-fucosidase, as might be anticipated for the cleavage of the ‘F’ side chain in dicot fucogalac- togalacturonan (Fig. 1). This may reflect compensation by exogenous or endogenous α-(1→2)-fucosidases2 or strain specialization for XyGs from individual plant sources. Indeed, XyGs from other Bacteroidetes species encode predicted α-fucosidases from families GH29 and GH95 (Fig. 2).

To provide further insight into XyG recognition by the keystone enzyme, we solved the three-dimensional structure of BoGH5A in complex with the heptasaccharide XXXG by X-ray crystallography (Fig. 4 and Extended Data Table 3). The tertiary structure comprised a 96-residue bacteroidetes-associated carbohydrate-binding often N-terminal (BACON) domain27–29 followed by a 372-residue C-terminal GH5 domain. The BACON domain comprised an eight-stranded, immunoglobulin-like β-sandwich fold and represents, to our knowledge, the first three- dimensional representative of this domain family. Notably, the two molecules in the asymmetric unit showed a large difference in the relative orientations of the BACON and GH5 domains (Fig. 4a). This suggested significant flexibility in the enzyme (Supplementary Video 1), which is anchored on the cell surface through N-terminal lipidation; mutation of the predicted lipidation site (Cys 1-Ala) ablates surface localization and hampers growth on XyG (Extended Data Fig. 3). Notably, we were unable to find experimental evidence in vitro that the BACON domain functions in substrate binding (Extended Data Fig. 4), nor that it mediates interactions with other proteins of the XyGUL (native polyacrylamide gel electrophoresis (PAGE) data not shown). By contrast, the observation of strong XyG binding (Extended Data Fig. 4) by the SusD-like protein and neighbouring gene product (Fig. 2; locus tags Bacova_02651 and Bacova_02650, respectively) indicates that, as in the archetypal Sus system24, polysaccharide binding is mediated by independently encoded, non-catalytic proteins of the XyGUL. Thus, the broad designation of BACON domains as ‘carbohydrate binding’, as inferred by bioinformatics alone, may be actively misleading2,30,31. In light of current experimental data, the most parsimonious conclusion is that the primary function of the BACON domain in BoGH5A may be to distance the catalytic module from the cell surface and confer additional mobility to the catalytic domain for attack of the polysaccharide.
in nature. XyG catabolism was rare, confined to 70 strains belonging to just 6 individual, phylogenetically dispersed species (Fig. 5). Of the 25 B. ovatus strains tested, all but one grew on tamarind XyG. Interestingly, none of the 18 strains of the closely related B. xylanisolvens exhibited growth on XyG, which suggested that the XyGUL has been independently acquired by the B. ovatus lineage. This is supported by a comparative genomic analysis of the XyGUL and surrounding chromosomal region in these two species (Extended Data Fig. 1). All ten strains tested of the more distantly related B. cellobioslyticus grew very well on XyG, comparable with the less proficient B. ovatus strains. Representing a third phylogenetic clade, the single B. fluxus strain also grew to high density on XyG. Of 37 strains of B. uniformis, which is closely related to B. ovatus and B. uniformis grew to high density on XyG. Of 37 strains of B. fluxus and among the most abundant organisms in the microbiota of Westerners18, 33 exhibited varying abilities to grow on XyG. Finally, two non-Bacteroides species, Dysgonomonas gadei and D. mossii, which are rarely observed in humans, but are abundant in termites, also possessed the capacity to degrade XyG, revealing that this phenotype has evolved in other lineages of Bacteroidetes.

A comparative analysis using genomic sequence data available for many of the strains tested showed a perfect concordance between the presence of an orthologous XyGUL and the ability to utilize XyG. Thus, we revealed putative XyGULs in B. cellobioslyticus, B. uniformis, B. fluxus, D. mossii and D. gadei, as well as in two sequenced species, B. coprococola and B. salanitronis, that were not available to test growth directly (Figs 2 and 5). These XyGULs were all activated during growth on XyG (Extended Data Fig. 2) and shared a similar overall organization with that of B. ovatus, but did not share similar flanking regions in their genomes, suggesting that they were acquired through separate events. In notable contrast with a recent report showing that conjugative transposons are capable of mediating the horizontal transfer of new PULs into Bacteroides genomes25, there is no evidence of such a mechanism for XyGUL acquisition. Rather, the XyGUL seems to have been precisely inserted between two ancestral genes that are shared by all sequenced strains. For example, on the left side of XyGUL (Fig. 2 and Extended Data Fig. 1), there are only 108 base pairs (bp) between the 3' end of BoG33A (Bacova_02644) and the end of the adjacent shared region; on the right side there are only 640 bp between the end of BoGH33B (Bacova_02659) and the second adjacent shared region.

To underscore the broad importance of dietary XyG metabolism, we surveyed public metagenomic data from 250 adult humans, revealing that 92% harbour at least one of five different Bacteroides XyGULs identified in this study (Extended Data Fig. 5). By contrast, the presence of a PUL involved in degrading the red algal polysaccharide porphyran (a component of the food product nori)26 was exceptionally rare in the same population, and largely confined to a small cohort of Japanese subjects. To demonstrate that a metabolic advantage is associated with possessing a XyGUL, a competition experiment was performed using germ-free mice co-colonized with wild-type B. ovatus and an isogenic XyGUL deletion mutant. Both strains competed equally when glucose was presented as the sole dietary carbohydrate or when XyG was presented in a mixture of polysaccharides from natural vegetable sources (a diet comprising equal amounts of cooked bell pepper, aubergine, tomato fruit and lettuce). However, when this complex diet was removed and XyG was maintained as the only exogenous polysaccharide via the drinking water, the B. ovatus XyGUL mutant exhibited a significant fitness defect over a three-week period (Extended Data Fig. 6).

The prevalence of XyGs in the human diet suggests that the mechanism by which bacteria degrade these complex polysaccharides is highly important to human energy acquisition. Moreover, the rarity of XyG metabolism (Fig. 5, see also ref. 8) highlights the significance of B. ovatus and other proficient XyG-degrading Bacteroidetes as key members of the human gut microbial consortium14. Additional work will now be required to determine whether bacteria from other groups have evolved strategies to attack this ubiquitous plant cell wall polysaccharide in competition or synergy with members of the Bacteroidetes8,30. An unexpected finding is the specific adaptation of the GH complement of some XyGULs to XyGs from different plant sources; this theme may extend to other groups of plant polysaccharides and host mucosal glycans with similar or greater degrees of fine-structural variation. Following the example presented here, defining cohorts of enzymes and binding proteins that coordinate to target other complex dietary polysaccharides will be essential to elucidate fully the systems biology of gut microbial catabolism. Such a refined mechanistic understanding will be essential in designing rational intervention strategies, including prebiotics, probiotics or microbial community transplants, which aim to manipulate the membership, function and stability of this important ecosystem4,9–12.

