Cell Reports

Molecular mechanism by which prominent human gut Bacteroidetes utilize mixed-linkage beta-glucans, major health-promoting cereal polysaccharides

--Manuscript Draft--

Manuscript Number:	CELL-REPORTS-D-17-01660R2
Full Title:	Molecular mechanism by which prominent human gut Bacteroidetes utilize mixed-linkage beta-glucans, major health-promoting cereal polysaccharides
Article Type:	Research Article
Keywords:	Polysaccharide utilization locus; Mixed-linkage beta-glucan; gut microbiota; Bacteroidetes; Carbohydrate-active enzyme; Glycoside hydrolase family 16
Corresponding Author:	Harry Brumer University of British Columbia Vancouver, BC CANADA
First Author:	Kazune Tamura
Order of Authors:	Kazune Tamura
	Glyn R Hemsworth
	Guillaume Dejean
	Theresa E Rogers
	Nicholas A Pudlo
	Karthik Urs
	Namrata Jain
	Gideon J Davies
	Eric C Martens
	Harry Brumer
Abstract:	Microbial utilization of complex polysaccharides is a major driving force in shaping the composition of the human gut microbiota. There is a growing appreciation that finely tuned polysaccharide utilization loci enable ubiquitous gut Bacteroidetes to thrive on the plethora of complex polysaccharides that constitute "dietary fiber". Mixed-linkage beta(1,3)/beta(1,4)-glucans (MLGs) are a key family of plant cell wall polysaccharides with recognized health benefits, but whose mechanism of utilization has remained unclear. Here, we provide molecular insight into the function of an archetypal MLG utilization locus (MLGUL) through a combination of biochemistry, enzymology, structural biology, and microbiology. Comparative genomics coupled with growth studies demonstrated further that syntenic MLGULs serve as genetic markers for MLG catabolism across commensal gut bacteria. In turn, we surveyed human gut metagenomes to reveal that MLGULs are ubiquitous in human populations globally, which underscores the importance of gut microbial metabolism of MLG as a common cereal polysaccharide.
Suggested Reviewers:	Antoni Planas antoni.planas@iqs.url.edu Expert in GH16 structure/function and MLG enzymology.
	Gurvan Michel gurvan.michel@sb-roscoff.fr Expert in bacterial polysaccharide utilization, genomics, structural biology (especially GH16 evolution).
	Mark McBride mcbride@uwm.edu Expert in Bacteroidetes genomics and polysaccharide utilization
	Harry Flint

	H.Flint@abdn.ac.uk Expert in the role of gut microorganisms in nutrition and health.
	Shinya Fushinobu asfushi@mail.ecc.u-tokyo.ac.jp Expert in carbohydrate enzyme structural biology of relevance to the human microbiota; http://dx.doi.org/10.1016/j.chembiol.2017.03.012
	Ed Bayer Ed.Bayer@weizmann.ac.il Expert in bacterial carbohydrate utilization systems, human gut microbiota
	Devin Rose drose3@unl.edu Expert in cereal polysaccharides and gut microbiota. http://foodsci.unl.edu/drose
	Birte Svensson bis@bio.dtu.dk Expert in polysaccharide enzymology, including in the human gut.
Opposed Reviewers:	

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

Michael Smith Laboratories
2185 East Mall
Vancouver, BC Canada V6T 1Z4

Phone +1 604 827 3738
Fax +1 604 822 2114
e-mail brumer@msl.ubc.ca

Professor
Harry Brumer, Ph.D.

11 September 2017

Ruth Zearfoss, Ph.D.
Associate Editor
Cell Reports

Dear Dr. Zearfoss,

Thank you for your email on September 5 informing me that our manuscript ("Molecular mechanism by which prominent human-gut Bacteroidetes utilize mixed-linkage beta-glucans, major health-promoting cereal polysaccharides"; CELL-REPORTS-D-17-01660) has been accepted for publication in *Cell Reports*. We herewith submit our individual manuscript, figure, and supplementary files for production.

Please note that although we have endeavoured to comply with all of the publication guidelines for *Cell Reports*, our manuscript is now over the recommended character count following revision according to the reviewers' original comments (however, the main text minus references is under the limit at 49000 characters including spaces). Also to comply with the reviewers' requests, we have also moved one supplemental figure to the main text (Fig. 5 in the current version), and added a new supplemental figure (Fig. S5 in the current version), which brings the total number of main-text figures and table to eight, with an equivalent number of supplemental figures. We hope that you will kindly grant us an exception on the text and figure limits, since the extra content has been included in direct response to specific reviewer comments. Indeed, we feel that these additions bring significant value, and that trimming the existing content to meet the character and figure guidelines would erode the overall presentation of the work.

Yours sincerely,

Prof. Harry Brumer

a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

Michael Smith Laboratories
2185 East Mall
Vancouver, BC Canada V6T 1Z4

Phone +1 604 827 3738
Fax +1 604 822 2114
e-mail brumer@msl.ubc.ca

Professor
Harry Brumer, Ph.D.

Reviewers' comments:

With reference to the email from Ruth Zearfoss, Associate Editor, on Sept. 5, there were no reviewer comments on our revised manuscript.

Molecular mechanism by which prominent human-gut Bacteroidetes utilize mixed-linkage beta-glucans, major health-promoting cereal polysaccharides

Kazune Tamura[1,2], Glyn R. Hemsworth[3], Guillaume Dejean[1], Theresa E. Rogers[4], Nicholas A. Pudlo[4], Karthik Urs[4], Namrata Jain[1,5], Gideon J. Davies[3], Eric C. Martens[4], Harry Brumer[1,2,5,6,7,*]

[1] Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC, V6T 1Z4, Canada

[2] Department of Biochemistry and Molecular Biology, University of British Columbia, 2350 Health Sciences Mall, Vancouver, BC, V6T 1Z3, Canada

[3] York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, United Kingdom

[4] Department of Microbiology and Immunology, University of Michigan Medical School, 1150 West Medical Center Dr, Ann Arbor, MI, 48109, USA

[5] Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, BC, V6T 1Z1, Canada

[6] Department of Botany, University of British Columbia, 3200 University Blvd, Vancouver, BC, V6T 1Z4, Canada

[7] Lead Contact

* Correspondence: Harry Brumer (brumer@msl.ubc.ca)

Summary

Microbial utilization of complex polysaccharides is a major driving force in shaping the composition of the human gut microbiota. There is a growing appreciation that finely tuned polysaccharide utilization loci enable ubiquitous gut Bacteroidetes to thrive on the plethora of complex polysaccharides that constitute "dietary fiber". Mixed-linkage β(1,3)/β(1,4)-glucans (MLGs) are a key family of plant cell wall polysaccharides with recognized health benefits, but whose mechanism of utilization has remained unclear. Here, we provide molecular insight into the function of an archetypal MLG utilization locus (MLGUL) through a combination of biochemistry, enzymology, structural biology, and microbiology. Comparative genomics coupled with growth studies demonstrated further that syntenic MLGULs serve as genetic markers for MLG catabolism across commensal gut bacteria. In turn, we surveyed human gut metagenomes to reveal that MLGULs are ubiquitous in human populations globally, which underscores the importance of gut microbial metabolism of MLG as a common cereal polysaccharide.

Introduction

The composition and homeostasis of the human gut microbiota has a profound and intimate connection to various aspects of our physiology, health and wellbeing (Littman and Pamer, 2011). Indeed, a multitude of diseases such as type 2 diabetes, inflammatory bowel diseases (IBDs), and cancer have been linked to alterations in the population and proportion of microbes in this highly complex and dynamic ecosystem that exists in our large intestine (Biedermann and Rogler, 2015; Fujimura et al., 2010; Kau et al., 2011; Schwabe and Jobin, 2013). The molecular mechanisms by which the microbiota exerts influence on human health are largely unresolved and undoubtedly complex, yet may hold the key to personalized medicine through therapeutics that target the gut microbial ecosystem (Blanton et al., 2016; Haak et al., 2017; Kootte et al., 2012; Subramanian et al., 2015).

A major factor in shaping the composition and physiology of the gut microbiota is the influx of complex glycans – popularly known as "dietary fibre" – that evade degradation by the limited set of human genome-encoded glycoside hydrolases (Hamaker and Tuncil, 2014; El Kaoutari et al., 2013; Koropatkin et al., 2012). Indeed, regular ingestion of plant polysaccharides is integral to maintaining a healthy balance of microbes in our lower gastrointestinal tract (De Filippo et al., 2010; Sonnenburg and Sonnenburg, 2014). Members of the Bacteroidetes, a dominant phylum in the human gut, possess an arsenal of Polysaccharide Utilization Loci (PUL) to target a wide range of complex glycans (El Kaoutari et al., 2013). Analogous to the archetypal *Bacteroides thetaiotaomicron* starch utilization system (Sus), a hallmark of all Bacteroidetes PULs is the organization of genes clustered around tandem *susC*/*susD* homologs (encoding a TonB dependent transporter, TBDT, and a cell-surface glycan-binding protein, SGBP, respectively) (Martens et al., 2009). Additional co-localized and co-regulated SGBP(s), a cohort of enzymes, and a transcriptional regulator typically make up a machinery that acts in concert to sense, break down, and import complex glycans (Grondin et al., 2017; Hemsworth et al., 2016). Many such PULs, each targeting specific glycan structures, have been identified by genomics and transcriptomics (see, *e.g.*, the seminal study by (Martens et al., 2011), but detailed functional characterization lags severely behind (reviewed in (Grondin et al., 2017; Martens et al., 2014)). Developing a precise understanding of the molecular details of complex glycan utilization by individual members of the microbiota is essential to designing targeted therapies based on prebiotics, probiotics, and synbiotics (Ciorba, 2012; Slavin, 2013), as well as novel therapeutic interventions.

Recently, comprehensive functional analysis has revealed the detailed molecular mechanisms by which individual PULs enable human gut Bacteroidetes to utilize predominant plant polysaccharides, including the matrix glycans, xyloglucan (Hemsworth et al., 2016; Larsbrink et al., 2014; Tauzin et al., 2016), xylan (Rogowski et al., 2015), β-mannan (Bågenholm et al., 2017), and rhamnogalacturonan II (Ndeh et al., 2017). Mixed-linkage β(1,3)/β(1,4)-glucans (MLGs, Fig. 1A) from cereal grains constitute an additional key group of dietary glycans, whose utilization by gut microbes was previously unresolved

at the molecular level. MLGs are abundant in the aleurone layer of common cereals, including oats (3-8 % dry weight) and barley (2-20 % dry weight) (El Khoury et al., 2012). Beyond their obvious potential to contribute to energy intake (Cummings and Macfarlane, 1997; McNeil, 1984), MLGs have been linked to a range of health benefits, *e.g.* promoting healthy cholesterol and blood glucose levels, ameliorating insulin resistance, and mitigating metabolic syndrome (El Khoury et al., 2012). In particular, the cholesterol lowering effect of oat MLG has long been recognized by the United States Food and Drug Administration (FDA) as well as the United Kingdom Joint Health Claims Initiative (JHCI), and been confirmed by subsequent studies (Othman et al., 2011).

The mechanisms behind these systemic benefits of MLG are, however, incompletely understood, in part due to a lack of understanding of MLG metabolism by individual members of the human gut microbiota. Thus, we report here the molecular characterization of a mixed-linkage glucan utilization locus (MLGUL) in the common symbiont *B. ovatus*. Identifying syntenic MLGUL in other Bacteroidetes revealed that as the archetype, this MLGUL serves as a molecular marker for MLG utilization across the Bacteroidetes phylum, thereby enabling future functional prediction across species.

Results

Identification of a multi-gene locus responsible for MLG utilization by *B. ovatus*

A putative MLGUL was previously identified in *B. ovatus* (Fig. 1B) based on the presence of a tandem *susC/susD* homolog signature (Martens et al., 2009) and high-level expression of select genes in the presence of bMLG (Martens et al., 2011). Individual genes in the locus, BACOVA_02741-02745, were all substantially upgregulated (125 to 298-fold) during growth on bMLG vs. glucose as sole carbon sources (Table S1). BACOVA_02742 and BACOVA_02743 encode the signature TBDT/SGBP pair with 28% and 19% protein sequence identity to SusC and SusD, respectively. The putative MLGUL was additionally predicted to encode a second, non-homologous SGBP (BACOVA_02744), a hybrid two-

component sensor/transcriptional regulator (HTCS, BACOVA_02740), and up to three glycoside hydrolases (GHs).

BoGH16$_{MLG}$ (BACOVA_02741) is a member of Glycoside Hydrolase Family 16 (GH16), a family of *endo*-β-glucanases in the Carbohydrate Active Enzymes (CAZy) classification (Cantarel et al., 2009). GH16 notably includes canonical bacterial MLG *endo*-glucanases (*endo*-MLGase) (Planas, 2000), along with a diversity of *endo*-glucanases and *endo*-galactanases (Eklof and Hehemann). BoGH3$_{MLG}$ (BACOVA_02745) is classified into Glycoside Hydrolase Family 3 (GH3), whose members include *exo*-β-glucosidases (Fincher et al.). Notably, we have determined that BACOVA_02738, which is predicted to encode a second GH3 *exo*-β-glucosidase, is unlikely to be part of the MLGUL for three reasons: (1) BACOVA_02738 was not significantly upregulated on MLG (1.6-fold vs. glucose control, Table S1), (2) a corresponding gene is not found among syntenic loci (Fig. 1B), and (3) the encoded enzyme was catalytically feeble compared to BoGH3$_{MLG}$ on β-glucosides relevant to MLG saccharification (*vide infra*).

To determine the correlation between the presence of the predicted MLGUL and growth of *B. ovatus* on MLG, we constructed an isogenic mutant of *B. ovatus Δtdk* (Larsbrink et al., 2014) in which a contiguous region of DNA encoding genes BACOVA_02738-02745 was deleted (*B. ovatus ΔMLGUL*). Vis-à-vis the parent strain, the *B. ovatus ΔMLGUL* was able to grow equally well on glucose as the sole carbon source, however the ability to grow on bMLG was completely abolished (Fig. S1). Thus, the putative MLGUL is necessary to confer *B. ovatus* the ability to utilize MLG.

Enzymology and structural biology of BoGH16$_{MLG}$, the vanguard MLGase

Cellular localization

The GH family membership of BoGH16$_{MLG}$ suggested a potential role as the vanguard enzyme catalyzing polysaccharide backbone cleavage at the cell surface as the essential first step in MLG utilization. Indeed, the presence of a predicted Type II signal sequence (determined with LipoP 1.0

(Juncker and Willenbrock, 2003)) suggested that the protein is membrane-anchored via lipidation of the N-terminal cysteine residue (Paetzel et al., 2002). To validate this prediction, *B. ovatus Δtdk* was grown on minimal medium with either glucose or bMLG as a sole carbon source prior to immunolocalization of $BoGH16_{MLG}$. As shown in Fig. 2A, $BoGH16_{MLG}$ was clearly visualized on the outer surface of cells in which the presence of the polysaccharide induced MLGUL expression, but was absent from cells grown on glucose (Fig. S2C, S2F). Further analysis of cellular fractions by Western blotting revealed the presence of $BoGH16_{MLG}$ in the membrane fraction, corroborating its attachment to the outer membrane (Fig. 2C). Interestingly, $BoGH16_{MLG}$ was also detected in the lysate supernatant (soluble periplasmic or cytoplasmic proteins) and in the culture supernatant (secreted protein) (Fig. 2C). While the former may represent anchored protein released into the soluble fraction during cell lysis, detection in the culture supernatant could result from packaging and release in outer membrane vesicles, which has previously been observed for other Bacteroidetes glycoside hydrolases (Elhenawy et al., 2014).

Substrate and product specificity

To verify the leading catalytic role of $BoGH16_{MLG}$ in MLG utilization and determine the specificity of the enzyme for individual β-glucans, recombinant $BoGH16_{MLG}$ produced in *E. coli* ($recBoGH16_{MLG}$, Fig. S3A, S3B) was screened for hydrolytic activity against a library of polysaccharides. No activity was observed on tamarind xyloglucan, beechwood xylan, wheat arabinoxylan, carob galactomannan, konjac glucomannan, synthetic carboxymethylcellulose, synthetic hydroxyethylcellulose, *Xanthomonas campestris* xanthan gum, or *Ulva sp.* ulvan. In this initial screen, $BoGH16_{MLG}$ was minimally active on all-β(1,3)-glucans, including *Laminaria digitata* laminarin, yeast β-glucan, and *Alcaligenes faecalis* curdlan, but exhibited high specific activity on bMLG. The optimum pH of 6.5 (consistent with function in the distal human colon) and maximum temperature range of 45 to 55 °C was determined using bMLG as substrate (data not shown).

Subsequent Michaelis-Menten kinetic analysis at the pH optimum and 37 °C demonstrated that BoGH16$_{MLG}$ is a highly predominant mixed-linkage β(1,3)/β(1,4)-glucanase (MLGase), with a 33-fold higher specificity constant, k_{cat}/K_m, for bMLG (Fig. 1A) over laminarin, an all-β(1,3)-glucan with single β(1,6)-linked glucosyl branches (Fig. 3A, Table S2) (Martin et al., 2007). BoGH16$_{MLG}$ was even less efficient on the other two all-β(1,3)-glucans for which activity was initially observed: The k_{cat}/K_m, was 147-fold higher for bMLG than yeast β-glucan (similar in structure to laminarin but with longer β(1,6)-linked glucose branches (Lowman et al., 2011)) and nearly four orders of magnitude higher than that of high curdlan, a 22 kDa, non-branched β(1,3)-glucan (Harada et al., 1968) (Fig. 3A, Table S2).

Detailed product analysis was employed to determine the mode of hydrolysis, *endo* vs. *exo*, and linkage specificity of recBoGH16$_{MLG}$ to gain information on the nature of the MLG cleavage products at the *B. ovatus* cell surface. HPLC analysis at selected time points in the hydrolysis showed the initial production of very large oligosaccharide fragments, which were progressively converted into shorter species and ultimately to two distinct oligosaccharides in the limit-digest (Fig. 3B). This product evolution indicates that BoGH16$_{MLG}$ operates through an *endo*-dissociative mode of action in which the MLG polysaccharide is stochastically cleaved along the backbone.

Comparison with oligosaccharide standards (Fig. 3B) and additional LC-MS analysis (data not shown) revealed that the limit-digest products were the mixed-linkage trisaccharide, G4G3G [Glcβ(1-4) Glcβ(1-3) Glc], and the mixed-linkage tetrasaccharide, G4G4G3G [Glcβ(1-4) Glcβ(1-4) Glcβ(1-3) Glc]. Thus, BoGH16$_{MLG}$ specifically hydrolyzes β(1,4)-linkages of glycosyl residues that are immediately preceded by a β(1,3)-linked glucosyl residue (toward the non-reducing chain end). This specificity is typical of bacterial *endo*-MLGases within GH16 (Gaiser et al., 2006; Mcgregor et al., 2017; Planas, 2000).

To provide more refined insight into BoGH16$_{MLG}$ substrate specificity, Michaelis-Menten kinetics were determined for a series of chromogenic glycosides (Fig. S4A, S4B; Table S3). recBoGH16$_{MLG}$ had

no activity on the *ortho*-chloro-*para*-nitrophenyl (CNP) β-glycosides of glucose (G-CNP), cellobiose (G4G-CNP), cellotriose (G4G4G-CNP), nor on *para*-nitrophenyl (*p*NP) β-glucoside (G-*p*NP). Weak activity was observed on the *p*NP and CNP β-glycosides of laminaribiose (G3G), consistent with a requirement for a β(1,3) linkage spanning the -2 to -1 active-site subsites (GH subsite nomenclature according to (Davies et al., 1997)), as was indicated by the bMLG limit-digest analysis (*vide supra*). Likewise, G4G3G-CNP and G4G4G3G-CNP were specifically and efficiently hydrolyzed to release the aglycone, with no cleavage of the internal glycosidic bonds. The specificity constants (k_{cat}/K_m values) for CNP release from these mixed-linkage tri- and tetrasaccharides were 800- and 1500-fold greater than that of G3G-CNP, respectively, which indicate that potential -3 and -4 subsites contribute 17 kJ/mol and 1.6 kJ/mol to transition state stabilization ($\Delta\Delta G^{\ddagger}$). Indeed, a very significant contribution from the -3 subsite is a common feature of GH16 *endo*-MLGases (Gaiser et al., 2006; McGregor et al., 2016; Planas, 2000).

BoGH16$_{MLG}$ tertiary structure

Three-dimensional structures of recBoGH16$_{MLG}$ were solved by X-ray crystallography to reveal the molecular basis for MLG recognition and hydrolysis. The apo structure of recBoGH16$_{MLG}$ was determined to a resolution of 1.8 Å by molecular replacement using the structure of *Zobellia galactanivorans* laminarinase ZgLamC$_{GH16\text{-}E142S}$ (PDB code 4CRQ) (Labourel et al., 2015) as a search model (See Table S4 for processing and refinement statistics). The crystal contained two polypeptide chains in the asymmetric unit corresponding to residues I35-L271 of wild-type BoGH16$_{MLG}$ for both chains (residue numbering is from transcriptional start site according to the genomic sequence). No electron density was observed for the N-terminal His$_6$-tag and subsequent 15 amino acids in either chain of the recombinant protein, which suggests that residues C19-D34 of the wild-type enzyme constitute a flexible linker sequence to distance the catalytic module from the outer membrane surface (residues M1-S18 comprise the predicted signal peptide); the sidechain of C19 would constitute the site of N-terminal lipidation (Paetzel et al., 2002). The overall fold of BoGH16$_{MLG}$ consists of a β-jelly roll architecture typical of other GH16 members (Davies and Sinnot, 2008): Two antiparallel β-sheets stack on top of each

other with the concave face forming the polysaccharide substrate binding cleft. The end-on arrangement of the two chains in the asymmetric unit hinted at the possibility of the formation of a dimer (Fig 4A). Size-exclusion chromatography, however, indicated that $BoGH16_{MLG}$ exists as a monomer in solution (data not shown) which, together with steric considerations of polysaccharide binding through the active-site cleft, indicates that end-on contacts observed between Chains A and B are artifacts of crystal packing.

The sidechains of the conserved GH16 catalytic residues (Planas, 2000), comprising Glu-143 (nucleophile), Asp-145 (electrostatic "helper") and Glu-148 (acid/base) are presented on the same face of one β-strand (β8), pointing into the active-site cleft. Notably, these residues are contained in a EXDXXE consensus sequence that is typical of bacterial GH16 laminarinases (β(1,3)-specific endo-glucanases). The insertion of an extra amino acid (underlined), typically methionine, results in a so-called "β-bulge" secondary structural motif that is not found among canonical bacterial MLGases, which instead possess a regular β-strand (Barbeyron et al., 1998; Michel et al., 2001).

Commensurate with this observation, the closest eight structural homologs identified using the Dali server (Holm and Rosenstrom, 2010) feature a β-bulge active-site motif (Table S5). Specifically, the top match (Z-score = 29.3) was the structure of laminarinase "$ZgLamC_{GH16\text{-}E142S}$" from *Zobellia galactanivorans* (PDB code 4CTE) (Labourel et al., 2015), which has 38% amino acid identity and superimposed with $BoGH16_{MLG}$ with a root mean square deviation (RMSD) value of 2.0 Å over 211 out of 231 Cα pairs. In comparison, the closest GH16 homolog with a regular active-site β-strand was the lichenase (MLGase) from *Paenibacillus macerans* (PDB code 1MAC) (Hahn et al., 1995), which has a comparable Z-score of 25.1 and an RMSD value of also 2.0 Å over 200 out of 212 Cα pairs, despite having only 22% amino acid identity with $BoGH16_{MLG}$.

Soaking crystals of the wild-type enzyme with G4G4G3G yielded a product complex with 1.8 Å resolution (Table S4). The complete tetrasaccharide was modelled in electron density spanning subsites -1 to -4 in the active-site cleft of Chain A, while the electron density for the fourth glucosyl residue in

subsite -4 was not resolved in chain B. This is most likely due to disorder of this residue since the corresponding -4 Glc in Chain A is fully solvent exposed, makes no contact with the enzyme, and has significantly higher B-factors (Fig. 4B). These structural observations are consistent with kinetic data for chromogenic MLG oligosaccharides (Table 1), which likewise suggest the existence of three primary negative subsites, -1 to -3, and a weakly interacting -4 subsite.

In both Chain A and B, the three glucosyl residues spanning subsites -1 to -3 are well defined and virtually identical. The reducing-end glucosyl residue in the -1 subsite is in the β-conformation, with the C1 hydroxyl hydrogen bonded to Tyr-181, which is observed in a dual conformation in both chains of the G4G4G3G-complex (Fig. 4C). Interestingly, this dual conformation is not observed in the apo-form of the enzyme; Tyr-181 is "swung in" to the active site in chain B, while it is "swung out" in chain A, the sidechain from chain A stacking on top of the chain B sidechain (Fig. 4D). The conformation of this sidechain will be key to determining the nature of the positive substrate binding subsites, indeed, comparison with other GH16 *endo*-glucanases clearly suggests the presence of two positive subsites (Gaiser et al., 2006; Planas, 2000). Whether the dynamics observed for Tyr-181 are an artefact of crystallisation, or perhaps play a role in substrate binding and product release is unclear in the absence of an enzyme-substrate complex spanning the positive subsites.

With regard to specific interactions in the negative subsites, subsite -1 is further characterized by hydrogen bonds between Glu-143 and the C2 hydroxyl, Trp-125 and the C6 hydroxyl, as well as Glu-148 and the ring oxygen and the C1 hydroxyl. This glucose is also positioned by a stacking interaction with Trp-125 and Trp-129 (Fig. 4E), both of which are conserved across all GH16 laminarinases. At subsite -2, highly conserved Arg-97 forms a hydrogen bond with the C6 hydroxyl, and Asn-60 hydrogen bonds to the C2 hydroxyl as well as to the glucosidic bond oxygen between the -1 and -2 sugars. Another conserved residue, Trp-138, serves as a platform that stacks with the subsite -2 glucose. In subsite -3, the main interaction is stacking against Trp-58, which also forms a hydrogen bond to the glucosidic bond

oxygen between the -3 and -4 sugars (Fig. 4E). Together, these interactions in subsite -3 are responsible for 17 kJ/mol of transition-state stabilization (*vide supra*).

Downstream saccharificaiton of mixed-linkage oligosaccharides produced by BoGH16$_{MLG}$

To elucidate the mechanism for the downstream conversion of the oligosaccharide products of BoGH16$_{MLG}$ to glucose for primary metabolism, we examined the biochemistry of the two predicted exo-β-glucosidases, BoGH3$_{MLG}$ and BACOVA_02738(GH3) associated with the MLGUL.

Cellular localization of BoGH3$_{MLG}$ and the BACOVA_02738(GH3) gene product

BoGH3$_{MLG}$ and BACOVA_02738(GH3) were unambiguously predicted by SignalP 4.0 (Petersen et al., 2011) to contain a secretion signal peptide, while LipoP 1.0 (Juncker and Willenbrock, 2003) additionally indicated a Type II lipoprotein signal sequence (Paetzel et al., 2002) in BoGH3$_{MLG}$ only. The same *B. ovatus Δtdk* cultures used for BoGH16$_{MLG}$ localization, containing glucose or bMLG as the sole carbon source, were probed using polyclonal antibodies independently raised against recombinant BoGH3$_{MLG}$ and the BACOVA_02738(GH3) gene product. Neither protein was detected on the cell surface by fluorescence microscopy, especially in the presence of bMLG which induces BoGH16$_{MLG}$ production (Fig. 2B, Fig. S2A). BoGH3$_{MLG}$ induction by bMLG was confirmed by a Western blot of cellular fractions, which also confirmed that this enzyme is membrane anchored (Fig. 2C).

In contrast, the BACOVA_02738(GH3) gene product was detected to a higher degree in *B. ovatus* cells grown in minimal medium with glucose as a sole carbon source compared to cells induced with bMLG (Fig. S2B). The lack of upregulation by bMLG is consistent with transcriptional analysis which showed a limited change in expression in bMLG vs. glucose (1.6-fold), which was two orders of magnitude lower than definitive MLGUL genes (Table S1). The higher detection in uninduced cells is explained by the high basal expression of BACOVA_02738(GH3) (more than an order of magnitude greater than all MLGUL members; Table S1). The lack of detection in minimal medium containing

bMLG is due to high amounts of induced MLGUL proteins diminishing the presence of the BACOVA_02738(GH3) gene product when normalized to total protein (Fig. S2B).

Substrate product specificity of BoGH3$_{MLG}$ and BACOVA_02738(GH3)

Initial activity screening on chromogenic *p*NP glycosides (see Experimental Procedures) revealed that both recBoGH3$_{MLG}$ and recBACOVA_02738(GH3) are specific *exo*-β-glucosidases (activity on other *p*NP glycosides was undetectable at micromolar enzyme concentrations). However, recBACOVA_02738(GH3) is strikingly feeble compared to recBoGH3$_{MLG}$ on G-β-*p*NP (k_{cat}/K_m values of 0.084 mM^{-1} s^{-1} versus 20 mM^{-1} s^{-1}; Fig. S4C, S4D, Table 1). Further, measuring Michealis-Menten kinetic parameters on cello- and laminari-oligosaccharides was not feasible due to overall poor activity and low production yields (data not shown). These kinetic results corroborate the above comparative genetic and transcriptional analyses, collectively suggesting BACOVA_02738(GH3) is not integral to the MLGUL. Hence, this enzyme was not characterized further.

To investigate oligosaccharide substrate preference of the BoGH3$_{MLG}$, we conducted initial-rate kinetics analyses on a series of gluco-oligosaccharides of distinct linkage composition and degrees of polymerization (d.p.). The non-reducing-end glucose was hydrolyzed from all-β(1,4)-linked cello-oligosaccharides (d.p. 2-6), all-β(1,3)-linked laminari-oligosaccharides (d.p. 2-5), and mixed-linkage β(1,3)/β(1,4)-gluco-oligosaccharides (d.p. 3-4, 5 examples) with comparable efficiencies, according to classic Michaelis-Menten saturation kinetics (Fig. S4E, S4F; Table 1). In this series, only cellobiose (G4G) was poorly hydrolyzed by BoGH3$_{MLG}$ *vis-à-vis* the β(1,3)-linked congener laminaribiose (G3G) and all other gluco-oligosaccharides (*e.g.*, G4G has a k_{cat}/K_m value 20-fold lower than G3G, Table 1). The β(1,6)-linked diglucoside gentiobiose (G6G) was also a very poor substrate, with a k_{cat}/K_m value 260-fold lower than that of G3G. Gluco-oligosaccharides with a β (1,3)-linked glucosyl unit at the non-reducing end all have slightly higher k_{cat} values compared to those with a β(1,4)-linkage in this position, which typically contributes to higher k_{cat}/K_m values for the former, when substrates of equal d.p. are compared.

281 However, the magnitude of these differences, which are often less than 2-fold, indicate that BoGH3$_{MLG}$ is
282 essentially agnostic to β(1,3) versus β(1,4) linkages. These results also suggest that in addition to a single
283 negative subsite (-1) commensurate with its *exo*-activity, BoGH3$_{MLG}$ has only two positive subsites that
284 contribute to catalysis: in each gluco-oligosaccharide series, tetrasaccharides and larger are hydrolyzed
285 with identical k_{cat}/K_m values to the corresponding trisaccharides.

286 Product analysis following extended incubation of BoGH3$_{MLG}$ with G4G4G3G and G4G3G
287 demonstrated that BoGH3$_{MLG}$ completely degrades the BoGH16$_{MLG}$ limit-digest products to glucose.
288 HPLC also revealed that laminaribiose (G3G) is the only new intermediate formed during the course of
289 hydrolysis (Fig. 3B). This demonstrates that BoGH3$_{MLG}$ sequentially hydrolyzes one glucose unit at a
290 time from the non-reducing end of MLG oligosaccharides, *viz.*: G4G4G3G → G + G4G3G (also present
291 in the starting mixture) → G+ G3G → G + G. Notably, the individual k_{cat} and K_m values for each step are
292 nearly identical (Table 1).

293 **BoGH3$_{MLG}$ and BACOVA_02738(GH3) primary structures**

294 Despite extensive efforts, we were unable to crystallize the key β-glucosidase BoGH3A$_{MLG}$ for
295 experimental tertiary structure determination. However, BoGH3$_{MLG}$ has 63% sequence identity to a *B.*
296 *ovatus* β-glucosidase (BoGH3B, PDB code 5JP0 (Hemsworth et al., 2016)), from the xyloglucan
297 utilization locus (Fig. S5A) and, as such, was amenable to tertiary structure homology modelling. Phyre2
298 (Kelley et al., 2015) utilized PDB code 5JP0 as the sole template and 728 out of 742 residues (98% of the
299 sequence, excluding the signal peptide) were modelled with 100 % confidence. The model suggests that
300 BoGH3$_{MLG}$ possesses a homologous three-domain architecture with the active site being formed at the
301 interface of the $(\alpha/\beta)_8$ TIM barrel and α/β sandwich domains (Fig. S5B). The predicted catalytic
302 nucleophile and acid/base residues are Asp-309 and Glu-453, respectively, based on primary and tertiary
303 alignment (Fig. S5A, S5C). Two tryptophan residues were modelled on opposite sides of the entrance to
304 the active site pocket (Fig. S5D), forming a narrow "coin slot", which may effect a preference towards

β(1,3)- and β(1,4)-linked glucans and poor activity against β(1,6)-linked gentiobiose (Table 1). In constrast, enzymes that lack a homologous Trp-453 have a more open entrance to the active site and show broad activity against β(1,2)- and β(1,6)-linked glucans in addition to β(1,3)- and β(1,4)-linked glucans (Karkehabadi et al., 2014; Pozzo et al., 2010).

In comparison, BACOVA_02738(GH3) possess catalytic residues homologous to $BoGH3_{MLG}$ and similar GH3 members, despite having only 31% sequence identity to $BoGH3_{MLG}$ (Fig. S5A). The most similar characterized GH3 member to BACOVA_02738(GH3) among ca. 300 members identified in the CAZy is a *Chrysosporium lucknowense* β-glucosidase with 39% sequence identity (Dotsenko et al., 2012).

Syntenic MLGUL are molecular markers of MLG utilization across the Bacteroidetes

Refined functional characterization of the catalytic specificity of GH components significantly increases confidence in the use of individual PULs as genetic markers of complex carbohydrate metabolism among Bacteroidetes (Cuskin et al., 2015; Larsbrink et al., 2014; Rogowski et al., 2015; Sonnenburg et al., 2010). The MLGUL characterized here represents the only PUL in *B. ovatus* that confers growth on MLG from cereals. To understand the wider distribution of MLG metabolic capacity among symbiotic Bacteroidetes in the human gut, we correlated the presence of a syntenic MLGUL with growth on bMLG for 354 individual Bacteroidetes strains representing 29 different species.