**METHODS SUMMARY**

Gene deletions in B. ovatus were conducted by allelic exchange in a Δtdk (thymidine kinase, Bacova_03071) derivative strain of ATCC 8483 using the vector pExchange-tdk. Growth was measured spectrophotometrically in minimal media including XyG or XyGOs as appropriate. The Bacteroidetes species phylogeny was constructed by concatenating nucleotide sequences of six core genes, alignment with CLUSTAL-W, trimming of poorly aligned base pairs with Gblocks, and maximum parsimony tree building with MEGA5. Genes encoding the GHs of the XyGUL were cloned into pET21a vectors (predicted signal–peptide-encoding sequences were removed) and overexpressed in E. coli BL21 (DE3) cells at 25 °C. Proteins were purified by immobilized nickel affinity and gel-filtration chromatography and expected molar masses were verified by mass spectrometry. Enzyme activities were determined against a range of aryl glycosides, polysaccharides and oligosaccharides using standard spectrophotometric, reducing-sugar and coupled enzyme assays; enzyme reaction products were analysed by high-performance anion exchange chromatography and matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry. Poly- and oligosaccharides were produced by extraction from natural sources and selective enzymatic hydrolysis using XyGUL and other enzymes as appropriate. X-ray diffraction data were collected at the

![Figure 5](image-url)
Diamond Light Source synchrotron, station I03, and the structure was solved by molecular replacement using chain A of the P. pacifica GH5 (PDB accession 2EFP) as the search model.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 February; accepted 22 November 2013.

Published online 19 January 2014.

1. McDougall, G. J., Morrison, J. M., Stewart, D. & Hillman, J. R. Plant cell walls as dietary fibre: range, structure, processing and function. J. Sci. Food Agric. 70, 130–150 (1996).
2. El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and activity of carbohydrate-active enzymes in the human microbiota. Nature Rev. Microbiol. 11, 497–504 (2013).
3. Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P. & Forano, E. Microbial degradation of complex carbohydrates in the gut. Gut Microbes 3, 289–306 (2012).
4. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. Nature Rev. Microbiol. 10, 323–335 (2012).
5. Hoffman, M. et al. Structural analysis of xyloglucans in the primary cell walls of plants in the subclass Asteridae. Carbohydr. Res. 340, 1826–1840 (2005).
6. Vogel, J. Unique aspects of the grass cell wall. Curr. Opin. Plant Biol. 11, 301–307 (2008).
7. Martens, E. C. et al. Recognition and degradation of plant cell wall polysaccharides by two human symbionts. PLoS Biol. 9, e1001221 (2011).
8. Hartemink, R., Vanlaere, K. M., J. Mertens, A. K. C. & Pombouts, F. M. Fermentation of xyloglucan by intestinal bacteria. Anaerobe 2, 223–230 (1996).
9. Kootte, R. S. et al. The therapeutic potential of manipulating gut microbiota in obesity and type 2 diabetes mellitus. Diabetes Obes. Metab. 14, 112–120 (2012).
10. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 124, 837–848 (2006).
11. van Nood, E. et al. Duodenal infusion of donor feces for recurrent Clostridium difficile. N. Engl. J. Med. 368, 407–413 (2015).
12. Petrov, E. et al. Stool substitute transplant therapy for the eradication of Clostridium difficile infection: “RePOPulating” the gut. Microbiome 1, 3 (2013).
13. Tasse, L. et al. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. Genome Res. 20, 1605–1610 (2012).
14. Cummings, J. H. & Macfarlane, G. T. Role of intestinal bacteria in nutrient metabolism and health. Proc. Nutr. Soc. 56, 495–504 (1997).
15. McNeil, N. I. The contribution of the large-intestine to energy supplies in man. Am. J. Clin. Nutr. 39, 338–342 (1984).
16. Smith, M. I. et al. Gut microbiomes of Malawian twin pairs discordant for kwashokorker. Science 339, 549–554 (2013).
17. Xu, J. et al. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. Science 299, 2074–2076 (2003).
18. Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464, 59–65 (2010).
19. Huttenhower, C. et al. Structure, function and diversity of the healthy human microbiome. Nature 486, 207–214 (2012).
20. Nelson, K. E. et al. A catalog of reference genomes from the human microbiome. Cell 148, 1254–1270 (2012).
21. Yaratoya, K. & Shirakawa, M. Xyloglucan: structure, rheological properties, biological functions and enzymatic modification. Curr. Trends Polym. Sci. 8, 72–77 (2003).
22. Hihe, Y. S. Y. & Harris, P. J. Xyloglucans of monocotyledons have diverse structures. Mol. Plant 2, 943–965 (2009).
23. Mello, L. V., Chen, X. & Rigden, D. J. Mining metagenomic data for novel domains: BACON, a new carbohydrate-binding module. FEBS Lett. 584, 2421–2426 (2010).
24. Cameron, E. A. et al. Multidomain carbohydrate-binding proteins involved in Bacteroides thetaiotaomicron starch metabolism. J. Biol. Chem. 287, 34614–34625 (2012).
25. Nakjang, S., Ndeh, D. A., Wipat, A., Bolam, D. N. & Hirt, R. P. A novel extracellular metalloprotease domain shared by animal host-associated mutualistic and pathogenic microbes. PLoS ONE 7, e30287 (2012).
26. Aspeborg, H., Coutinho, P. M., Wang, Y., Brummer, H. & Henrissat, B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol. Biol. 12, 186 (2012).
27. Oostrazer, T. M. et al. Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12. J. Biol. Chem. 282, 19177–19189 (2007).
28. Hehemann, J. H., Kelly, A. G., Pudlo, N. A., Martens, E. C. & Boraston, A. B. Bacteria of the human gut microbiota catalyze red seaweed glycans into carbohydrate-active enzyme updates from extrinsic microbes. Proc. Natl Acad. Sci. USA 109, 19786–19791 (2012).
29. Hehemann, J. H. et al. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature 464, 908–912 (2010).
30. Wegmann, U. et al. Complete genome of a new Firmicutes species belonging to the dominant human colonic microbiota (“Ruminococcus bicircularis”) reveals two chromosones and a selective capacity to utilize plant glucans. Environ. Microbiol. http://dx.doi.org/10.1111/1462-2920.12117 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to the following colleagues for providing materials or experimental assistance: F. Villalpana (neutral sugar analysis of XyG preparations), G. Sundqvist (protein mass spectrometry), F. Buttin (aryl glycoside syntheses), S. Prexler (protein production and purification), S. Tuomivainen and W. York (production of acetylated tomato XyG samples), the staff at the University of Michigan Germfree Laboratory (technical assistance with gnotobiotic mouse experiments) and the staff at the Diamond Light Source (provision of data collection facilities). Work in Stockholm was supported by the Mijutani Foundation for Glycoscience, The Swedish Research Council Formas (via CarboMat—the KTH Advanced Carbohydrate Materials Centre), The Swedish Research Council (Vetenskapsrådet; salary support to H.B.), and the Wallenberg Wood Science Centre (salary support to O.S. and L.S.M.). Work in Vancouver was supported by faculty funding from the Michael Smith Laboratories, University of British Columbia; the Natural Sciences and Engineering Research Council of Canada (Discovery Grant); the Canada Foundation for Innovation and the British Columbia Knowledge Development Fund; and the Department of Biotechnology and Biological Sciences Research Council under reference BB/I014802/1. Work in Ann Arbor was supported by National Institutes of Health grants DK084214 and GM099513. T.E.R. was supported in part by the Global Probiotics Council Young Investigator Grant for Probiotics Research awarded to E.C.M.

Author Contributions J.L. performed gene cloning, recombinant production and biochemical/ enzymatic characterization for all enzymes. T.E.R. constructed B. ovatus genetic mutants and tested mutant growth phenotypes. G.R.H. performed all protein X-ray crystallography. L.S.M. performed enzyme kinetic analyses and product analyses on select enzymes and substrates (BoGH3A and 3B, BoGH5A, BoGH9A, BoGH43A and 43B). A.S.T. performed all experiments and data analysis relating to the BACON domain and carbohydrate-binding proteins. G.S. and S.K. performed initial gene cloning and production of all enzymes, and enzymatic characterization of BoGH5A and BoGH2A. N.A.P. performed high-throughput screening of various Bacteroides strains, including B. ovatus deletion mutants (BoGH3A2), on XyG oligo- and polysaccharides and analysis in vivo analysis of active enzyme updates from extrinsic microbes.