A total of 121 strains across just 7 of the species were able to grow on bMLG (Fig. 5). In particular, 33 of 33 *B. ovatus* strains (including the type strain ATCC 8483) grew well on bMLG, as well as 44 of 45 strains of the closely related *B. xylanisolvens*. The majority of *B. uniformis* strains tested (33 out of 35) were also competent bMLG utilizers. The limited penetrance of the MLGUL across the genus clearly demonstrates nutrient-niche specialization among individual *Bacteroides* species.

Comparative analysis of available genomic sequences revealed that strains able to grow on bMLG as the sole carbon source harbor a syntenic MLGUL (Fig. 1B). Previous transcriptional analysis

329 demonstrated that the syntenic MLGUL found in *B. cellulosilyticus* is also activated during growth on
330 bMLG (McNulty et al., 2013). Concordance between the presence of a syntenic MLGUL and the ability
331 to utilize MLG is further highlighted by the lack of a MLGUL in the *B. uniformis* ATCC 8492, one of
332 only two strains of *B. uniformis* that could not grow on bMLG. Based on this analysis, we can also predict
333 MLG utilization ability in two sequenced species of *Prevotella*, *Pr. copri* DSM 18205 and *Pr. multiformis*
334 DSM16608, important members of the Bacteroidetes from the human gut and oral cavity, respectively
335 (Fig. 1B).

336 # Discussion

337 ## A molecular model for MLG utilization by *B. ovatus*

338 Our current suite of data suggests a model by which the MLGUL gene products work in concert
339 to enable the utilization of MLG (Fig. 6), analogous to that of other PUL-encoded systems (Grondin et al.,
340 2017). Thus, BoGH16$_{MLG}$ is anchored to the outer membrane where it plays a leading role in fragmenting
341 large MLG polysaccharide chains (typical d.p. 700 – 5000, depending upon the plant species of origin
342 (Lazaridou et al., 2004; Zheng et al., 2011)) into oligosaccharides that can be imported into the periplasm
343 via the TBDT. Notably, the specific limit-digest products of BoGH16$_{MLG}$ *endo*-hydrolysis identified here,
344 *viz.* the trisaccharide G4G3G and the tetrasaccharide G4G4G3G (Fig. 3B), have been shown previously to
345 bind the periplasmic sensor domain of the HTCS encoded by BACOVA_02740 (K_D 300 μM and 400 μM,
346 respectively), while the intact MLG polysaccharide does not (Martens et al., 2011). Monomeric glucose,
347 all-β(1,4)-linked cello-oligosaccharides, and all-β(1,3)-linked laminari-oligosaccharides are also not
348 bound by the HTCS (Martens et al., 2011), indicating that the unique linkages present in MLG are central
349 to inducing the MLGUL system. Thus, there is an essential, yet distant, interplay between the outer-
350 membrane localized MLGase and the HTCS in specific nutrient sensing.

351 It is therefore likely that the BoGH16$_{MLG}$ limit-digest products, or minimal repeats of these
352 structures [(G4G4G3G)$_m$(G4G3G)$_n$], comprise the main products transported through the TBDT *in vivo*.

353 Recent studies on inulin (β(2,1)-fructan) utilization suggest that some TBDTs are able to import longer
354 polysaccharide chains (e.g. d.p. >20) (Rakoff-Nahoum et al., 2016). Regardless of length, the efficient
355 *exo*-hydrolytic activity of BoGH3$_{MLG}$ in the periplasm is sufficient to completely saccharify all imported
356 oligosaccharides to glucose (Fig. 3B), to feed primary metabolism in the cytosol. In this process, the
357 trisaccharide substrate of the HTCS, G4G3G, will always be generated regardless of the imported
358 saccharide chain length, ensuring continual production of the MLGUL up-regulation signal until substrate
359 is exhausted. Interestingly, BoGH3$_{MLG}$ will never encounter cellobiose (G4G), towards which it has
360 relatively weak activity (Fig. S4F; Table 1), in this process; the final step of saccharification of MLGOs is
361 the hydrolysis of the competent substrate laminaribiose (G3G).

362 **Structural enzymology reveals complex trajectories for the evolution of MLG activity in GH16**

363 Previous phylogenetic analyses of GH16 have suggested that the evolution of the active-site β-
364 bulge motif EXDXXE, which is widespread among Clan GH-B (comprising GH16 and GH7), to a regular
365 β-strand motif EXDXE is a defining feature that delineates *endo*-β(1,3)-glucanases (laminarinases, EC
366 3.2.1.39 and EC 3.2.1.6) from mixed-linkage *endo*-β(1,3)/β(1,4)-glucanases (licheninases, EC 3.2.1.73),
367 respectively (Barbeyron et al., 1998; Michel et al., 2001). In this context, the observation that
368 BoGH16$_{MLG}$ is highly specific for MLG, despite having a β-bulge motif, was surprising.

369 Using the CAZy Database as a starting point (http://www.cazy.org/GH16_characterized.html)
370 together with mining of the primary literature, we generated a contemporary maximum-likelihood
371 phylogeny of all biochemically characterized GH16 members (Fig. S6). This analysis indicates that
372 although canonical, normal β-strand MLGases do form a monophyletic group as previously observed,
373 MLGase activity is in fact widespread among the historical "laminarinase" group, in which BoGH16$_{MLG}$
374 is itself positioned. Despite currently limited and disparate kinetic data for individual enzymes, it also
375 appears that it is not possible to define further substrate-specific clades within this group based on
376 molecular phylogeny alone, in light of weak bootstrap support. This precludes defining any single

evolutionary event giving rise to unique trajectories for the further diversification of extant laminarinases and MLGases in this clade. Instead, it appears that diverse, subtle mutations have allowed the independent evolution of predominant laminarinase or MLGase activity numerous times across a flat evolutionary landscape. As such, we suggest that this GH16 subgroup should be more generally referred to as the "laminarinase/MLGase group" going forward.

Detailed tertiary structural comparison of 10 β-bulge-containing members of this laminarinase/MLGase group revealed, however, that predominant laminarinases harbor a significantly more protruding loop (which is often, but not always, longer) connecting strands β2 and β3 than predominant MLGases (Fig. S7A, S7B). Structural superposition with the $BoGH16_{MLG}$:G4G4G3G complex indicates that this loop in predominant laminarinases would clash with MLG in the negative subsites, instead favoring binding of an all-β(1,3)-glucan that curves away from this loop. Such curvature is exemplified by the $ZgLamC_{GH16\text{-}E142S}$:thio-β-1,3-trisaccharide structure (Fig. S7A, PDB code 4CTE) (Labourel et al., 2015). Indeed, Ilari *et al.* observed that shortening the homologous loop in LamA from the archeon *Pyrococcus furiosus* (Fig. S7A, PDB code 2VY0) by 4 amino acids increased the activity towards MLG by 10-fold (Ilari et al., 2009). Likewise, BglF from *Nocardiopsis* sp. F96 (Fig. S6B, PDB code 2HYK) and LamR from *Rhodothermus marinus* (Fig. S7B, PDB code 3ILN), which have a 3.3- and 8.5-fold greater specificity constant and specific activity, respectively, toward MLG than laminarin, also have a smaller loop, similar to $BoGH16_{MLG}$, in this position. The canonical, regular-β-strand MLGase from *Paenibacillus macerans* (Fig. S7C, PDB code 1MAC) and *Bacillus licheniformis* (Fig. S7C, PDB code 1GBG), similarly have a small loop at this position.

Taken together, these analyses reveal a complex evolutionary landscape that computational phylogenetic analysis fails to resolve. Despite using a manually curated, structure-based sequence alignment as input, the maximum-likelihood numerical approach did not delineate the members of the laminarinase/MLGase group on the basis of the distinct active-site loop differences observed in tertiary

structures (Fig. S7). Instead, the phylogeny was likely obfuscated by diverse, random variations in amino acid composition across the entire β-sandwich domain, which clearly limits large-scale, unsupervised phylogenetic analysis of these GH16 members. Moreover, analysis of both MLG and laminarin specificity (as a minimum) for individual members of this group, in light of their tertiary structures, is essential to avoid potential mis-annotation of these enzymes.

Mining metagenomic data reveals the ubiquity of MLG utilization in the human gut and beyond

Using syntenic MLGULs as genetic markers, we surveyed the publicly available metagenome data of 426 adults to understand the capacity of human populations to derive nutrition from cereal MLGs. We were able to distinguish the species of origin based on nucleotide sequence except for MLGULs from *B. ovatus* and *B. xylanisolvens*, which were strikingly similar at 97% nucleotide identity. The *B. ovatus*/*B. xylanisolvens* and *B. uniformis* MLGULs are the most prevalent; both are observed in about 70% of the total human cohort (Fig. 7). The *Pr. copri* MLGUL is more often the sole MLGUL of an individual than that of *B. cellulosilyticus* when only one is present (Fig. 7, cyan lines), despite the latter being more frequent in total. Overall, 92.5% of the subjects harbor at least one of the five different MLGULs identified in this study, irrespective of nationality or whether they have been diagnosed with IBD. MLGULs are ubiquitously detectable despite variability in sampling depth across different metagenomics sequencing projects (Fig. 7). The prevalence of MLGULs across different nationalities is consistent with MLG from cereal grains being a ubiquitous component of the human diet. Indeed, the importance of MLG utilization is underscored by the upregulation of the MLGUL in the ceca of mice fed a complex plant cell wall diet (Martens et al., 2011). Similar widespread global distribution in human populations has been observed for xyloglucan utilization loci (Larsbrink et al., 2014). These observations are sharply contrasted by the PUL that mediates utilization of the red algal polysaccharide porphyran, which is essentially confined to subjects from Japan, where seaweed is a common part of the diet (Hehemann et al., 2010; Larsbrink et al., 2014). Interestingly, we were unable to detect MLGULs in four unweaned

infants sampled in the Japanese metagenome project (data not included in our analysis of adult metagenomes). This is consistent with a dearth of Bacteroidetes in infants who receive the bulk of their nutrition from milk and are not yet consuming plant polysaccharides (Urokawa et al., 2007)

Moving beyond the human microbiota, we can likewise predict MLG utilization ability in *Dysgonomonas gadei* and *Pr. oryzae* (formerly *Xylanibacter oryzae*) based on the presence of a syntenic MLGUL. These species are commonly found in the termite hindgut and decomposing rice straw, respectively. This provides direct evidence that an analogous MLG utilization system is employed by Bacteroidetes operating in environments beyond the human gut.

Conclusion

Complex carbohydrates that promote the growth of beneficial microbes in our distal large intestine are a cornerstone of a healthy diet. MLGs in particular have long been known to impart healthful effects (Othman et al., 2011), yet its mechanism of utilization for fermentation by gut microbes was unknown. Our work here sheds light on the fine-tuned mechanism that *B. ovatus* and other Bacteroidetes has evolved to efficiently utilize MLGs in the highly competitive environment of the human gut microbiota. The finding that a majority of humans possess microbes that can utilize this ubiquitous cereal polysaccharide highlights the relevance of potential therapeutic interventions based on MLG utilization to the general population. The present study also sets the stage for future work to understand the quantitative contributions of individual members of the microbiota and their cognate enzymes to MLG utilization in the human gut (Patrascu et al., 2017; Zhong et al., 2015).

Experimental Procedures

Microbiology. *B. ovatus* gene deletions were constructed using allelic exchange as previously described (Koropatkin et al., 2008). Anaerobic growth profiles were measured as previously described (Martens et

al., 2011). Details of localization analysis by immunofluorescence microscopy and immunoblotting are provided in the Supplemental Experimental Procedures.

Cloning, expression, and purification of recombinant enzymes. The genes encoding $BoGH16_{MLG}$, $BoGH3_{MLG}$, and BACOVA_02738(GH3) were cloned into expression vectors for recombinant production in *E. coli*. Details of cloning strategies, production, and purification are provided in the Supplemental Experimental Procedures.

Enzyme kinetics and product analysis. Thorough kinetic analysis was conducted on a panel of polysaccharides, oligosaccharides, and chromogenic substrates. Products of enzymatic reactions were analyzed by HPAEC-PAD and HILIC-MS. Details of enzymatic assays, analytical methods, as well as sources of commercial substrates are provided in the Supplemental Experimental Procedures.

X-ray crystallography. Crystals of $BoGH16_{MLG}$ were screened and optimized by sitting drop vapor diffusion method. The structures of the apo- and G4G4G3G-$BoGH16_{MLG}$ were solved by molecular replacement. Details of crystallization, data collection, and structure solution are provided in the Supplemental Experimental Procedures.

Bioinformatics. Phylogenetic anlaysis of select GH16 and GH3 sequences was conducted based on structure-guided alignment. Metagenomic survey was carried out by nucleotide BLAST of MLGUL sequences against various metagenome sequence data. Details are provided in the Supplemental Experimental Procedures.

Statistical analysis. All kinetic assays were done in triplicate. Michaelis-Menten parameters are reported as fitted values ± standard deviation throughout the article. All growth experiment results are averages of two biological replicates.

Author contributions

KT cloned, expressed, and purified recombinant enzymes, conducted and analyzed kinetics for hydrolysis of polysaccharides, oligosaccharides, and chromogenic substrates, determined hydrolysis products, conducted phylogenetic and structural analyses, carried out metagenomics survey, and wrote the article. GRH determined the crystal structures of apo- and G4G4G3G-complexed $BoGH16_{MLG}$. GD conducted enzyme localization analyses. TR, NP, and KU conducted reverse genetics and growth analyses. NJ synthesized G3G-CNP. GJD, ECM, and HB designed and directed research and co-wrote the article with input from all authors.

Accession numbers

The atomic coordinates and structure factors of apo- and G4G4G3G-complexed $BoGH16_{MLG}$ reported in this paper can be accessed through the PDB with identifiers 5NBO and 5NBP, respectively.

Acknowledgements

We thank Diamond Light Source (Harwell, UK) for access to beamlines I02 and I03 (proposal mx9948) that contributed to the results presented here. Work in Vancouver was supported by Operating Grants from the Canadian Institutes for Health Research (MOP-137134 and MOP-142472) and infrastructure support from the Canadian Foundation for Innovation and the British Columbia Knowledge Development Fund. We thank Constance M. Bahr (Koropatkin group, U. Michigan) for invaluable assistance with microscopy. We thank Adriana Cabrera (Brumer group) for preparing laminaritol by sodium borohydride reduction of laminarin. We thank Nicholas McGregor (Brumer group) for assistance with LC-MS. We thank Alexander H. Viborg (http://research.ahv.dk/) for access to CAZy database tools. We thank Hila Behar (Brumer group) for assistance with BLAST analysis of human metagenome sequences.

References

Bågenholm, V., Reddy, S.K., Bouraoui, H., Morrill, J., Kulcinskaja, E., Bahr, C.M., Aurelius, O., Rogers, T., Xiao, Y., Logan, D.T., et al. (2017). Galactomannan Catabolism Conferred by a Polysaccharide Utilization Locus of Bacteroides ovatus. *292*, 229–243.

Barbeyron, T., Gerard, A., Potin, P., Henrissat, B., and Kloareg, B. (1998). The Kappa-Carrageenase of the Marine Bacterium Cytophaga drobachiensis. Structural and Phylogenetic Relationships Within Family-16 Glycoside Hydrolases. Mol. Biol. Evol *15*, 528–537.

Biedermann, L., and Rogler, G. (2015). The intestinal microbiota: its role in health and disease. Eur. J. Pediatr. *174*, 151–167.

Blanton, L. V., Barratt, M.J., Charbonneau, M.R., Ahmed, T., and Gordon, J.I. (2016). Childhood undernutrition, the gut microbiota, and microbiota-directed therapeutics. Science. *352*, 1533–1533.

Cantarel, B.I., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. *37*, 233–238.

Ciorba, M. (2012). A Gastroenterologist's Guide to Probiotics. Clin. Gastroenterol. Hepatol. *10*, 960–968.

Cummings, J.H., and Macfarlane, G.T. (1997). Role of intestinal bacteria in nutrient metabolism. JPEN. J. Parenter. Enteral Nutr. *21*, 357–365.

Cuskin, F., Lowe, E.C., Temple, M.J., Zhu, Y., Cameron, E.A., Pudlo, N.A., Porter, N.T., Urs, K., Thompson, A.J., Cartmell, A., et al. (2015). Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. Nature *517*, 165–169.

Davies, G.J., and Sinnot, M.L. (2008). Sorting the diverse. Biochem. J. *30*, 26–32.

Davies, G.J., Wilson, K.S., and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. Biochem. J. *321*, 557–559.

Dotsenko, G.S., Sinitsyna, O.A., Hinz, S.W.A., Wery, J., and Sinitsyn, A.P. (2012). Bioresource Technology Characterization of a GH family 3 β-glycoside hydrolase from Chrysosporium lucknowense and its application to the hydrolysis of β-glucan and xylan. Bioresour. Technol. *112*, 345–349.

Eklof, J., and Hehemann, J.-H. "Glycoside Hydrolase Family 16" in CAZypedia, available at URL http://www.cazypedia.org/, accessed 10 February 2017.

Elhenawy, W., Debelyy, M.O., and Feldman, M.F. (2014). Preferential packing of acidic glycosidases and proteases into Bacteroides outer membrane vesicles. mBio *5*, 1–12.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc. Natl. Acad. Sci. U. S. A. *107*, 14691–14696.

Fincher, G., Mark, B., and Brumer, H. "Glycoside Hydrolase Family 3" in CAZypedia, available at URL http://www.cazypedia.org/, accessed 10 February 2017.

Fujimura, K., Slusher, N., Cabana, M., and Lynch, S. (2010). Role of the gut microbiota in defining human health. Expert Rev Anti Infect Ther *8*, 435–454.

Gaiser, O.J., Piotukh, K., Ponnuswamy, M.N., Planas, A., Borriss, R., and Heinemann, U. (2006). Structural basis for the substrate specificity of a Bacillus 1,3-1,4-β-glucanase. J. Mol. Biol. *357*, 1211–

1225.

Grondin, J.M., Tamura, K., Déjean, G., and Abbott, D.W. (2017). Polysaccharide Utilization Loci: Fueling Microbial Communities. J. Bacteriol. *199*, 1–15.

Haak, B.W., Levi, M., and Wiersinga, W.J. (2017). Microbiota-targeted therapies on the intensive care unit. Curr. Opin. Crit. Care *23*, 1–8.

Hahn, M., Pons, J., Planas, a, Querol, E., and Heinemann, U. (1995). Crystal structure of Bacillus licheniformis 1,3-1,4-beta-D-glucan 4-glucanohydrolase at 1.8 A resolution. FEBS Lett. *374*, 221–224.

Hamaker, B.R., and Tuncil, Y.E. (2014). A perspective on the complexity of dietary fiber structures and their potential effect on the gut microbiota. J. Mol. Biol. *426*, 3838–3850.

Harada, T., Misaki, A., and Saito, H. (1968). Curdlan: A bacterial gel-forming β-1, 3-glucan. Arch. Biochem. Biophys. *124*, 292–298.

Hehemann, J.-H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature *464*, 908–912.

Hemsworth, G.R., Thompson, A.J., Stepper, J., Sobala, Ł.F., Coyle, T., Larsbrink, J., Spadiut, O., Goddard-Borger, E.D., Stubbs, K.A., Brumer, H., et al. (2016). Structural dissection of a complex Bacteroides ovatus gene locus conferring xyloglucan metabolism in the human gut. Open Biol. *6*, 1–14.

Holm, L., and Rosenstrom, P. (2010). Dali server: Conservation mapping in 3D. Nucleic Acids Res. *38*, 545–549.

Ilari, A., Fiorillo, A., Angelaccio, S., Florio, R., Chiaraluce, R., Van Der Oost, J., and Consalvi, V. (2009). Crystal structure of a family 16 endoglucanase from the hyperthermophile Pyrococcus furiosus - Structural basis of substrate recognition. FEBS J. *276*, 1048–1058.

Juncker, A., and Willenbrock, H. (2003). Prediction of lipoprotein signal peptides in Gram negative bacteria. Protein Sci. *12*, 1652–1662.

El Kaoutari, A., Armougom, F., Gordon, J.I., Raoult, D., and Henrissat, B. (2013). The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat. Rev. Microbiol. *11*, 497–504.

Karkehabadi, S., Helmich, K.E., Kaper, T., Hansson, H., Mikkelsen, N., Gudmundsson, M., Piens, K., Fujdala, M., Banerjee, G., Scott-craig, J.S., et al. (2014). Biochemical Characterization and Crystal Structures of a Fungal Family 3 β-Glucosidase , Cel3A from Hypocrea jecorina. *289*, 31624–31637.

Kau, A.L., Ahern, P.P., Griffin, N.W., Goodman, A.L., and Gordon, J.I. (2011). Human nutrition, the gut microbiome and the immune system. Nature *474*, 327–336.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling , prediction and analysis. Nat. Protoc. *10*, 845–858.

El Khoury, D., Cuda, C., Luhovyy, B.L., and Anderson, G.H. (2012). Beta glucan: Health benefits in obesity and metabolic syndrome. J. Nutr. Metab. *2012*, 1–28.

Kootte, R.S., Vrieze, A., Holleman, F., Dallinga-Thie, G.M., Zoetendal, E.G., de Vos, W.M., Groen, A.K., Hoekstra, J.B.L., Stroes, E.S., and Nieuwdorp, M. (2012). The therapeutic potential of manipulating gut microbiota in obesity and type 2 diabetes mellitus. Diabetes, Obes. Metab. *14*, 112–120.

Koropatkin, N.M., Martens, E.C., Gordon, J.I., and Smith, T.J. (2008). Starch Catabolism by a Prominent Human Gut Symbiont Is Directed by the Recognition of Amylose Helices. Structure *16*, 1105–1115.

Koropatkin, N.M., Cameron, E. a., and Martens, E.C. (2012). How glycan metabolism shapes the human gut microbiota. Nat. Rev. Microbiol. *10*, 323–335.

Labourel, A., Jam, M., Legentil, L., Sylla, B., Hehemann, J.H., Ferrières, V., Czjzek, M., and Michel, G. (2015). Structural and biochemical characterization of the laminarinase ZgLamCGH16 from Zobellia galactanivorans suggests preferred recognition of branched laminarin. Acta Crystallogr. Sect. D Biol. Crystallogr. *71*, 173–184.

Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tauzin, A.S., Spadiut, O., Klinter, S., Pudlo, N. a, Urs, K., Koropatkin, N.M., et al. (2014). A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. Nature *506*, 498–502.

Lazaridou, A., Biliaderis, C.G., Micha-Screttas, M., and Steele, B.R. (2004). A comparative study on structure-function relations of mixed-linkage (1→3), (1→4) linear β-D-glucans. Food Hydrocoll. *18*, 837–855.

Littman, D.R., and Pamer, E.G. (2011). Role of the commensal microbiota in normal and pathogenic host immune responses. Cell Host Microbe *10*, 311–323.

Lowman, D.W., West, L.J., Bearden, D.W., Wempe, M.F., Power, T.D., Ensley, H.E., Haynes, K., Williams, D.L., and Kruppa, M.D. (2011). New insights into the structure of (1-3,1-6)-β-D-glucan side chains in the Candida glabrata cell wall. PLoS One *6*, 1–10.

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

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N., et al. (2011). Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol. *9*, 1–16.

Martens, E.C., Kelly, A.G., Tauzin, A.S., and Brumer, H. (2014). The devil lies in the details: How variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. J. Mol. Biol. *426*, 3851–3865.

Martin, K., McDougall, B.M., McIlroy, S., Jayus, Chen, J., and Seviour, R.J. (2007). Biochemistry and molecular biology of exocellular fungal β-(1,3)- and β-(1,6)-glucanases. FEMS Microbiol. Rev. *31*, 168–192.

Mcgregor, N., Yin, V., Tung, C., Petegem, F. Van, and Brumer, H. (2017). Crystallographic insight into the evolutionary origins of xyloglucan endotransglycosylases and endohydrolases. Plant J. *89*, 651–670.

McGregor, N., Morar, M., Fenger, T.H., Stogios, P., Lenfant, N., Yin, V., Xu, X., Evdokimova, E., Cui, H., Henrissat, B., et al. (2016). Structure-function analysis of a mixed-linkage β-glucanase/xyloglucanase from the key ruminal bacteroidetes prevotella bryantii B14. J. Biol. Chem. *291*, 1175–1197.

McNeil, N.I. (1984). The contribution of the large-intestine to energy supplies in man . Am. J. Clin. Nutr. *39*, 338–342.

McNulty, N.P., Wu, M., Erickson, A.R., Pan, C., Erickson, B.K., Martens, E.C., Pudlo, N.A., Muegge, B.D., Henrissat, B., Hettich, R.L., et al. (2013). Effects of Diet on Resource Utilization by a Model Human Gut Microbiota Containing Bacteroides cellulosilyticus WH2, a Symbiont with an Extensive Glycobiome. PLoS Biol. *11*, 1–20.

Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B., and Dideberg, O. (2001). The κ-carrageenase of P. carrageenovora features a tunnel-shaped active site: A novel insight in the

609 evolution of clan-B glycoside hydrolases. Structure *9*, 513–525.

610 Ndeh, D., Rogowski, A., Cartmell, A., Luis, A.S., Baslé, A., Gray, J., Venditto, I., Briggs, J., Zhang, X.,
611 Labourel, A., et al. (2017). Complex pectin metabolism by gut bacteria reveals novel catalytic functions.
612 Nature *544*, 65–70.

613 Othman, R.A., Moghadasian, M.H., and Jones, P.J.H. (2011). Cholesterol-lowering effects of oat β-
614 glucan. Nutr. Rev. *69*, 299–309.

615 Paetzel, M., Karla, A., Strynadka, N.C.J., and Dalbey, R.E. (2002). Signal peptidases. Chem. Rev. *102*,
616 4549–4579.

617 Patrascu, O., Béguet-Crespel, F., Marinelli, L., Chatelier, E. Le, Abraham, A., Leclerc, M., Klopp, C.,
618 Terrapon, N., Henrissat, B., Blottière, H.M., et al. (2017). A fibrolytic potential in the human ileum
619 mucosal microbiota revealed by functional metagenomic. Sci. Rep. *7*, 1–15.

620 Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal
621 peptides from transmembrane regions. Nat. Methods *8*, 785–786.

622 Planas, A. (2000). Bacterial 1,3-1,4-β-glucanases: structure, function and protein engineering. Biochim.
623 Biophys. Acta *1543*, 361–382.

624 Pozzo, T., Pasten, J.L., Karlsson, E.N., and Logan, D.T. (2010). Structural and Functional Analyses of β-
625 Glucosidase 3B from Thermotoga neapolitana: A Thermostable Three-Domain Representative of
626 Glycoside Hydrolase 3. J. Mol. Biol. *397*, 724–739.

627 Rakoff-Nahoum, S., Foster, K.R., and Comstock, L.E. (2016). The evolution of cooperation within the
628 gut microbiota. Nature *533*, 255–259.

629 Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Baslé, A., Morland,
630 C., Day, A.M., Zheng, H., et al. (2015). Glycan complexity dictates microbial resource allocation in the
631 large intestine. Nat. Commun. *6*, 1–15.

632 Schwabe, R.F., and Jobin, C. (2013). The microbiome and cancer. Nat. Rev. Cancer *13*, 800–812.

633 Slavin, J. (2013). Fiber and prebiotics: Mechanisms and health benefits. Nutrients *5*, 1417–1435.

634 Sonnenburg, E.D., and Sonnenburg, J.L. (2014). Starving our microbial self: The deleterious
635 consequences of a diet deficient in microbiota-accessible carbohydrates. Cell Metab. *20*, 779–786.

636 Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., Bolam, D.N., and
637 Sonnenburg, J.L. (2010). Specificity of polysaccharide use in intestinal bacteroides species determines
638 diet-induced microbiota alterations. Cell *141*, 1241–1252.

639 Subramanian, S., Blanton, L. V., Frese, S.A., Charbonneau, M., Mills, D.A., and Gordon, J.I. (2015).
640 Cultivating healthy growth and nutrition through the gut microbiota. Cell *161*, 36–48.

641 Tauzin, A.S., Kwiatkowski, K.J., Orlovsky, N.I., Smith, C.J., Creagh, A.L., Haynes, C.A., Wawrzak, Z.,
642 Brumer, H., and Koropatkin, N.M. (2016). Molecular Dissection of Xyloglucan Recognition in a
643 Prominent Human Gut Symbiont. mBio *7*, e02134-15.

644 Urokawa, K.K., Toh, T.I., Uwahara, T.K., Shima, K.O., Oh, H.T., Oyoda, A.T., Ori, H.M., Gura, Y.O.,
645 Hrlich, D.S.E., Toh, K.I., et al. (2007). Comparative Metagenomics Revealed Commonly Enriched Gene
646 Sets in Human Gut Microbiomes. DNA Res. *14*, 169–181.

647 Zheng, X., Li, L., and Wang, Q. (2011). Distribution and molecular characterization of β-glucans from

hull-less barley bran, shorts and flour. Int. J. Mol. Sci. *12*, 1563–1574.

Zhong, Y., Marungruang, N., Fak, F., and Nyman, M. (2015). Effects of two whole-grain barley varieties on caecal SCFA, gut microbiota and plasma inflammatory markers in rats consuming low- and high-fat diets. Br. J. Nutr. *113*, 1558–1570.

Figure Legends

Figure 1. Cereal mixed-linkage β(1,3)/β(1,4)-glucan (MLG) and MLG Utilization Locus (MLGUL) structures. A: Chemical structure of MLG, consisting of a linear glucan chain of β(1,4)-linked cellotriosyl and cellotetraosyl units linked by β(1,3) bonds. MLGs from various sources (barley, oat, lichenin, etc) vary in the ratio of cellotriose to cellotetraose units (Lazaridou et al., 2004). Arrows indicate the specific site of hydrolysis by the vanguard endo-glucanase of the MLGUL, BoGH16$_{MLG}$. B: Genetic organization of the *B. ovatus* MLGUL and syntenic loci in select Bacteroidetes species. Homologous genes are connected by colored bars and the locus tag of the TBDT of each syntenic MLGUL is given on the right. See also Figure S1; Table S1.

Figure 2. Enzyme localization analysis. Phase contrast microscopy and corresponding fluorescence microscopy images of *B. ovatus Δtdk* cells grown in minimal medium with bMLG as the sole carbon source probed with custom polyclonal antibodies against recBoGH16$_{MLG}$ (A) and recBoGH3$_{MLG}$ (B). C: Western blots of protein collected from the culture supernatant, cell lysate supernatant, and cell lysate membrane fraction of *B. ovatus Δtdk* cells grown in minimal medium with glucose or bMLG as a sole carbon source. See also Figure S2.

Figure 3. BoGH16$_{MLG}$ kinetics and MLGUL GHs product analysis. A: BoGH16$_{MLG}$ initial-rate kinetics curves fitted to the Michaelis-Menten equation for β-glucan polysaccharide substrates on which it is active. Laminarin was reduced to laminaritol by sodium borohydride reduction to reduce background in the BCA assay. Curve fitting was done on OriginPro 2015 and error bars represent standard deviations from the mean. B: Chromatograms of bMLG and its hydrolysis products by BoGH16$_{MLG}$ and BoGH3$_{MLG}$ separated by HPAEC-PAD. Red: full length bMLG polysaccharide. Dark blue: reaction progress time course and limit digest of bMLG hydrolysis by 10 nM BoGH16$_{MLG}$. Cyan: reaction progress time course and limit digest of BoGH16$_{MLG}$ products hydrolysis by 25 nM BoGH3$_{MLG}$. Standards are shown below in

black: solid lines are those corresponding to limit digest products and dotted line to intermediate products. See also Figures S3, S4 and S5; Tables S2 and S3.

Figure 4. BoGH16$_{MLG}$ structural biology. A: the overall structure of the BoGH16$_{MLG}$:G4G4G3G asymmetric unit containing two polypeptide chains shown from orthogonal views with the bound oligosaccharides in yellow and the transparent surface representation in white. Chain A cartoon is shown in cyan, and chain B cartoon is shown in slate blue throughout the figure. B: Mixed-linkage tetrasaccharide ligand modelled into chain A of BoGH16$_{MLG}$ with the opaque surface representation in gray and the oligosaccharide colored according to B-factors. The glucose in subsite -4 is outside of the active site cleft and has significantly higher B factor than the glucose units in subsites -1 to -3. C: Tyr-181 rotamers observed in the complex structure with the 2Fo-Fc map of the tyrosines shown contoured at 0.5σ in grey. D: Tyr-181 residues observed in the apo structure with the 2Fo-Fc map of the tyrosines shown contoured at 0.5σ in grey. E: Wall-eyed stero view of the active site of chain A of the BoGH16$_{MLG}$:G4G4G3G complex. Hydrogen bonding interactions are shown as dashed black lines, sugars are shown in yellow with its 2Fo-Fc map contoured at 1σ in orange, and the conserved GH16 active site residues shown in purple. Hydrophobic stacking interactions in addition to hydrogen bonds position the mixed-linkage oligosaccharide in the negative subsite of BoGH16$_{MLG}$. See also Figures S5 and S6; Tables S4 and S5.

Figure 5. Penetrance map of MLG utilization ability across diverse human gut Bacteroidetes. The phylogenetic tree was constructed from fully sequenced strains of the species shown. The number of strains of each species tested for growth is depicted to scale as a black circle at each leaf. The number of those strains that grew on bMLG as a sole carbon source is shown to scale in red within the black circle.

Figure 6. Model of mixed-linkage β-glucan saccharification by the concerted action of the MLGUL machinery. Gene products are colored analogously to the gene locus in Fig. 1. The cell surface localized endo-MLGase BoGH16$_{MLG}$ cleave large mixed-linkage β-glucan polysaccharides into shorter fragments

which are imported into the periplasm via the TonB dependent transporter, BoTBDT. This glycan capture and transport process at the cell surface is aided by the two surface glycan binding proteins BoSGBP-A and BoSGBP-B. The smaller mixed-linkage β-glucan fragments in the periplasm bind the sensor domain of the hybrid two-component sensor BoHTCS to induce upregulation of the system. Periplasmic exo-β-glucosidases BoGH3$_{MLG}$ and BACOVA_02738(GH3) act from the non-reducing ends to liberate individual glucose monomers which are imported into the cell and metabolized.