We are grateful to the following colleagues for providing valuable materials and expertise: A. N. P. performed growth profiling of various Bacteroides strains, including B. ovatus deletion mutants (BoGH3A2), on XyG oligo- and polysaccharides and analysis in vivo competition data by qPCR. K. J. U. analysed growth data from 292 Bacteroides isolates on XyG and other substrates, and assisted with metagenomic surveys. N.M.K. provided advice and assistance on XyGUL recombinant carbohydrate-binding protein production, A.L.C. and C.A.H. assisted with calorimetry and data analysis, A.G.K. assisted with comparative genomic locus identification and performed Bacteroides phylogenetic analysis. S.N.C. assisted with recombinant production of all enzymes and performed stability studies. E.C.M. constructed the B. ovatus ΔtktA strain and the N-terminal lipidation mutant, performed corresponding cellular localization and growth studies, and performed comparative genomic analyses and metagenomic surveys. H.B., E.C.M. and G.J.D. conceived the study, directed research, analysed data, and wrote the article, including significant data analysis and writing input from J.L., T.E.R. and L.S.M.

Author Information Atomic coordinates and structure factors for the BoGH5A-XXG complex structure have been deposited at the PDB under accession 3ZMR. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.B. (brummer@msl.ubc.ca), E.C.M. (emartens@umich.edu) or G.J.D. (gideon.davies@york.ac.uk).

©2014 Macmillan Publishers Limited. All rights reserved.
METHODS

Growth analysis of Bacteroides species and reverse genetics. The locus tags of the genes involved in the present study and the corresponding DNA primers used for knockout, sequence alteration and qPCR studies are given in Supplementary Table 1. To construct gene deletions in *B. ovatus* strain ATCC 8483, a deletion of the gene encoding thymidylate kinase (*tdk*; Bacova_03071) was first constructed using an identical strategy to that used previously for *B. thetaotaomicron*41. All subsequent gene deletions and sequence modifications were conducted in a Δtdk strain background by allelic exchange using the vector pExchange-tdk7 and primers listed in Supplementary Table 1. Wild-type *B. ovatus* ATCC 8483, all mutant derivatives and other Bacteroides species tested were grown in tryptone yeast extract glucose (TYG) medium, brain-heart infusion agar supplemented with 10% horse blood, or minimal media (MM) supplemented with appropriate carbohydrates as previously described31. Antibiotics were added as needed: gentamicin (200 μg ml⁻¹), erythromycin (25 μg ml⁻¹) and 5-fluoro-2-deoxyuridine (200 μg ml⁻¹). To prepare cells for exposure to glycans, wild-type and mutant *B. ovatus* were grown in TYG, subcultured into MM-glucose, grown to mid-exponential phase (A₆₀₀₅₉₀ nm 0.6–0.8), then washed and resuspended in 2X MM before addition of the appropriate glycan. XyG substrates are described later and all carbohydrate stocks were prepared at 10 mg ml⁻¹ in diH₂O and sterilized by autoclaving. All quantitative growth was performed at 37 °C in an anaerobic chamber ( Coy Manufacturing, 5% CO₂ and 95% N₂) in an automated plate-reading device as described previously. In instances where bacterial growth data are either used to show growth curves or quantify differences in growth ability, strains were grown in 2–3 biological replicates. Fluorescence microscopy was performed on fixed *B. ovatus* cells grown to early exponential phase (A₆₀₀₅₉₀ nm 0.25–0.35) in MM containing a 9:1 mixture of limit-digest XGOS to slightly longer (average degree of polymerization (dp) = 14) XGOS. These conditions circumvented the reduced growth rate of the AGHS mutant on limit-digest XGOS. Cells were fixed in formalin and stained with a polyclonal antibody raised in rabbit against purified recombinant BoGH5A (Cocalico Biologicals), following the method previously reported32. The same anti-serum was used to probe GH5 presence by western blot.

Construction of a Bacteroides phylogeny. Six Bacteroides core genes that encode anthuranilate synthase, glucose-6-phosphate isomerase, glycerol-3-phosphate dehydrogenase, DNA-directed RNA polymerase subunit β, chorismate synthase, 3-dehydroquinase synthase and 16S rDNA were used to construct a phylogenetic tree for each species. Nucleotide sequences were concatenated in the same order, aligned using CLUSTAL-W33, and the resulting alignment was used to construct a maximum parsimony tree with bootstraps using MEGA5. Analysis of metagenomic data sets for presence of XyGULs. Human metagenomic data sets46,50,57–59 were searched by BLAST for the presence of XyGUL nucleotide sequences from *B. ovatus* (27,767 kb), *B. uniformis* (28,433 kb for PUL1, 59,874 kb for PUL2), *B. cellulosolitica* (26,173 kb) and *B. flaccum* (25,357 kb), plus the porphyran utilization PUL from *B. plebeius* (54,476 kb)60. Each BLAST probe was first searched against the NCBI Refseq genomes database to determine the background thresholds for BLAST hits to other sequenced genomes that do not contain a XyGUL using a word size of 11; this analysis failed to reveal any hits with E values < 20, and nucleotide identities >90% over a length >75 bp. Thus, in subsequent searches, we considered a metagenome to be positive for a particular XyGUL probe if it returned two or more hits ≥100 bp in length with ≥90% identity and E value ≤ 20. Because the most recent HMP metagenomic data sets assembled much larger contigs than previous studies (often aligning to over 20 continuous kb of the respective XyGUL query), we considered hits in these data sets to be positive if they harboured just one hit that was ≥10 kb with the same identity and E-value cut-offs listed earlier.

Germ-free mouse experiments and diets. All animal experiments were approved by the University Committee on Use and Care of Animals at the University of Michigan and were supervised by a veterinarian. Germ-free mouse experiments were conducted in a total of 12, 6–8-week-old male and female Swiss Webster mice from two litters that were born 9 days apart. Mice were randomly assigned into four groups of four animals (each was considered to be a single biological replicate) by a technician who was not familiar with the project. The investigators were not blinded to the identities of the treatments during the group experiments. Four mice per group was chosen as the sample size based on previous studies61,62 that used similar dietary treatment regimens and found significant alterations in competitive index using 3–7 animals per group (see Extended Data Fig. 6 legend for a description of statistical tests). A custom XyG-rich diet was constructed by pureeing equal amounts by wet weight of locally purchased Roman tomatoes, lettuce leaves, common aubergine with skins, and green bell peppers with seeds. This mixture was brought to a gentle simmer for 1 h by heating in a stainless steel vessel to simulate the cooking process for these items in human foods. The mixture was cooled and combined at a ratio of 4:1 (wet weight mixed/proportion weight mouse diet) with a custom polysaccharide-free rodent diet that contained sucrose as the only additional carbohydrate (Harlan Teklad)39. This diet mixture was dried into biscuits and autoclaved before feeding to germ-free mice. A different custom ‘polysaccharide-free diet’ (Harlan Teklad TD.13043) 27% protein, 44.5% glucose, 15% fat and 8% cellulose; because *B. ovatus* cannot grow on cellulose it is assumed that this diet does not contribute usable polysaccharides) was used for longer-term feeding studies and served as the base diet for inoculating and stabilizing the input mixture of wild-type and ΔXyGUL *B. ovatus* strains. The relative ratios of competing *B. ovatus* strains were measured from total DNA extracted from freshly voided faecal pellets as previously described32.