Figure 7. Bacteroidetes MLGULs from a survey of 426 adult human gut metagenomes. Vertical lines represent the presence (cyan when unique, blue when one of multiple) or absence (black) of a corresponding species-related MLGUL in a single individual. The total number of MLGULs observed in an individual is shown in the bottom row, colored according to the legend in the top left corner. The frequency of MLGUL occurrence across all 426 individuals is shown on the right. Variation in sequencing depth in megabase pair is illustrated in the graph below: grey lines show the depth for individual subjects and black lines show the average depth of each metagenomics project.

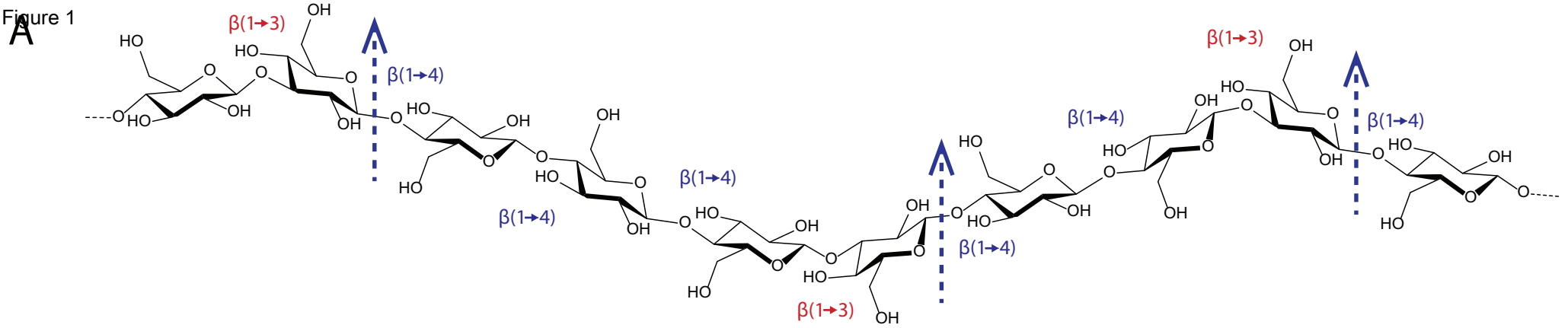
Table

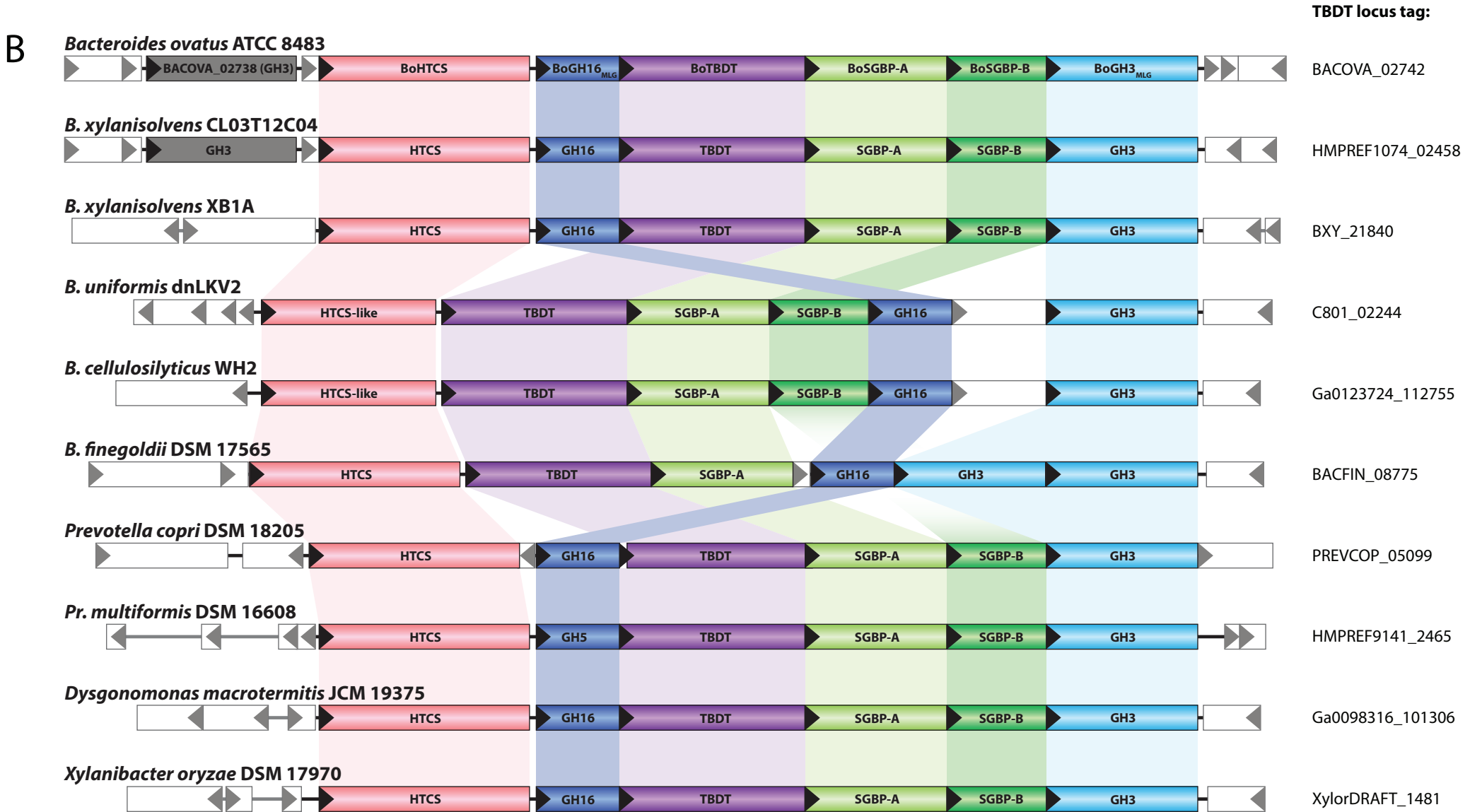
Table 1.

Enzyme	Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})	Assay
$BoGH3_{MLG}$	β-Glc-*p*NP	59.5 ± 1.46	2.95 ± 0.14	20.2	*p*NP
	gentiobiose (G6G)	ND	ND	0.0571	HK/G6PDH
	cellobiose	5.52 ± 0.19	7.47 ± 0.48	0.739	HK/G6PDH
	cellotriose	22.1 ± 0.3	0.859 ± 0.033	25.7	HK/G6PDH
	cellotetraose	17.3 ± 0.5	0.687 ± 0.044	25.2	HK/G6PDH
	cellopentaose	19.4 ± 0.8	0.777 ± 0.060	25.0	HK/G6PDH
	cellohexaose	17.4 ± 0.4	0.747 ± 0.041	23.3	HK/G6PDH
	laminaribiose	28.0 ± 1.1	1.90 ± 0.12	14.7	HK/G6PDH
	laminaritriose	34.2 ± 1.0	0.911 ± 0.052	37.5	HK/G6PDH
	laminaritetraose	31.3 ± 2.3	0.898 ± 0.135	34.9	HK/G6PDH
	laminaripentaose	39.5 ± 3.4	1.27 ± 0.20	31.1	HK/G6PDH
	$MLGO_3$ A (G3G4G)	61.6 ± 1.6	0.997 ± 0.040	61.8	HK/G6PDH
	$MLGO_3$ B (G4G3G)	24.7 ± 1.3	0.521 ± 0.064	47.4	HK/G6PDH
	$MLGO_4$ A (G3G4G4G)	55.7 ± 2.7	1.33 ± 0.12	41.9	HK/G6PDH
	$MLGO_4$ B (G4G4G3G)	30.8 ± 2.0	0.736 ± 0.106	41.8	HK/G6PDH
	$MLGO_4$ C (G4G3G4G)	15.7 ± 0.3	0.601 ± 0.031	26.1	HK/G6PDH
BACOVA_02738 (GH3)	β-Glc-*p*NP	0.212 ± 0.004	2.53 ± 0.13	0.0838	*p*NP

ND: not determined (in cases where Michealis-Menten curve fitting was not feasible, individual k_{cat} and K_m values are not reported and k_{cat}/K_m value was determined from linear curve fit to initial-rate data in the [S] << $K_{m(apparent)}$ range). Data are represented as the fitted parameters ± standard deviation. Highlighted in bold are the biologically relevant substrates that $BoGH3_{MLG}$ encounters in the periplasmic space. See also Figure S4.

Figure 1


A
β(1→3)
β(1→4)
β(1→4)
β(1→4)
β(1→3)
β(1→4)
β(1→4)
β(1→3)


B
TBDT locus tag:
Bacteroides ovatus ATCC 8483
BACOVA_02738 (GH3)
BoHTCS
BoGH16MLG
BoTBDT
BoSGBP-A
BoSGBP-B
BoGH3MLG
BACOVA_02742
B. xylanisolvens CL03T12C04
GH3
HTCS
GH16
TBDT
SGBP-A
SGBP-B
GH3
HMPREF1074_02458
B. xylanisolvens XB1A
HTCS
GH16
TBDT
SGBP-A
SGBP-B
GH3
BXY_21840
B. uniformis dnLKV2
HTCS-like
TBDT
SGBP-A
SGBP-B
GH16
GH3
C801_02244
B. cellulosilyticus WH2
HTCS-like
TBDT
SGBP-A
SGBP-B
GH16
GH3
Ga0123724_112755
B. finegoldii DSM 17565
HTCS
TBDT
SGBP-A
GH16
GH3
GH3
BACFIN_08775
Prevotella copri DSM 18205
HTCS
GH16
TBDT
SGBP-A
SGBP-B
GH3
PREVCOP_05099
Pr. multiformis DSM 16608
HTCS
GH5
TBDT
SGBP-A
SGBP-B
GH3
HMPREF9141_2465
Dysgonomonas macrotermitis JCM 19375
HTCS
GH16
TBDT
SGBP-A
SGBP-B
GH3
Ga0098316_101306
Xylanibacter oryzae DSM 17970
HTCS
GH16
TBDT
SGBP-A
SGBP-B
GH3
XylorDRAFT_1481

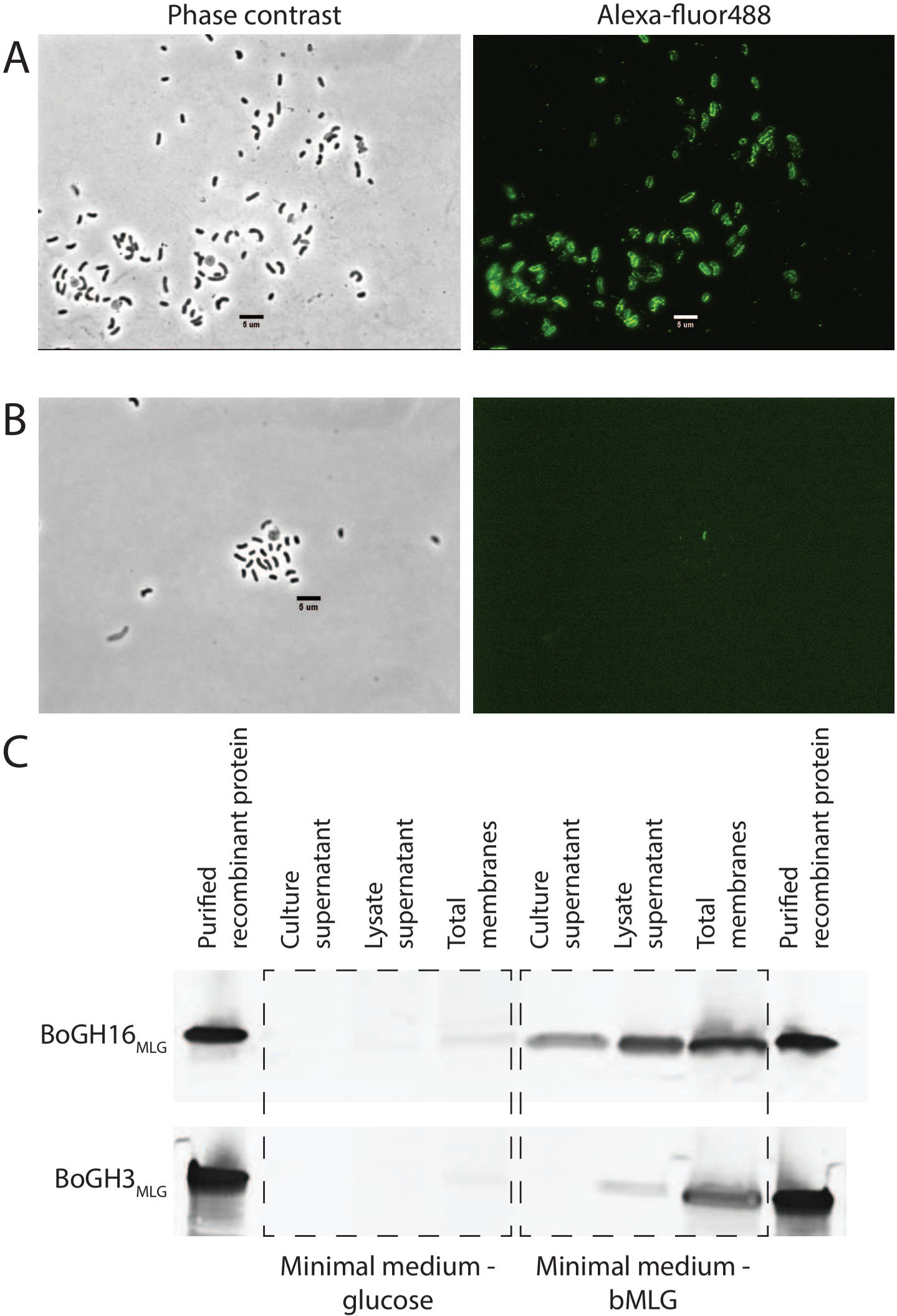

Phase contrast
Alexa-fluor488
A
5 um
5 um
B
5 um
C
Purified recombinant protein
Culture supernatant
Lysate supernatant
Total membranes
Culture supernatant
Lysate supernatant
Total membranes
Purified recombinant protein
BoGH16MLG
BoGH3MLG
Minimal medium - glucose
Minimal medium - bMLG

Figure 2

Figure 3

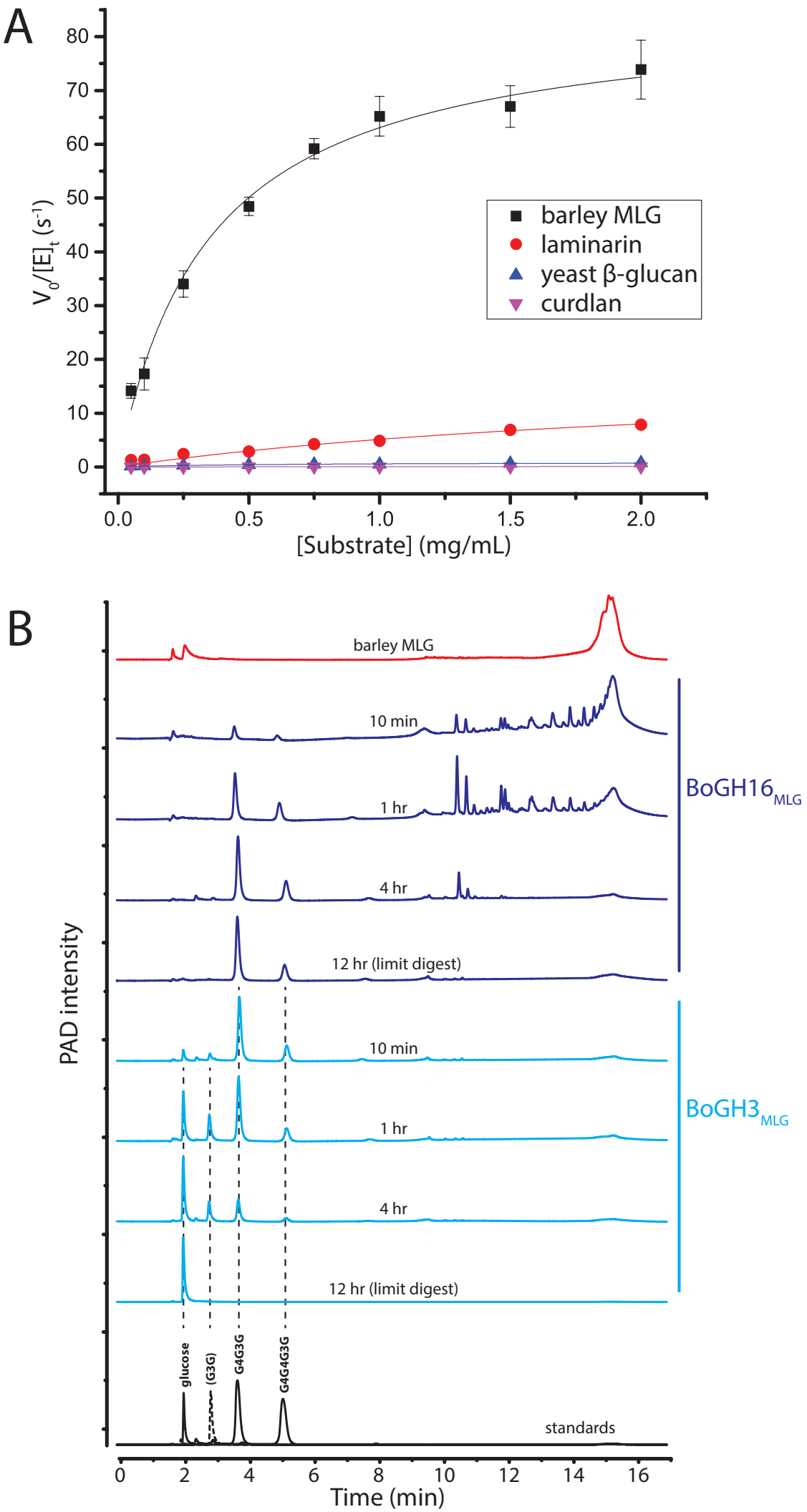

A
80
70
60
50
40
30
20
10
0
$V_0/[E]_t$ (s^{-1})
barley MLG
laminarin
yeast β-glucan
curdlan
0.0
0.5
1.0
1.5
2.0
[Substrate] (mg/mL)
B
barley MLG
10 min
1 hr
4 hr
12 hr (limit digest)
BoGH16MLG
10 min
1 hr
4 hr
12 hr (limit digest)
BoGH3MLG
PAD intensity
glucose
(G3G)
G4G3G
G4G4G3G
standards
0
2
4
6
8
10
12
14
16
Time (min)

Figure 4

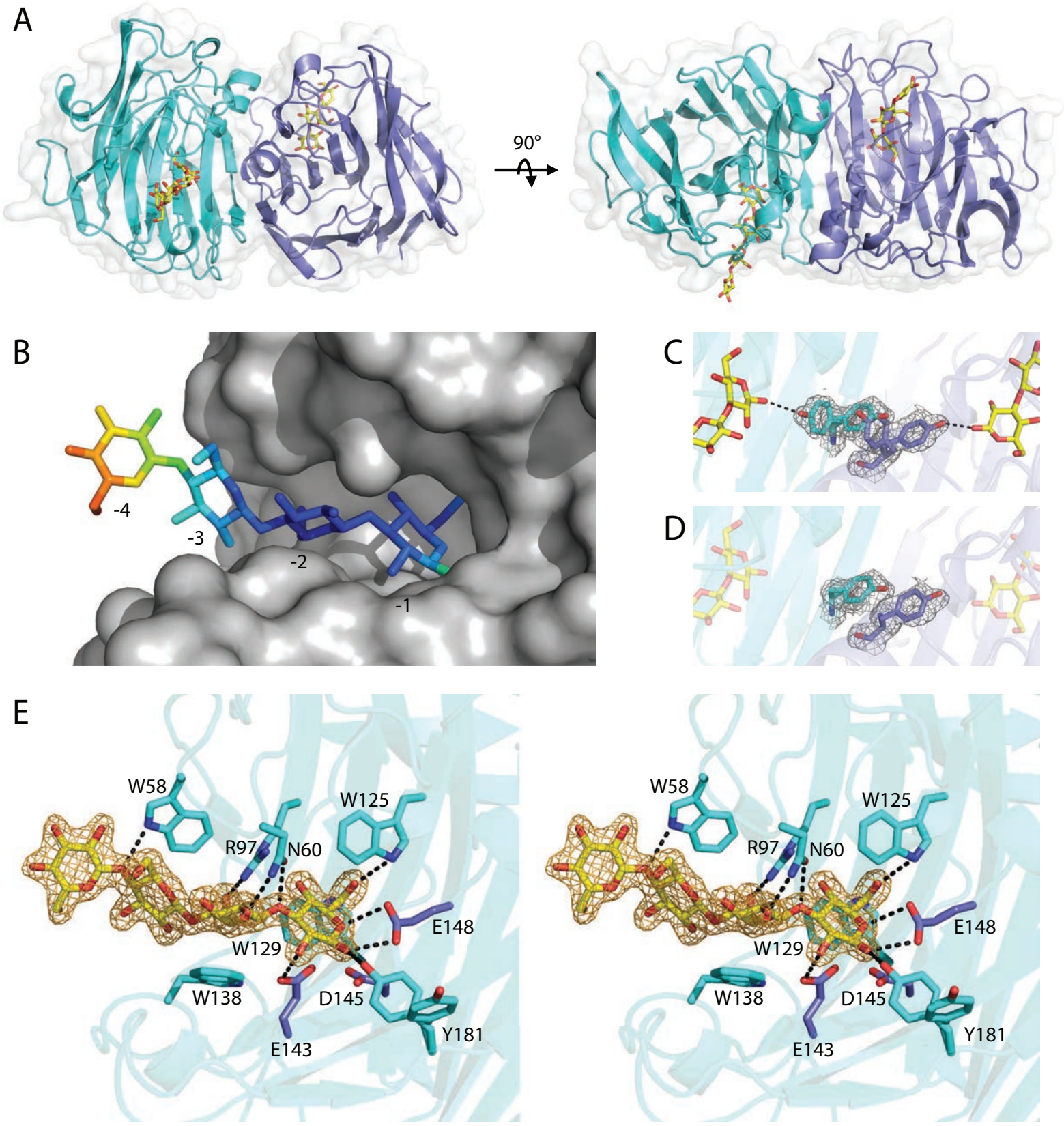

A
90°
B
-4
-3
-2
-1
C
D
E
W58
W125
R97
N60
E148
W129
W138
D145
Y181
E143
W58
W125
R97
N60
E148
W129
W138
D145
Y181
E143

Figure 5

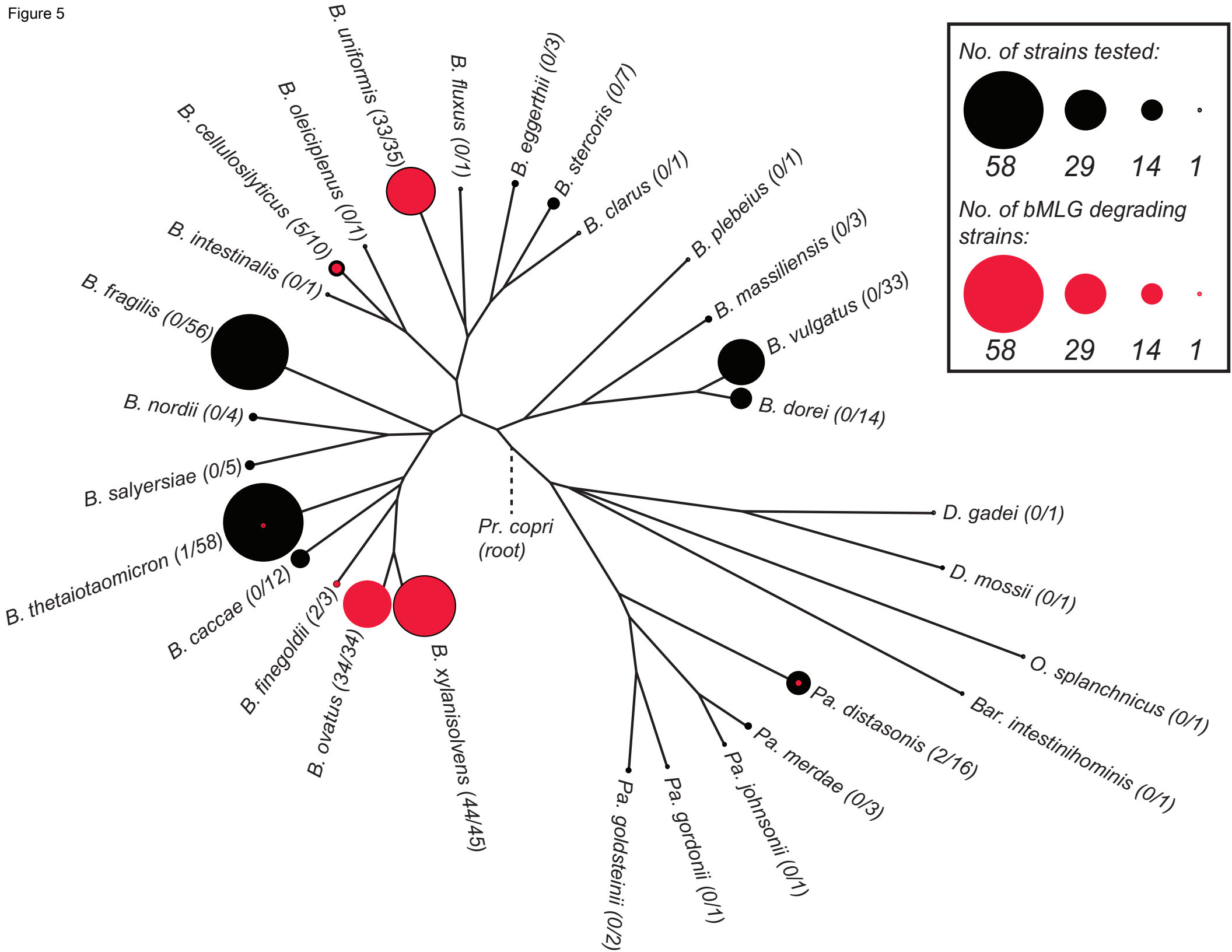

No. of strains tested:
58
29
14
1
No. of bMLG degrading strains:
58
29
14
1
B. uniformis (33/35)
B. fluxus (0/1)
B. eggerthii (0/3)
B. stercoris (0/7)
B. clarus (0/1)
B. plebeius (0/1)
B. massiliensis (0/3)
B. vulgatus (0/33)
B. dorei (0/14)
B. oleiciplenus (0/1)
B. cellulosilyticus (5/10)
B. intestinalis (0/1)
B. fragilis (0/56)
B. nordii (0/4)
B. salyersiae (0/5)
B. thetaiotaomicron (1/58)
B. caccae (0/12)
B. finegoldii (2/3)
B. ovatus (34/34)
B. xylanisolvens (44/45)
Pr. copri (root)
D. gadei (0/1)
D. mossii (0/1)
O. splanchnicus (0/1)
Bar. intestinihominis (0/1)
Pa. distasonis (2/16)
Pa. merdae (0/3)
Pa. johnsonii (0/1)
Pa. gordonii (0/1)
Pa. goldsteinii (0/2)

Figure 6

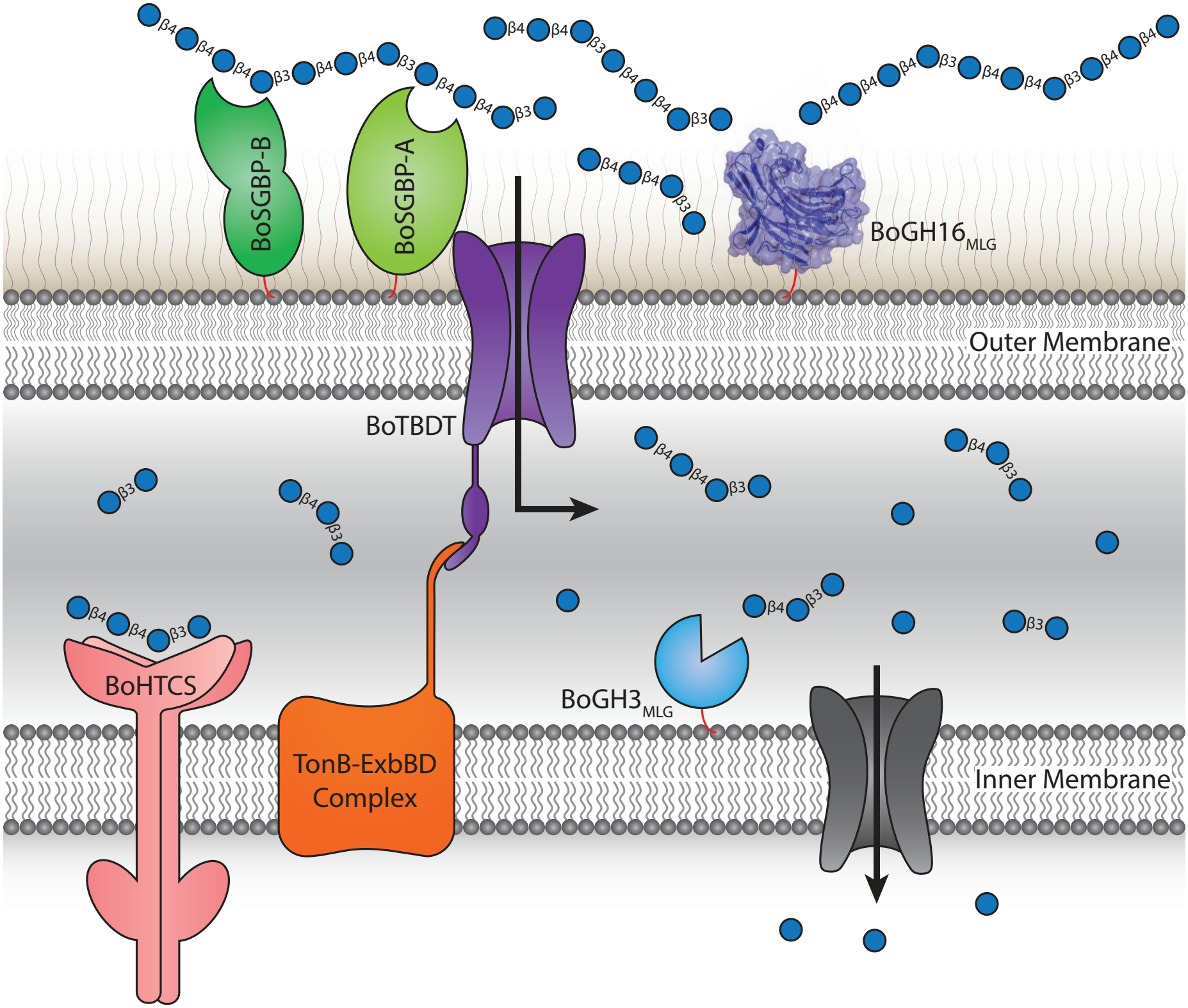
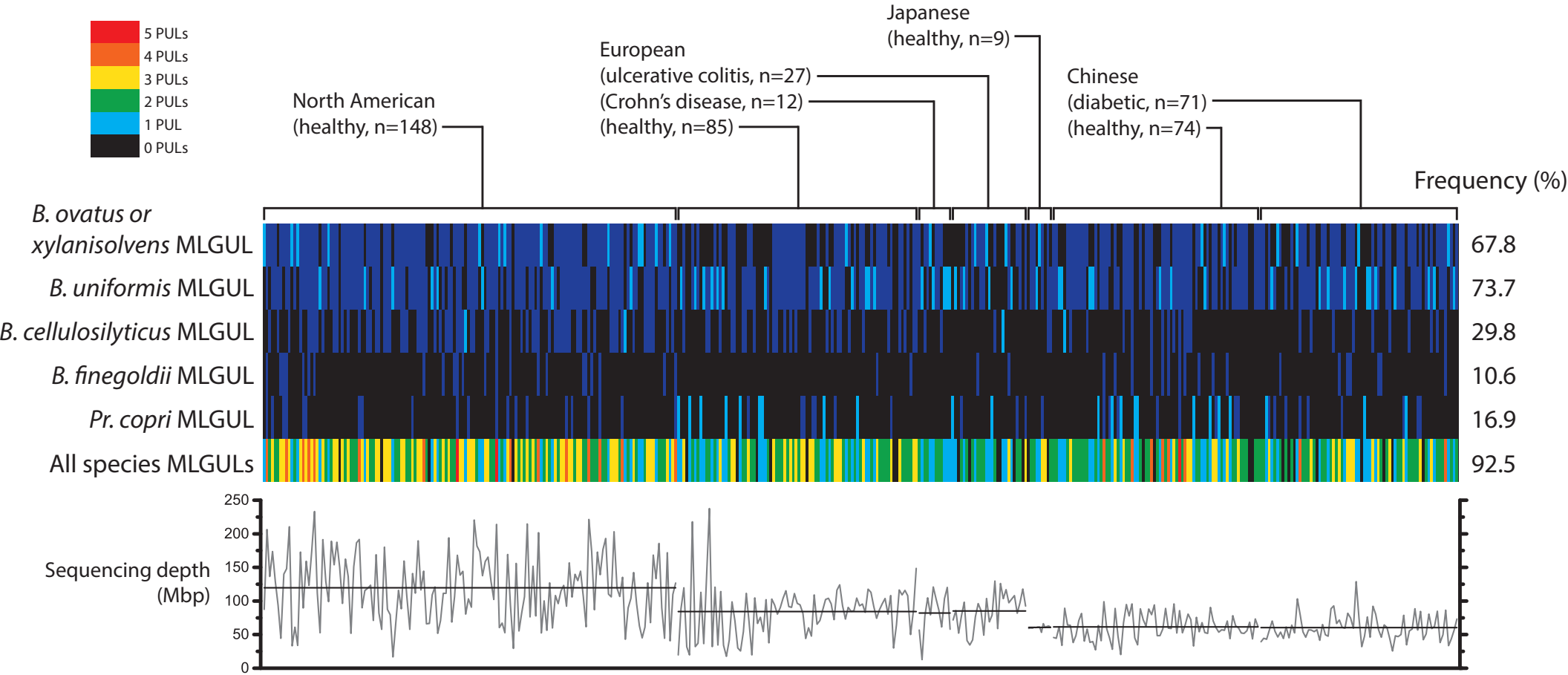

BoSGBP-B
BoSGBP-A
BoGH16MLG
Outer Membrane
BoTBDT
BoHTCS
TonB-ExbBD
Complex
BoGH3MLG
Inner Membrane