GPhs. The genes of the GPhs of this PUL were amplified by PCR using forward primers including NdeI restriction sites and reverse primers including XhoI restriction sites, except for GH43B, which had Sall restriction sites in the reverse primers (Supplementary Table 2). Constructs truncated to exclude predicted signal peptides and N-terminal lipidation sites42 were generated. The PCR products were digested with NdeI and XhoI (NdeI and Sall for GH43B), and ligated into similarly digested PET21a vectors, followed by transformation into electrocompetent *E. coli* TOP10 cells. The cells were grown in lysogeny broth (LB) overnight and plasmids were extracted using the MiniPrep kit (Qiagen) and sequenced (Eurofins MWG Operon) to identify positive clones. Plasmids containing the PUL hydrolase genes were transformed into *E. coli* BL21 (DE3) cells by electroporation. The cells were grown at 37 °C with shaking in Terrific Broth containing 100 μM ampicillin to an A₆₀₀₅₉₀ nm of 0.4–0.6, when protein expression was induced by addition of 0.2 mM isopropyl β-D-galactopyranoside (IPTG) and the temperature was lowered to 25 °C. Protein expression continued for 2–3 days, after which the cells were collected by centrifugation at 4,000g for 10 min. The cells were resuspended in buffer A (20 mM sodium phosphate pH 7.4, 500 mM sodium chloride, 20 mM imidazole) and disrupted by either sonication or by passing twice through a French press, followed by centrifugation at 27,000g for 45 min. The supernatant was loaded onto 5 ml Hitrap IMAC FF columns (GE Healthcare) using an AKTA FPLC (GE Healthcare) and washed thoroughly with buffer A. Each protein was purified on a separate column to eliminate the risk of cross-contamination.Tagged proteins were eluted using a linear gradient of 0–100% buffer B (20 mM sodium phosphate pH 7.4, 500 mM sodium chloride, 500 mM imidazole) over typically four column volumes. Eluted proteins were concentrated using Amicon Ultra centrifugal filters (Millipore) and further purified if needed by size-exclusion chromatography on a HiPrep 26/60 Sephacryl S-200 column into 50 mM sodium phosphate, pH 7.0. Liquid chromatography electrospray ionisation mass spectrometry (LC-ESI-MS) was used to verify the correct protein mass of purified proteins as described previously (data not shown)44. To improve the stability and solubility of certain of the enzymes, additives were included in the protein stock buffers, and these were present for all assays; BoGH5A was stored in 300 mM ammonium sulphate, whereas BoGH3A was stored in 20% glycerol.

Non-catalytic proteins and BoGH5A truncation variants. Expression vector PET21a (Novagen) containing BoGH5A (pBOGH5) was used for subsequent cloning of the BACON domain (residues 37–137 from transcription start), the GH5 catalytic domain (residues 138–502 from transcription start), and two active-site mutants, generating pBACON, pCAT, pBOGH5-E430A and pCAT-E430A, by using the complete genomic DNA encoding BoGH5A and GH5 domains amplified by PCR using forward primers including NdeI restriction sites and reverse primers including XhoI (Supplementary Table 3). Both contained a C-terminal (His) tag, and a TEV-cleavage site was added before the tag at the C-terminal end of the BACON domain insert. The catalytic nucleophile mutants of full-length BoGH5A and the GH5 domain were generated using Q5 High-Fidelity DNA Polymerase (New England Biolabs) from pBOGH5 and pCAT. The gene fragments corresponding to amino acids (from transcription start) 28–546 of SusD-like Bacova_02651 and 34–489 of Bacova_02650 gene products were amplified from *B. ovatus* genomic DNA by PCR using forward primers including NdeI restriction sites and reverse primers including SalI restriction sites, and the gene products were ligated into a modified version of pET-28a (EMD Biosciences) containing a recombinant tobacco etch virus (rTEV) protease recognition site. Heterologous protein production in *E. coli* BL21 and purification was subsequently performed essentially as described for the GH5s of the XyGUL, SDS–PAGE was used to confirm the purity of the proteins.

MALDI-TOF analysis of oligosaccharides. XyGO products were analysed by MALDI-TOF mass spectrometry on an LTQ Plus mass spectrometer (SAI) operated by the MALDI Mainframe2, Maldi Control software (version 1.03.51; SAI). A matrix of 2,5-dihydroxybenzoic acid (DHB) (10 mg ml⁻¹ in water) was used. The number of laser pulses was varied and the resulting individual spectra were summed to optimize signal-to-noise.

Analytical carbohydrate chromatography. Oligo- and monosaccharides were analysed by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAECPAD) on a Dionex ICS-3000 HPLC system operated by Chromelon software version 6.80 (Dionex) using a Dionex CarboPac PA200
column. Solvent A was water, solvent B 1 M sodium hydroxide and solvent C 1 M sodium acetate. Depending on the analytes, the following gradients were used. Gradient A: 0–5 min, 10% B, 2% C, 5–12 min, 10% B and a linear gradient from 2–30% C; 12–11.5 min, 50% B, 50% C; 12.1–13 min, an exponential gradient (curve setting 9) of B and C back to initial conditions; 13–17 min, initial conditions. Gradient B: 0–4 min, 10% B, 6% C; 4–17 min, 10% B and a linear gradient from 0–25% C; 17–17.1 min, 50% B, 50% C; 17.1–18 min, an exponential gradient (curve setting 9) of B and C back to initial conditions; 18–22 min, initial conditions. Gradient C: columns pre-conditioned before injection by -13 to -3 min, 12% B, 6.8% C; 3 – 0 min, 100% A; 0–25 min, 100% A.

Substrates. The 4-nitrophényl glycoside substrates used in this study (Glc-β-PNP, Glc-β-PNP, Gal-β-PNP, Gal-β-PNP, Xyl-β-PNP, Xyl-β-PNP, Tyr-α-PNP, Tyr-α-PNP) were purchased from Sigma-Aldrich. The 2-chloro-4-nitrophényl oligosaccharide substrates GGGG-β-CNP, XGGG-β-CNP and XLGG-β-CNP were synthesized as described previously and -Fuc-β-PNP was purchased from Carbo-synth. XyG (tamarind seed), arabinobioxy, barley mixed linkage glucan, curdian, galactan, gluco- and galactomannan, lichenan, carboxymethyl cellulose and hydroxyethyl cellulose, XXXG, isopimaran, cello-oligosaccharides, and XyG0s with a higher degree of polymerization (average degree of polymerization = 14) were purchased from Megazyme.

A mixture of oligosaccharides from tamarind XyG was produced by dissolving 1 g of the polysaccharide in 100 ml water at 55 °C under vigorous stirring. The temperature was lowered to 37 °C and the endo-xylanoglucanase BoGH5A (1 μL; see later) and phosphate buffer (to 50 mM, pH 7.0) was added, and the reaction proceeded overnight to completion, as assessed by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) and LC-ESI-MS. The solution was passed through a Biogel-P2 column using water as an eluent to remove phosphate buffer. Fractions containing XyG0s were collected, dialysed extensively to remove salts in deionized water. The resulting hemicellulose polysaccharides were removed by repeated 70% ethanol wash and filtration steps. The neutralized solution was dialysed extensively to remove salts in deionized water. The resulting hemicellulose fractions were each incubated with 135 units of the Cellobiose xylanase CjCBM22–Xyn10A and 45 units of mannanase 26A (both purchased from NZYtech) for 16 h at 37 °C, to hydrolyse contaminating xylan and mannan, respectively. The final yield of tomato XyG was 170 mg, and of lettuce XyG was 180 mg.

Before analysis by HPAEC-PAD (gradient C), the products (1 mg) were hydrolysed by incubation with 2 M trifluoroacetic acid (TFA) for 3 h at 120 °C (1 ml total volume). Hydrolysis products were vacuum dried and resuspended in deionized water, then filtered (0.2 μm filter). The tomato XyG comprised Glc, Xyl, Gal and Ara. The lettuce XyG comprised Glc, Xyl, Gal and Fuc. Both materials were found to also contain GaLA and Rha, at levels indicating pectin contamination of up to 50%. An anion exchange method adapted from one previously described was used to remove pectins. Dry material (84 mg lettuce XyG and 63 mg tomato XyG) was redissolved in 10 mM imidazole–HCl buffer, pH 7.0. This solution was mixed thoroughly with Q-separase fast-flow resin (Sigma–Aldrich), to which anionic pectic polysaccharides and any proteins bound. The flow-through, containing neutral polysaccharides (XyGs and any trace remaining amounts of mannans and xylangs), was collected by applying to a vacuum. Neutral sugar analysis showed a marked decrease in pectin contamination, to around 5%.

One hundred milligrams of crude XyG polysaccharide obtained from tomato or lettuce was dissolved in water (20 ml) and incubated with 0.2 μM of BoGH5A in 50 mM phosphate buffer (pH 7.0) at 37 °C overnight. Ethanol was added to a final concentration of 70% to precipitate remaining polysaccharides, and the soluble oligosaccharide products were recovered by filtration and re-dissolved in the filter. The XyG0s were analysed by HPAEC-PAD (gradient B) and MALDI-TOF. The final yield was 64 mg from the tomato XyG and 58 mg from the lettuce XyG.

Assays. All assays were carried out at 37 °C, at or near the buffer (50 mM) and pH optimum of the individual enzyme, unless otherwise stated. Curve fitting was performed using Origin 8 (OriginLab). For all quantitative enzyme assays at least two technical replicates were performed at each assay condition (for example, pH value, substrate concentration, and so on).

Phenol spectrophotometric assay (DNSA) versus a standard curve of Glc (1–6 mM). Reaction samples were added to an equal volume of 200 mM NaCO3 to terminate the reactions by raising the pH to 11.0. An extinction coefficient of 18,500 M−1 cm−1 was used to calculate product concentration from absorbance values. For some very slow reactions, a continuous assay was used in a Cary 300 spectrophotometer; an extinction coefficient of 10,500 M−1 cm−1 (pH 7.0) was used.

Assays using CNP-glycosides were monitored continuously for the release of 2-chloro-4-nitrophenolate using a Cary 300 spectrophotometer (Agilent Technologies). An extinction coefficient of 12,936 M−1 cm−1, determined from a standard curve, was used to calculate product concentration from A490 values. Determination of reducing sugars. Hydrolysis of polysaccharide substrates was measured by an increase in reducing sugars, using the 3,5-dinitrosalicylic acid reducing-sugar assay (DNSA) versus a standard curve of Glc (1–6 mM). Reaction samples were added to an equal volume of 2% (w/v) sodium citrate pH 6.22, which was read continuously using a Cary 300 spectrophotometer.

Assays for the detection of specific monosaccharide release. A linked galactose dehydrogenase/galactose mutarotase assay (Megazyme) was used to quantify the release of galactose or arabinose from XyG or XyG0s. The release of galactose or arabinose led to the stoichiometric reduction of NAD+ to NADH, giving an increase in A340 (e = 6,230 M−1 cm−1 at pH 7.0)M, which was read continuously using a Cary 300 spectrophotometer. A second linked assay (Glucose/Mannose/Fructose detection kit, Megazyme) was used to quantify the release of glucose monosaccharides. The protocol provided in the manufacturer’s instructions was modified for use as a continuous assay. The release of a glucose monosaccharide corresponds stoichiometrically with the reduction of a molecule of NAD+ to NADH, which leads to an increase in A340 (e = 6,220 M−1 cm−1)M, observed continuously using a Cary 300 spectrophotometer. Reactions were carried out in the triethyramine buffer (pH 7.6) provided with the assay kit.
Assays using HPAEC-PAD analysis. Enzymatic reactions on xylogluco- and cello-oligosaccharides were performed in 50 μl of 50 mM buffer and were stopped by addition of 2 μl of 5 M sodium hydroxide before HPAEC-PAD analysis. Gradient A was used for reactions with XXXG or XLLG and gradient B was used for assays against XyG polysaccharides and cello-oligosaccharides.

Affinity gel electrophoresis. Affinity gel electrophoresis was performed for 90 min at room temperature (20 °C) on non-denaturing 10% (w/v) polyacrylamide gels, essentially as previously described.

Isothermal titration calorimetry. Isothermal titration calorimetry (ITC) was performed essentially as previously described using a MicroCal VP-ITC calorimeter. Proteins were dialysed into 20 mM HEPES 100 mM NaCl pH 7.0 and 2 mM imidazole) and cells were lysed using three 15 s pulses of sonication at maximum amplitude using an MSE Soniprep 150 sonicator. Cell debris was removed by centrifugation at 5,000 r.p.m. and the resulting heat of reaction was recorded. Data were analysed using the Origin software program.

BoGH5A preparation for crystallization. Typically ~5 g of cell paste were resuspended in 5 X volumes of buffer C (50 mM HEPES pH 7.0, 0.5 M NaCl, 30 mM imidazole) and cells were lysed using three 15 s pulses of sonication at maximum amplitude using an MSE Soniprep 150 sonicator. Cell debris was removed by centrifugation at 38,000 for 30 min at 4 °C. The supernatant was applied to a 5 ml GE Healthcare Hitrap Nickel NTA column that had been equilibrated in buffer C at 10 °C using an Äktä Xpress purification system. The column was washed with five column volumes of buffer C before a 100 ml gradient of buffer D (buffer C plus 300 mM imidazole) was applied to elute the protein. Fractions of 1.6 ml were collected along the gradient. Peak fractions containing the enzyme were pooled and concentrated using a 30 kDa molecular weight cut-off Savartis concentrator by centrifugation at 5g to a volume of approximately 1 ml. The protein was then applied to a GE Healthcare 16/60 HiLoad Superdex 200 column, which had been equilibrated with gel filtration buffer (10 mM Tris pH 8.0, 250 mM NaCl). After passage of the void volume (45 ml), 1.6 ml fractions were collected. Fractions containing BoGH5A were combined and buffer exchanged into 10 mM Tris pH 8.0, 50 mM NaCl before final concentration to 53 mg ml⁻¹ as judged by A₂₈₀nm using an extinction coefficient of 79,300 M⁻¹ cm⁻¹ and a molecular weight of 51,674 Da. To prepare the protein-XyGO complex, 53 mg ml⁻¹ of the protein was mixed with 100 mM XyGO (10% XLLG, 90% XXLG) to give final concentrations of 48 mg ml⁻¹ protein and 10 mM XyGO.

Structure solution and refinement. Crystallization trials were set up in sitting drops using a Mosquito robot (TTP Labtech). Diffraction-quality crystals were obtained in 0.1 M Bis-Tris pH 6.5, 20% w/v polyethylene glycol monomethyl ether 5,000. Crystals were cryo-protected by soaking in mother liquor supplemented with 20% (v/v) ethylene glycol for 30 s before flash freezing in liquid N₂.