Figure 7


5 PULs
4 PULs
3 PULs
2 PULs
1 PUL
0 PULs
North American
(healthy, n=148)
European
(ulcerative colitis, n=27)
(Crohn's disease, n=12)
(healthy, n=85)
Japanese
(healthy, n=9)
Chinese
(diabetic, n=71)
(healthy, n=74)
Frequency (%)
B. ovatus or xylanisolvens MLGUL
67.8
B. uniformis MLGUL
73.7
B. cellulosilyticus MLGUL
29.8
B. finegoldii MLGUL
10.6
Pr. copri MLGUL
16.9
All species MLGULs
92.5
Sequencing depth (Mbp)
250
200
150
100
50
0

Supplemental Information

Supplemental Figures

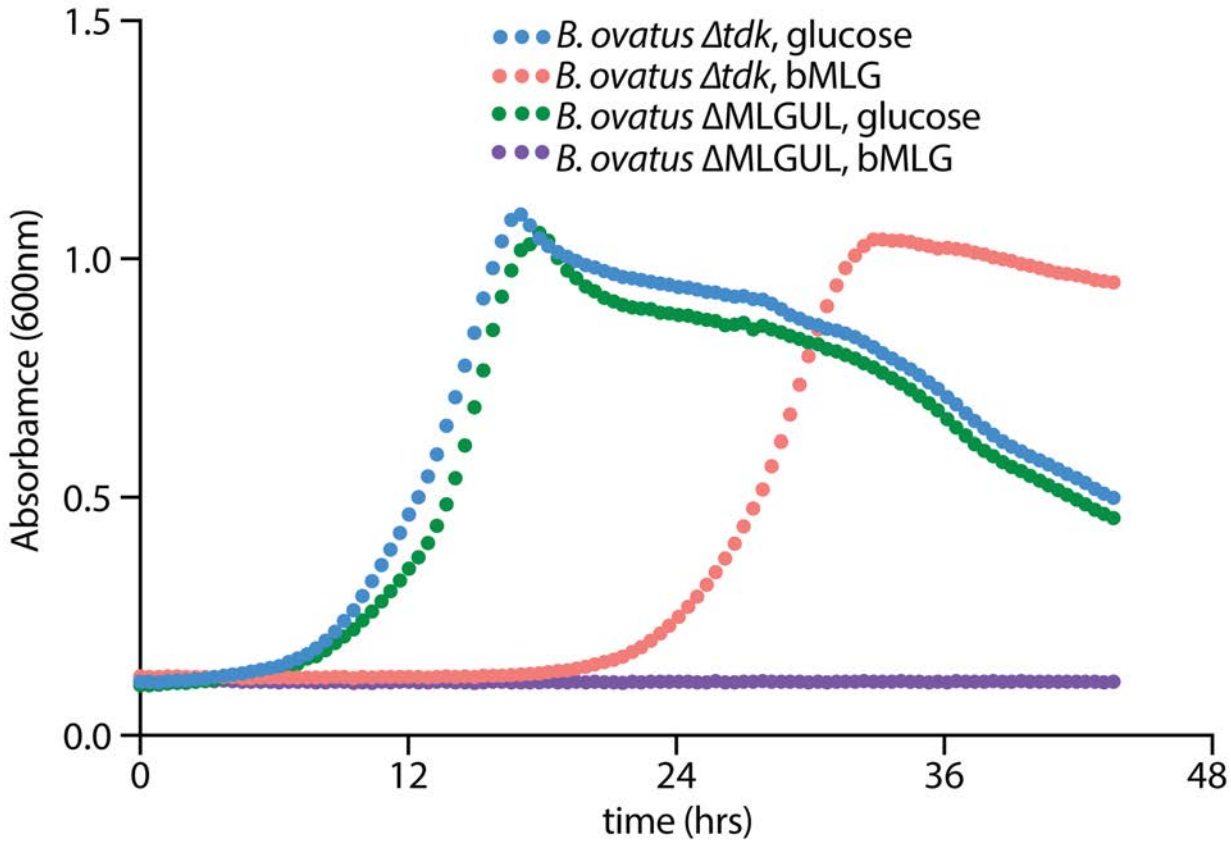


Figure S1, related to figure 1. Growth profile of *B. ovatus* strains.

B. ovatus Δtdk (wt) and *B. ovatus ΔMLGUL* (whole-PUL knockout) were cultured in minimal medium containing either glucose or bMLG as the sole carbon source (average of n = 2 growths per strain).

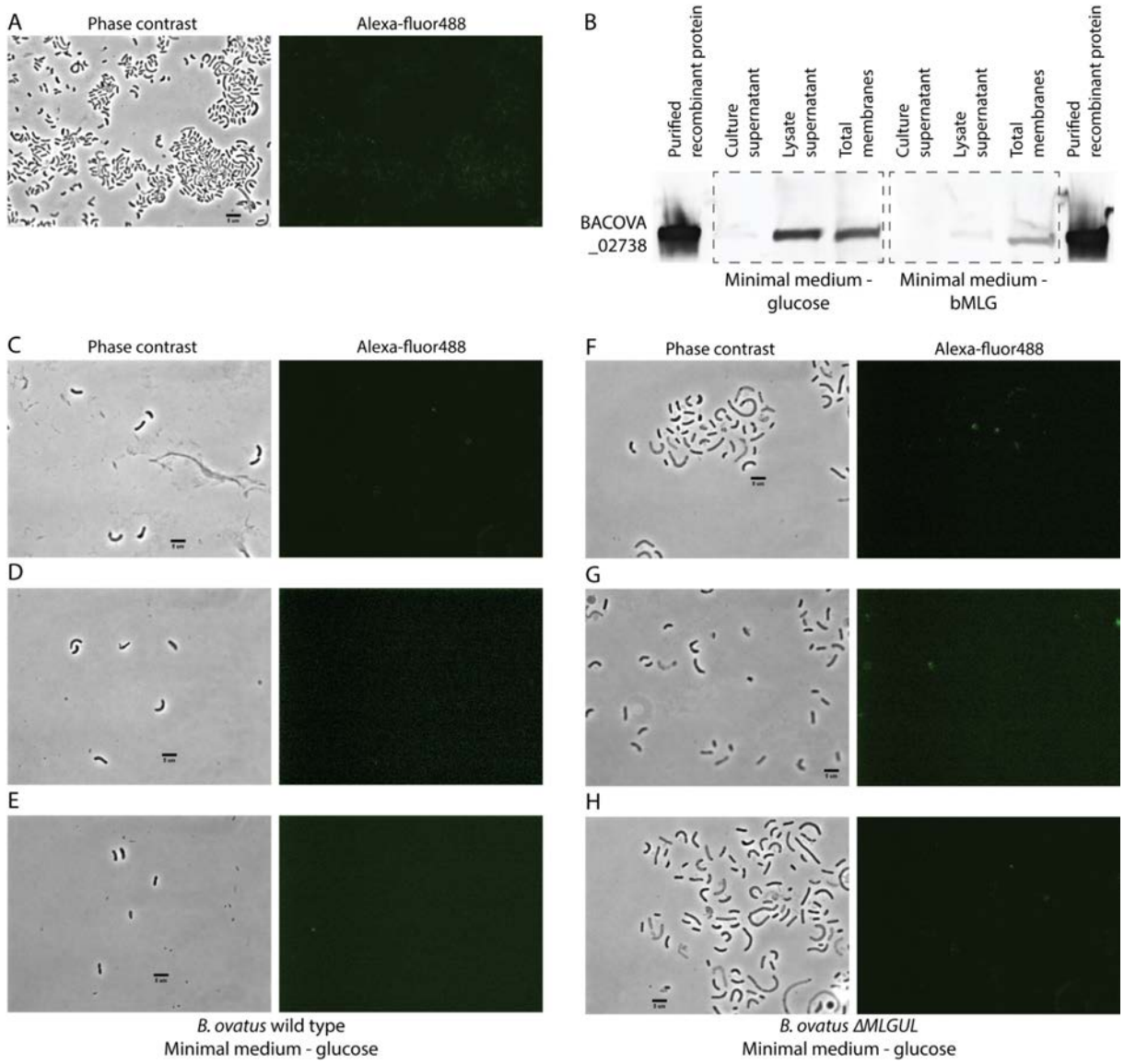


Figure S2, related to Figure 2. Enzyme localization analysis.

A: phase contrast and corresponding fluorescence microscope images of *B. ovatus* Δtdk cells grown in minimal medium with bMLG as the sole carbon source probed with custom polyclonal antibodies against recBACOVA_02738 (GH3). B: Western blot of protein collected from the culture supernatant, cell lysate supernatant, and cell lysate membrane fraction of *B. ovatus* Δtdk cells grown in minimal medium with glucose or bMLG as a sole carbon source. Phase contrast and corresponding fluorescence microscope images of wild type *B. ovatus* cells grown in minimal medium with glucose as the sole carbon source probed with custom polyclonal antibodies against recBoGH16$_{MLG}$ (C), recBoGH3$_{MLG}$ (D), and recBACOVA_02738(GH3) (E). Phase contrast and corresponding fluorescence microscope images of *B. ovatus ΔMLGUL* cells grown in minimal medium with glucose as the sole carbon source probed with custom polyclonal antibodies against recBoGH16$_{MLG}$ (F), recBoGH3$_{MLG}$ (G), and recBACOVA_02738(GH3) (H).

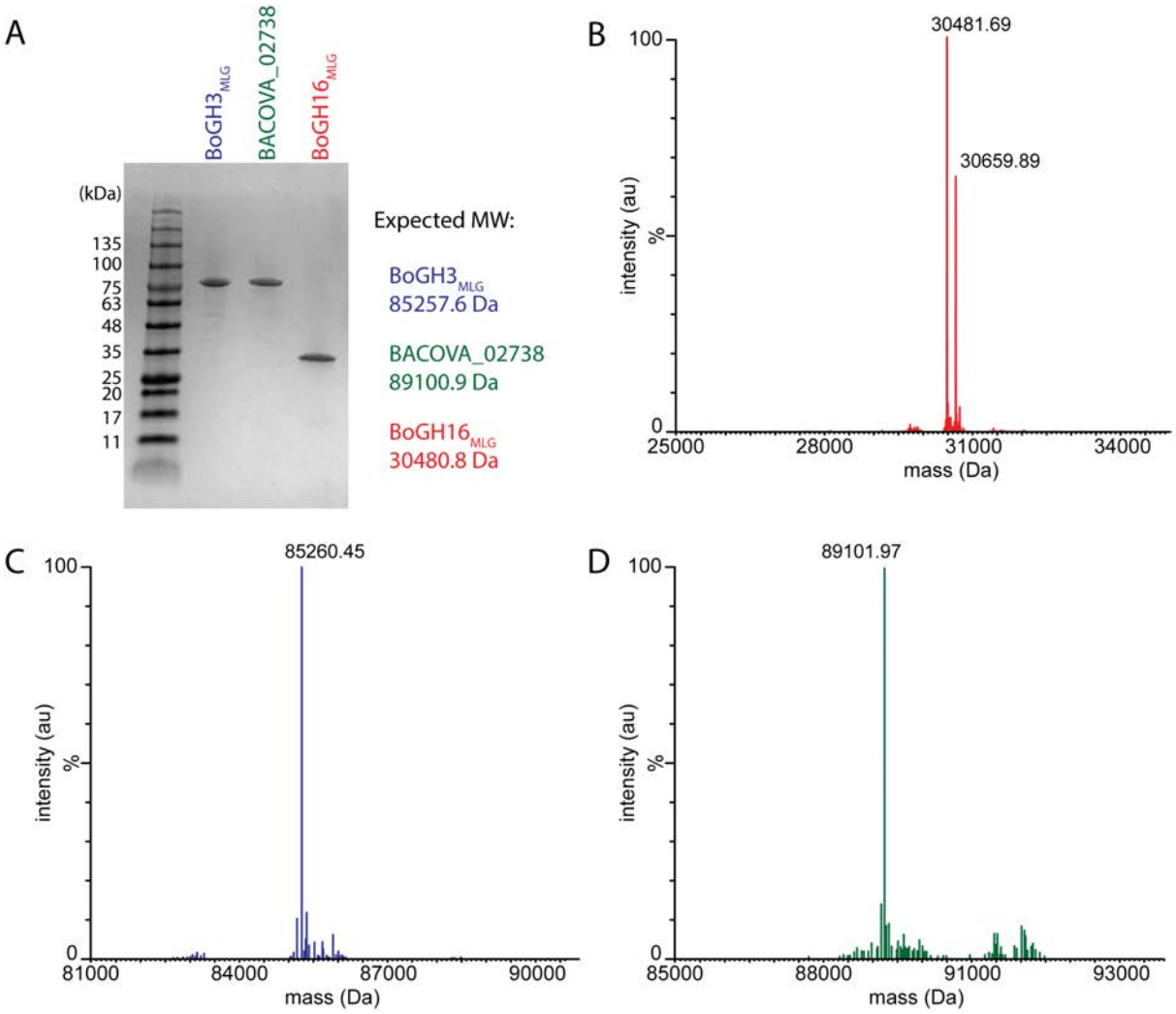


Figure S3, related to Figure 3. Purity and molecular mass of recombinant MLGUL proteins.

A: SDS-PAGE of recBoGH16$_{MLG}$, recBoGH3$_{MLG}$, and recBACOVA_02738(GH3). B: Reconstructed mass spectrum of intact recBoGH16$_{MLG}$. The second peak at +178.2 Da from the main peak is a species that has been spontaneously phosphogluconoylated at the N-terminal his-tag during production in *E. coli* (Geoghegan et al., 1999). C: Reconstructed mass spectrum of intact recBoGH3$_{MLG}$. D: Reconstructed mass spectrum of intact recBACOVA_02738(GH3).

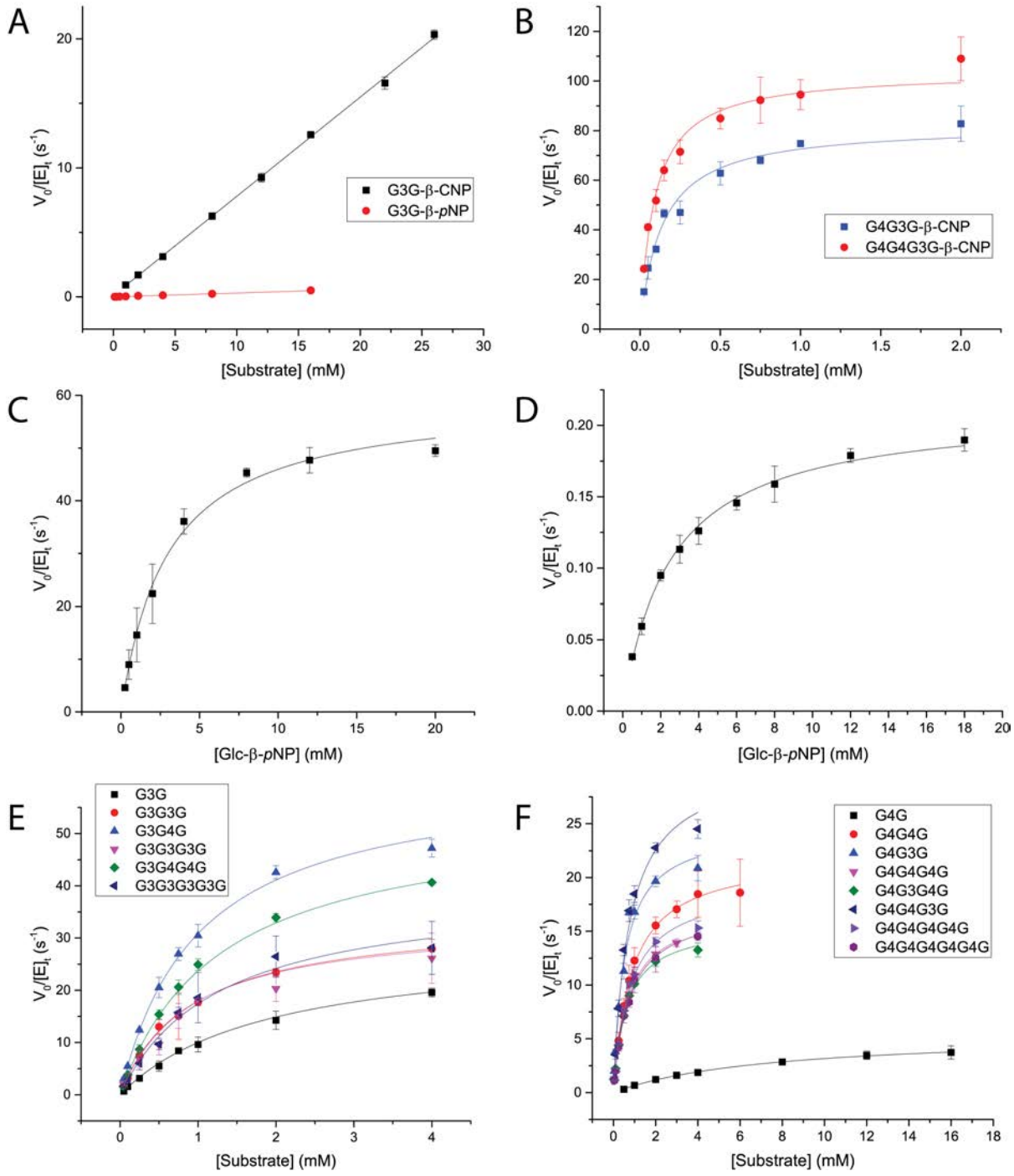


Figure S4, related to Figure 3. Initial-rate kinetic analysis of MLGUL GHs.