X-ray diffraction data were collected at the Diamond Light Source synchrotron, station I03. Data were processed using XDS. The structure was determined by molecular replacement in PHASER using chain A of the P. paludis GH5 (PDB accession 2IEP) as the search model. Two molecules were located in the asymmetric unit with log-likelihood gains of 99 and 364 for the first and second molecules, respectively. The crucial step in model building was to harness three cycles of phase improvement, including NCS averaging, in PHASER, before model construction using ARP/wARP. Subsequent model building and refinement were performed using REFMAC and Coot, respectively, with the validity of the model monitored using MolProbity before deposition (PDB accession 3ZMR).

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

32. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe 4, 447–457 (2008).

33. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680 (1994).

34. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577 (2007).

35. Tamura, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739 (2011).

36. Turnbaugh, P. J. et al. A core gut microbiome in obese and lean twins. Nature 457, 480–484 (2009).

37. Kurokawa, K. et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res. 14, 169–181 (2007).

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

39. Kamada, N. et al. Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. Science 336, 1325–1329 (2012).

40. Juncker, A. S. et al. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. 12, 1652–1662 (2003).

41. Sundqvist, G., Stenval, M., Berglund, H., Ottosson, J. & Brumer, H. A general, robust method for the quality control of intact proteins using LC-ESI-MS. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 852, 188–194 (2007).

42. Ibitubin, F. M., Baumann, M. J., Greffe, L. & Brumer, H. Kinetic analyses of retaining endo-xyloglucanases from plant and microbial sources using new chromogenic xylogluco-oligosaccharide aryl glycosides. Biochemistry 47, 7762–7769 (2008).

43. Martínez-Fleites, C. et al. Crystal structures of Clostridium thermocellum xyloligo-oligosaccharide XGH74A, reveal the structural basis for xyloligocan recognition and degradation. J. Biol. Chem. 281, 24922–24933 (2006).

44. Greffe, L., Bassuelle, L., Bulone, V. & Brumer, H. Synthesis, preliminary characterization, and application of novel surfactants from highly branched xyloligocan oligosaccharides. Glycobiology 15, 437–445 (2005).

45. Mopper, K. & Gindler, E. A new noncorrosive dye reagent for automatic sugar chromatography. Anal. Biochem. 56, 440–442 (1973).

46. McFeaters, R. F. A manual method for reducing sugar determinations with 2,2'-bichinoninic reagent. Anal. Biochem. 103, 302–306 (1980).

47. Brumer, H., Sims, P. F. G. & Sinnott, M. L. Lignocellulose degradation by Phanerochaete chrysosporium: purification and characterization of the main β-galactosidase. Biochem. J. 339, 43–53 (1999).

48. Cartmell, A. et al. The structure and function of an arabian-specific α-L-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases. J. Biol. Chem. 286, 15483–15495 (2011).

49. Miller, G. L. The use of dinitrosalicylic acid for the determination of reducing sugar. Anal. Chem. 31, 426–428 (1959).

50. Freeove, A. C. J., Bolam, D. N., White, P., Hazlewood, G. P. & Gilbert, H. J. A novel carbohydrate-binding module of the clostridial xyloglucan-specific carbohydrate binding modules. BMC Biotechnol. 9, 32 (2009).

51. Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. Carbohydrate-binding protein is a component of the plant cell wall-degrading complex of Pichromyces chrysosporium. BMC Bacteriol. 9, 28 (2009).

52. Maceda, J. E., O’Dwyer, M. R., Nutman, R. & Linstrom, V. S. Automated protein model building combined with iterative structure refinement. Nature Struct. Biol. 6, 458–463 (1999).

53. Mambudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D 53, 240–255 (1997).

54. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D 60, 2126–2132 (2004).

55. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 35, W375–W383 (2007).
Extended Data Figure 1 | Evolution of the genomic region containing the XyGUL and corresponding growth on XyG as a sole carbon source.

**a**, Genomic organization of 11 representative strains from 3 different species of *Bacteroides*. **b**, Growth of these strains on tamarind XyG with glucose and xylose as controls (average of \( n = 2 \) growths per strain). The observation that one *B. ovatus* strain (SD CMC 3f) lacks a corresponding XyGUL, as do all *B. xylanisolvens* (the closest cultured relative of *B. ovatus*), suggests that the XyGUL entered *B. ovatus* after it diverged. Also note that two other unrelated flanking PULs show variable presence at this locus, suggesting that it is a ‘hotspot’ for PUL evolution.
**Extended Data Figure 2** | Activation of putative XyGULs in Bacteroidetes isolates via growth on XyG. 

**a.** Sentinel susC-like gene expression ($n = 3$) from separately grown cultures relative to a minimal medium plus glucose control. 

**b.** Growth profiles of the corresponding strains in medium containing tamarind XyG with glucose and xylose as controls (average of $n = 2$ growths per strain). Error bars represent standard error of the mean.
Extended Data Figure 3 | Cell-surface localization of BoGH5A and effect of localization on B. ovatus growth.  a, Staining of fixed wild-type and mutant B. ovatus strains. b, Western blot indicating that BoGH5A is still produced in the C1A lipidation site mutant, albeit in multiple degraded forms (residue number corresponds to the mature protein, equivalent to C33 numbered from translation start). c, Growth of wild-type (positive control), ΔBoGH5A (negative control) and BoGH5A-C1A strains on tamarind XyG. The BoGH5A-C1A mutant exhibits -2.6-fold slower exponential growth than the wild-type. Vertical error bars on each curve indicate the standard deviation (s.d.) of the mean (n = 3 replicates). The residual growth ability of the BoGH5A-C1A strain, despite mislocalization, is unlikely to be explained by the presence of BoGH5A enzyme accumulation in the supernatant, which was only detected by western blot for wild-type bacteria expressing BoGH5A on the cell surface. Detection of BoGH5A in panels a and b was achieved with a rabbit polyclonal antibody raised against the recombinant protein produced in E. coli (representative data from two experiments each that yielded very similar results).
Extended Data Figure 4 | Non-catalytic interaction of BoGH5A variants, SusD-like Bacova_02651 and Bacova_02650 of the XyGUL with polysaccharides. a, SDS–PAGE of recombinant proteins (representative data from at least three preparations for each protein are shown). b, Affinity gel electrophoresis (representative data from at least two gels for each experimental condition). c, Isothermal titration calorimetry (ITC); the top graph in each pair shows the raw heat during titration, whereas the bottom graph shows the integrated heats after correction. d, Association constants and thermodynamic parameters obtained from ITC data. Bovine serum albumin (BSA) was used as a non-interacting negative control protein. Other protein names were abbreviated as follows: BACON, residues Asp 37–Tyr 137 corresponding to the BACON domain of BoGH5A; Bacova_02651, SusD-like XyGUL gene product; Bacova_02650, SusE-positioned XyGUL gene product; BoGH5A E430A, full-length (Asp 37–Asn 502) catalytic nucleophile mutant of BoGH5A; CAT E430A, catalytic nucleophile mutant of the BoGH5A catalytic domain only (Ile 138–Asn 502). Reducing-sugar assays confirmed that the catalytic mutants had no detectable hydrolytic activity on XyG (data not shown), whereas an active variant (that is, E430) of CAT had a twofold higher specific activity than the full-length, wild-type BoGH5A at saturating XyG concentrations (0.5–3 mM).
| Metagenomic project                              | Frequency (%) | Sequencing depth |
|------------------------------------------------|---------------|------------------|
| North American twins (healthy, n=18)           | 92.0          |                 |
| North American Human Microbiome Project (healthy, n=98) | 83.2          |                 |
| European MetaHit Project (healthy, n=86)       | 82.8          |                 |
| Japanese metagenome (healthy, n=11)            | 53.6          |                 |
| European MetaHit Project (ulcerative colitis, n=27) | 34.4          |                 |
| European MetaHit Project (Crohn's disease, n=12) | 1.2           |                 |
| B. plebeius porphyran PUL                       | 6.4           |                 |

Heatmap color key:
- 5 PLUs
- 4 PLUs
- 3 PLUs
- 2 PLUs
- 1 PLUs
- 0 PLUs
Extended Data Figure 5 | Abundance of *Bacteroides* XyGULs in human from a survey of metagenomic sequencing data from a total of 250 adult human samples. The samples were from 211 healthy individuals, 27 with ulcerative colitis and 12 with Crohn’s disease (see Methods for references). Data sets were individually queried by BLAST using the entire XyGUL nucleotide sequence from each of the four *Bacteroides* species listed at the top (compare with Fig. 2) and a PUL involved in degrading the red algal polysaccharide porphyran. Each horizontal line represents the presence or absence of a hit in a single individual. The leftmost column summarizes the total XyGUL content in each person (annotated according to the colour key in the top right corner). The XyGUL frequency across all 250 samples is shown at the bottom for each condition. The graph at the far right illustrates the variation in sequencing depth for each sample/study: black lines show the average depth in megabases (Mb) for each study; the light grey line shows the depth for each individual sample.
Presence of the XyGUL confers a fitness advantage to *B. ovatus* in the presence of dietary XyG, but only when other dietary polysaccharides are eliminated. 

**a.** MALDI analysis of BoGH5A-digested alkaline extract from a custom mouse diet that contained a large amount of XyG from natural vegetable sources (equal amounts of cooked bell pepper, aubergine, tomato and lettuce; see Methods), indicating the presence of both solanaceous arabinogalactoxyloglucan and fucogalactoxyloglucan. 

**b.** qPCR analysis of XyGUL sentinel gene expression in wild-type *B. ovatus* grown on extracted polysaccharides from the XyG-rich custom diet, demonstrating that it significantly activates XyGUL expression over a glucose control (error bars show the s.d. of the mean of three biological replicates for both growth conditions). 

**c.** *In vitro* growth of wild-type and ΔXyGUL *B. ovatus* strains in the polysaccharide extract from the XyG-rich diet, including glucose and tamarind XyG as positive and negative control substrates, respectively. Compared with growth on tamarind XyG (middle panel), the incomplete growth defect of the ΔXyGUL mutant on the food extract (right panel) indicates that the food contains other polysaccharides that are accessible by *B. ovatus*. Vertical error bars on each curve indicate the s.d. of the mean of three replicates.

**d.** *In vivo* competition between wild-type and ΔXyGUL *B. ovatus* strains in mice consuming various amounts of dietary XyG. All mice were initially fed a synthetic diet containing glucose as the sole digestible carbohydrate for 1 week and then gavaged with a 7:3 ratio of ΔXyGUL:wild type (based on independent culture optical densities, total of $10^8$ viable *B. ovatus*) and the communities were allowed to equilibrate for 3 days. Despite the initial ratio being biased in favour of the ΔXyGUL strain, the communities equilibrated in the range 5:5–4:6, but thereafter remained stable while mice were maintained on the XyG-free diet (blue boxes in three competition plots). After community stabilization, three different dietary regimens were analysed. Left, mice were maintained on the control diet (glucose only, devoid of XyG) between days 5–37, but switched to water containing 0.25% purified XyG for days 15–37 (grey box). Middle, mice were switched to a XyG-rich, custom diet from natural food sources while simultaneously drinking water containing 0.25% purified XyG (green box). These mice were then switched to the glucose-only, XyG-free control diet while remaining on water containing 0.25% XyG (grey box). Right, mice were switched to the XyG-rich diet between days 3 and 15 but given normal water (yellow box); these mice were not continued further on any dietary regimen. Maintenance on either XyG food/XyG water (middle panel) or XyG food only (right panel) does not exert a measurable fitness pressure on the competing wild-type and ΔXyGUL strains. However, when the complex natural food polysaccharides were withheld while 0.25% XyG was maintained in water, a clear fitness pressure was observed through the significant, sequential reduction of the ΔXyGUL mutant between days 15 and 37. These data suggest that although the XyGUL confers an advantage to *B. ovatus* by broadening its substrate range to include XyG, the presence of alternative oligo- and polysaccharides (for example, other hemicelluloses, pectins) in a complex vegetable-based diet is nonetheless sufficient to support strains lacking this locus *in vivo*. Each data point is the mean abundance of the indicated strain in four separate mice and error bars represent 1 s.d. Measurements conformed to a normal distribution based on the observation that 67% of all assay values were within 1 s.d. of their respective means. Asterisks indicate statistically significant alterations ($P < 0.01$; Student’s *t*-test, one-tailed, paired) in strain abundance relative to the day 15 samples, which immediately preceded the diet switch aimed at isolating XyG as the sole exogenous polysaccharide.
### Extended Data Table 1 | Summary of kinetic analyses of XyGUL glycoside hydrolases

| Enzyme | Substrate* | $k_{\text{cat}}$ (s$^{-1}$) | $K_m$ (mM)$^\dagger$ | $k_{\text{cat}}/K_m$ (s$^{-1}$ mM$^{-1}$)$^{\dagger\ddagger}$ | Assay          |
|--------|------------|-----------------------------|----------------------|------------------------------------------------|----------------|
| BoGH2A | Gal-β-PNP  | $4.0 \pm 0.092$             | $0.087 \pm 0.0074$   | 46.0                                              | PNP            |
|        | cellulbiose| Trace activity              |                      |                                                  |                |
|        | cellotetraose| $65.0 \pm 3.81$            | $2.49 \pm 0.42$     | 21.3                                              | Glc-kit       |
|        | cellobiose  | Trace activity              |                      |                                                  |                |
|        | cellohexaose| $0.62 \pm 0.15$            | $0.666 \pm 0.52$    | 0.90                                              | Glc-kit       |
|        | GLLG        | ND$^\ddagger$               | ND                   | 0.15                                              | Glc-kit       |
|        | GXXG        | ND$^\ddagger$               | ND                   | 3.41                                              | Glc-kit       |
| BoGH3A | Glc-β-PNP   | $1.74 \pm 0.21$            | $0.155 \pm 0.039$   | 11.2                                              | PNP            |
|        | Gal-β-PNP   | Trace activity              |                      |                                                  | PNP            |
|        | cellbiose   | $1.57 \pm 0.11$            | $0.68 \pm 0.16$     | 2.3                                               | Glc-kit       |
|        | cellotetraose| $3.99 \pm 0.31$            | $0.23 \pm 0.060$    | 17.3                                              | Glc-kit       |
|        | cellobiose  | Trace activity              |                      |                                                  |                |
|        | cellohexaose| $4.91 \pm 0.71$            | $0.47 \pm 0.21$     | 10.4                                              | Glc-kit       |
|        | GLLG        | ND$^\ddagger$               | ND                   | 0.18                                              | Glc-kit       |
|        | GXXG        | ND$^\ddagger$               | ND                   | 3.34                                              | Glc-kit       |
| BoGH5A | XXXG-β-CNP  | $10.5 \pm 0.14$            | $0.036 \pm 0.0027$  | 291.7                                             | CNP            |
|        | XLLG-β-CNP  | $11.1 \pm 1.13$            | $0.145 \pm 0.024$   | 76.5                                              | CNP            |
|        | GGGG-β-CNP  | $0.12 \pm 0.025$           | $3.59 \pm 1.27$     | 0.034                                             | CNP            |
|        | Tamarind XyG| $435.3 \pm 25.6$           | $0.82 \pm 0.17$ mg ml$^{-1}$ | 534.0 s$^{-1}$ mg$^{-1}$ ml | DNSA           |
|        | Lettuce XyG | ND$^\ddagger$               | ND                   | 543.1 s$^{-1}$ mg$^{-1}$ ml | DNSA           |
|        | Tomato XyG  | ND$^\ddagger$               | ND                   | 501.5 s$^{-1}$ mg$^{-1}$ ml | DNSA           |
| BoGH9A | Tamarind XG | Active (MALDI-MS data)     |                      |                                                  |                |
|        | Lettuce XG  | Active (MALDI-MS data)     |                      |                                                  |                |
|        | Tomato XG   | Active (MALDI-MS data)     |                      |                                                  |                |
| BoGH3A | Xyl-α-PNP   | $1.60 \pm 0.031$           | $7.7 \pm 0.273$     | 0.21                                              | PNP            |
|        | Glc-α-PNP   | $0.071 \pm 0.015$          | $31.8 \pm 5.1$      | 0.0022                                            | PNP            |
|        | XXYG        | $32.6 \pm 2.1$             | $0.223 \pm 0.046$   | 146.2                                             | HPLC           |
|        | XLLG        | $31.0 \pm 1.9$             | $0.378 \pm 0.075$   | 82.0                                              | HPLC           |
|        | Isoprimeverose| $1.78 \pm 0.21$           | $38.1 \pm 7.1$      | 0.047                                             | Glc-kit       |
| BoGH43A| L-Araf-α-PNP| $0.057 \pm 0.001$          | $0.71 \pm 0.03$     | 0.081                                             | PNP            |
|        | Xyl-β-PNP   | $0.26 \pm 0.005$           | $6.58 \pm 0.21$     | 0.039                                             | PNP            |
|        | Tomato XyGOs| ND$^\ddagger$               | ND                   | 0.013 s$^{-1}$ mg$^{-1}$ ml | GDh-kit        |
|        | Tomato XyG  | Active (MALDI-MS data)     |                      | 0.0024 s$^{-1}$ mg$^{-1}$ ml | GDh-kit        |
| BoGH43B| L-Araf-α-PNP| $5.0 \times 10^{-4} \pm 8.1 \times 10^{-5}$ | $6.6 \pm 2.3$     | $7.6 \times 10^{-5}$ s$^{-1}$ mg$^{-1}$ ml     | Cont. assay    |
|        | Tomato XyGOs| ND$^\ddagger$               | ND                   | $0.0024$ s$^{-1}$ mg$^{-1}$ ml | GDh-kit        |
|        | Tomato XyG  | Active (MALDI-MS data)     |                      |                                                  | GDh-kit        |

$^\dagger$ Data are only given for those substrates on which a given enzyme was active. For example, BoGH5A and BoGH9A were not active on arabinoxylan, barley mixed-linkage glucan, curdlan, galactan, gluco- and galactomannan, lichenan, carboxymethyl cellulose, or hydroxyethyl cellulose. The full panel of chromogenic and natural substrates against which enzymes were tested is described in Methods.

$^\dagger\ddagger$ These units apply to all values in the column, unless otherwise stated. Standard errors of means are indicated.

$^\ddagger$ For substrates for which individual $k_{\text{cat}}$ and $K_m$ values are not reported, $k_{\text{cat}}/K_m$ values were obtained from linear curve fitting to initial rate data in the $[S] < K_m$ regime, where the standard Michaelis–Menten equation reduces to $v_o = k_{\text{cat}}[E][S]/K_m$.

ND, not determined.

pH-rate profiles, full kinetic data plots, and carbohydrate product analyses are available in Supplementary Figs 1–20 for all enzymes.
Extended Data Table 2 | Growth of mutant *B. ovatus* strains on XyG relative to wild-type *B. ovatus* 8384

| Strain          | Relative Growth* (%) | Relative Rate (ΔA<sub>600</sub>/Δtime) | Relative Lag (Δtime<sub>400</sub>/Δtime<sub>600</sub>) |
|-----------------|----------------------|---------------------------------------|--------------------------------------------------------|
| ΔGH5 tamarind xyloglucan | No Growth            |                                       |                                                        |
| dp14 tamarind xyloglucan   | 139 ± 17†           | 2.2 ± 0.2†                           | 1.2 ± 0.1                                              |
| GH5-digested tamarind xyloglucan | 70 ± 9               | 0.6 ± 0.2‡                           | 1.2 ± 0.2                                              |
| XGOs (Megazyme)           | 71 ± 6‡              | 0.6 ± 0.1†                           | 1.3 ± 0.3                                              |
| ΔGH9 tamarind xyloglucan  | 119 ± 11             | 1.2 ± 0.2                            | 1.0 ± 0.1                                              |
| dp14 tamarind xyloglucan   | 114 ± 18             | 0.9 ± 0.1                            | 1.1 ± 0.1                                              |
| GH5-digested tamarind xyloglucan | 105 ± 6              | 1.0 ± 0.1                            | 1.3 ± 0.3                                              |
| XGOs (Megazyme)           | 95 ± 13              | 0.9 ± 0.1                            | 1.2 ± 0.2                                              |
| ΔGH31 tamarind xyloglucan | 18 ± 2†              | 0.06 ± 0.05†                         | 3.3 ± 0.8†                                              |
| dp14 tamarind xyloglucan   | No Growth            |                                       |                                                        |
| GH5-digested tamarind xyloglucan | No Growth           |                                       |                                                        |
| XGOs (Megazyme)           | No Growth            |                                       |                                                        |

* Growth conditions are described in Methods. Relative growth values are average per cent of wild-type *B. ovatus* 8384 growth density. No growth was defined as an increase of OD<sub>600</sub> < 0.05, measured from the initial baseline to the maximum achieved. Standard errors of the mean values for three biological replicates are indicated.
† P value = 0.05 as determined by unpaired t-test.
‡ P value = 0.05–0.1 as determined by unpaired t-test.
### Extended Data Table 3 | Data collection and refinement statistics for BoGH5A–XXXG complex

**BoGH5A:XXXG (PDB ID 3zmR)**

| **Data collection**          |  |
|------------------------------|---|
| Space group                  | P2₁ |
| Cell dimensions              |  |
| \(a, b, c\) (Å)              | 46.8, 147.2, 84.1 |
| \(\alpha, \beta, \gamma\) (°) | 90.0, 92.8, 90.0 |
| Resolution (Å)               | 72.99 - 1.43 (1.47 – 1.43) * |
| \(R_{merge}\)                | 0.058 (0.575) |
| \(I / \sigma\)              | 12.2 (2.2) |
| Completeness (%)             | 99.5 (99.8) |
| Redundancy                   | 2.1 (2.0) |

| **Refinement**               |  |
| Resolution (Å)               | 72.99 - 1.43 |
| No. reflections (Work/Free)  | 197432/10458 |
| \(R_{work} / R_{free}\)     | 0.12/0.16 |
| No. atoms                    |  |
| Protein                      | 7403 |
| Ligand/ion                   | 226 |
| Water                        | 1342 |
| Avg B-factors                |  |
| Protein                      | 19 |
| Ligand/ion                   | 38 |
| Water                        | 32 |
| R.m.s. deviations            |  |
| Bond lengths (Å)             | 0.011 |
| Bond angles (°)              | 1.46 |

* Values in parentheses are for highest-resolution shell.