A: BoGH16$_{MLG}$ against G3G-β-CNP and G3G-β-*p*NP fitted to a linear equation. B: BoGH16$_{MLG}$ against G4G3G-β-CNP and G4G4G3G-β-CNP fitted to the Michealis-Menten equation. C: BoGH3$_{MLG}$ against glucose-β-*p*NP fitted to the Michealis-Menten equation. D: BACOVA_02738(GH3) against glucose-β-*p*NP fitted to the Michealis-Menten equation. E: BoGH3$_{MLG}$ against oligosaccharides with a β(1,3) bond at the non-reducing end. D: BoGH3$_{MLG}$ with oligosaccharides with a β(1,4) bond at the non-reducing end. Curve fitting was done on OriginPro 2015 and error bars represent standard deviations from the mean.

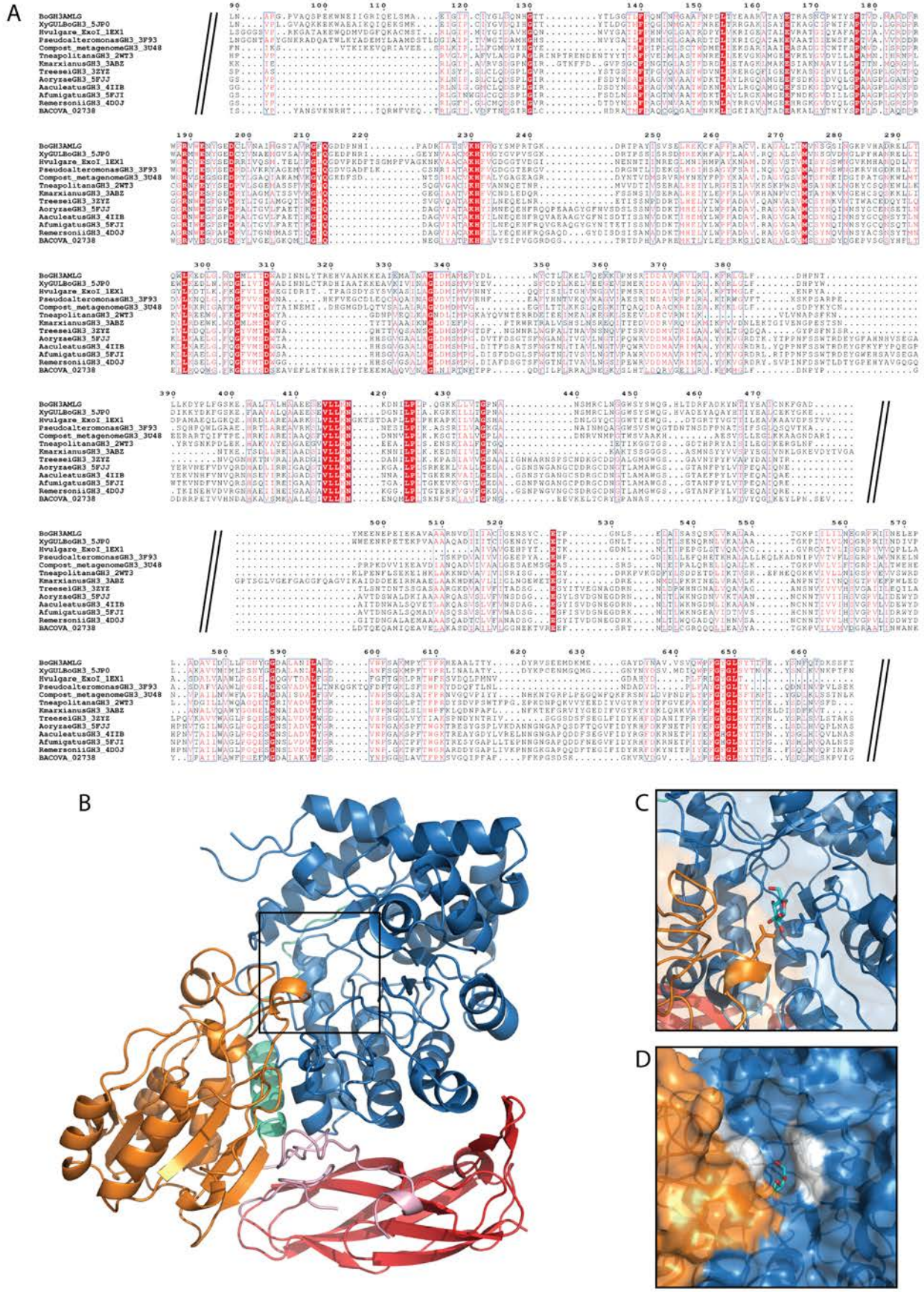

A
B
C
D

Figure S5, related to Figure 3. Primary and tertiary structure analysis of BoGH3$_{MLG}$.

A: Structure-based sequence aligmnment of BoGH3$_{MLG}$ and BACOVA_02738(GH3) with structurally characterized GH3 β-glucosidases. Portions of the alignment have been removed for brevity and breaks are indicated by double hash lines. Red highlights indicate invariant positions, blue outlines indicate similar positions and green arrows indicate catalytic residues. Alignment illustration created with ESPript (Gouet et al., 2003). B: Homology model of BoGH3$_{MLG}$ generated by Phyre2. The $(\alpha/\beta)_8$ TIM barrel (blue) is connected by a linker (teal) to a central α/β sandwich (orange), in turn connected by a linker (pink) to a C-terminal fibronectin type-III (FN-III) domain (red). The black box indicates the location of the active site. C: Catalytic site of BoGH3$_{MLG}$ with a bound glucose from the XyGUL BoGH3B overlay in the active site pocket. The catalytic nucleophile (Asp-309) and catalytic acid/base (Glu-527) are shown as sticks. D: Surface representation of the entrance to the BoGH3$_{MLG}$ active site with the two tryptophan residues that line the positive subsite shown in white.

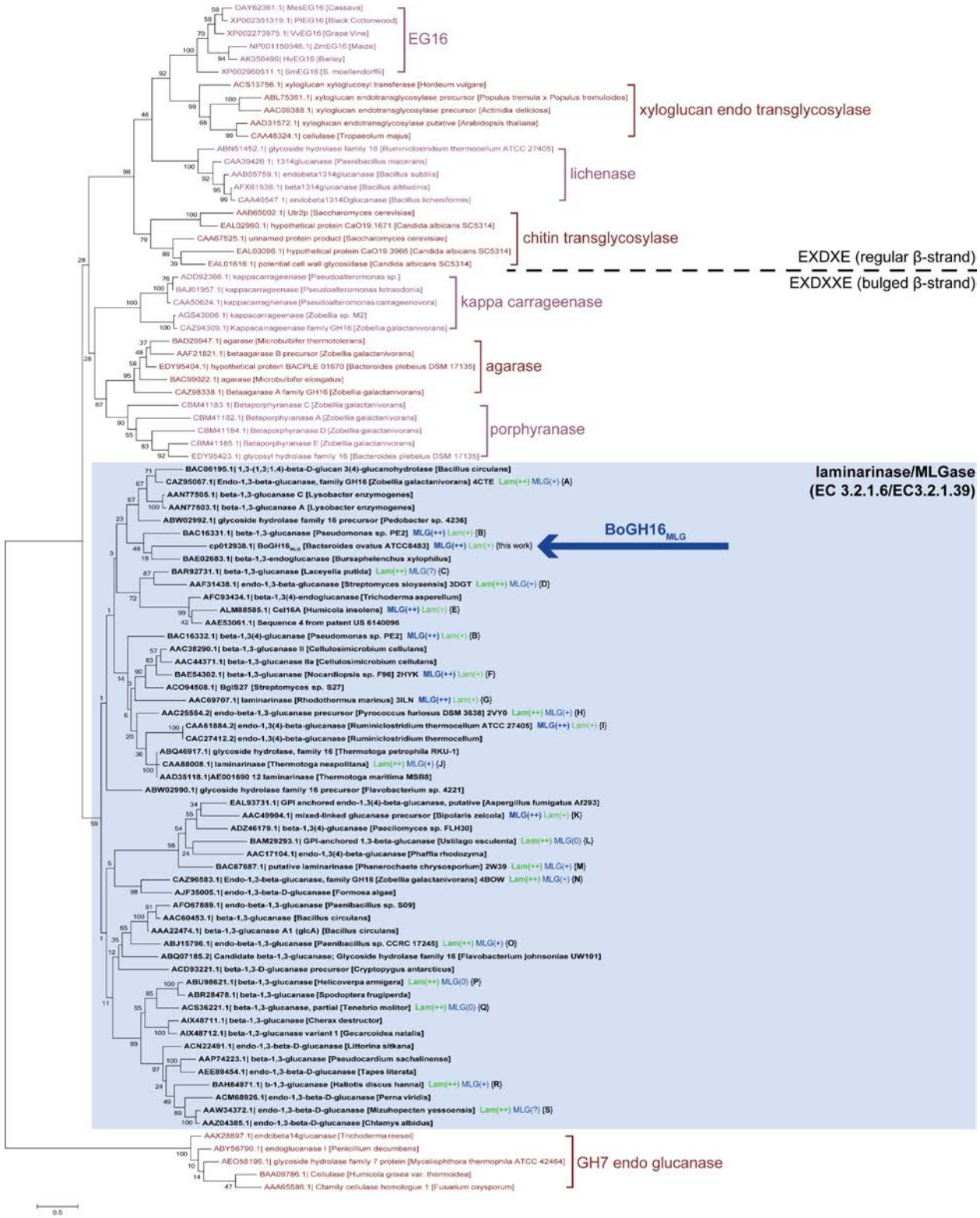

EG16
xyloglucan endo transglycosylase
lichenase
chitin transglycosylase
EXDXE (regular β-strand)
EXDXXE (bulged β-strand)
kappa carrageenase
agarase
porphyranase
laminarinase/MLGase
(EC 3.2.1.6/EC3.2.1.39)
BoGH16MLG
GH7 endo glucanase
0.5

Figure S6, related to Figure 4. Phylogenetic tree of characterized Glycoside Hydrolase Family 16 sequences.

Clades distinguishing major substrate specificities (see https://www.cazypedia.org/index.php/GH16) are represented by at least 5 members. Leaf names contain GenBank accession number, enzyme name, organism of origin, and a PDB code where available. Bootstrap values are shown for each node. The dotted horizontal line separates GH Family 16 enzymes with active-site residues on a regular β-strand from those with a bulged β-strand. The clade highlighted with a blue background is traditionally referred to as a "laminarinase" (EC 3.2.1.39) group, yet our current literature analysis also indicates the presence of predominant MLGases (EC 3.2.1.6). Where known, biochemically determined activities are shown next to select sequences, with the predominant activity displayed in bold type; (++) indicates better activity; (+) indicates poorer activity, (0) indicates no activity, and (?) indicates that specific activity data is unavailable. Reported activities are based on the following references, A: (Labourel et al., 2014), B: (Kitamura et al., 2002), C: (Kobayashi et al., 2016), D: (Hong et al., 2008), E: (Li et al., 2015), F: (Masuda et al., 2006), G: (Krah et al., 1998), H: (Ilari et al., 2009) , I: (Fuchs et al., 2003), J: (Zverlov et al., 1997), K: (Gorlach et al., 1998), L: (Nakajima et al., 2012), M: (Kawai et al., 2006), N: (Labourel et al., 2015), O: (Hong and Meng, 2003), P: (Pauchet et al., 2009), Q: (Genta et al., 2009), R: (Kumagai and Ojima, 2009), S: (Kovalchuk et al., 2006)

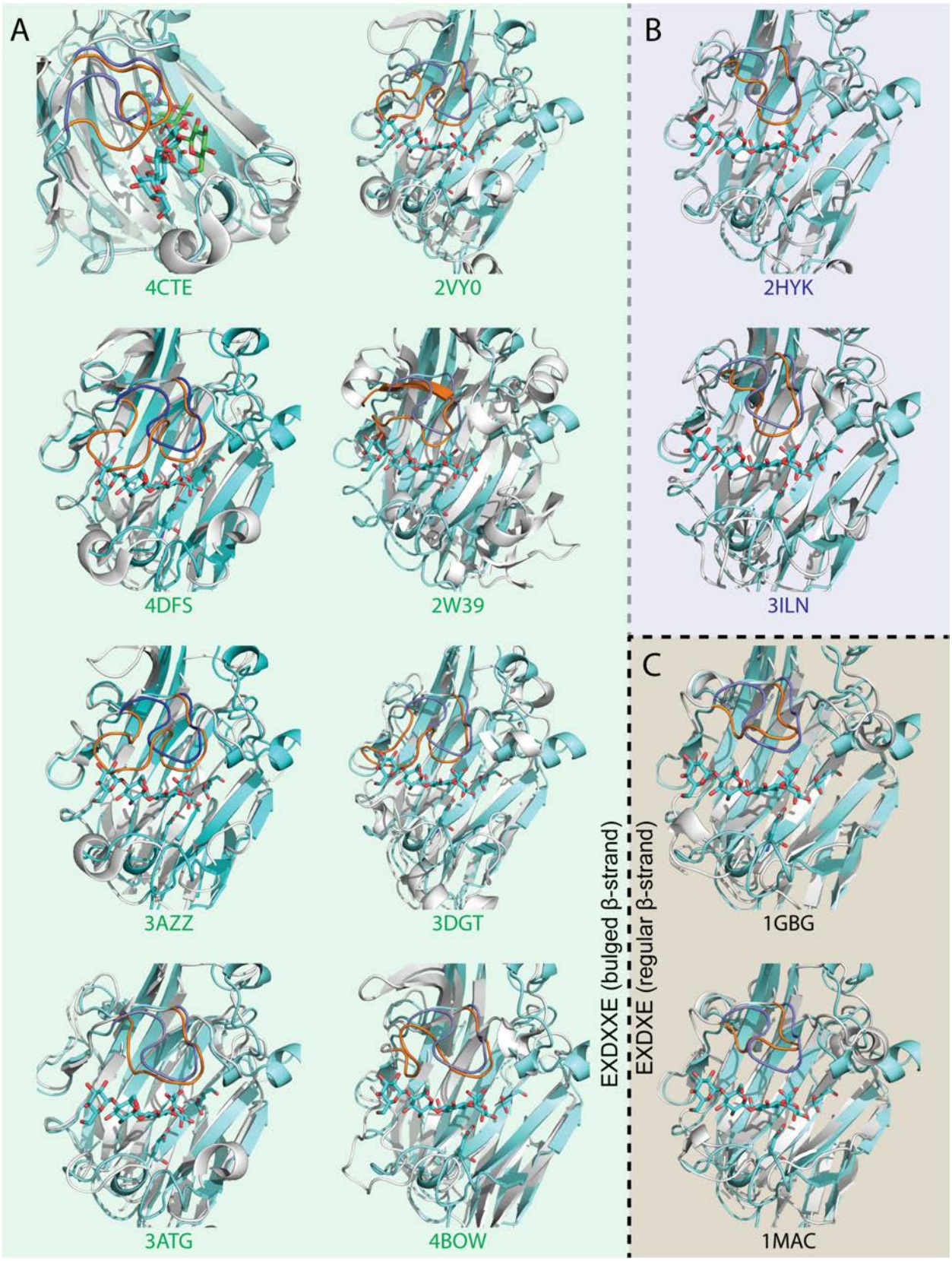

A
4CTE
2VY0
4DFS
2W39
3AZZ
3DGT
3ATG
4BOW
B
2HYK
3ILN
C
1GBG
1MAC
EXDXXE (bulged β-strand)
EXDXE (regular β-strand)

Figure S7, related to Figure 4. $BoGH16_{MLG}$ loop comparison with laminarinases and MLGases.

$BoGH16_{MLG}$ (cyan) is structurally aligned with all 10 available β-bulge-containing laminarinases/MLGases and two representative regular β-stranded canonical lichenases (white); PDB ID of the compared structure shown below each alignment. The $BoGH16_{MLG}$ loop is colored blue and the loop of the compared structure is colored orange. The $BoGH16_{MLG}$ catalytic residues and mixed-linkage oligosaccharide G4G4G3G in complex with $BoGH16_{MLG}$ are also shown in cyan. The thio-β-1,3-trisaccharide in complex with $ZgLamC_{GH16\text{-}E142S}$ (PDB code 4CTE) is shown in green
A: β-bulge-containing laminarinases, B: β-bulge-containing MLGases, C: regular β-stranded MLGases (canonical lichenases).

Supplemental Tables

Table S1, related to Figure 1. Transcriptional expression of MLGUL and neighboring genes.[a]

Locus tag	Putative protein ID	bMLG	MM-Glc	bMLG/MM-Glc
BACOVA_02736		879.4 ± 29.4	511.6 ± 94.3	1.7
BACOVA_02737		1072.3 ± 23.0	590.4 ± 147.6	1.8
BACOVA_02738	GH3	355.7 ± 46.9	229.0 ± 55.5	1.6
BACOVA_02739	Sigma 70, region 4	22.7 ± 0.2	37.2 ± 8.9	0.6
BACOVA_02740	**HTCS**	**376.8 ± 4.4**	**197.6 ± 0.1**	**1.9**
BACOVA_02741	**GH16$_{MLG}$**	**3657.6 ± 77.7**	**16.0 ± 16.5**	**228.8**
BACOVA_02742	**TBDT**	**3401.1 ± 375.5**	**13.1 ± 5.7**	**259.7**
BACOVA_02743	**SGBP-A**	**1837.5 ± 111.1**	**14.7 ± 6.3**	**124.9**
BACOVA_02744	**SGBP-B**	**2748.9 ± 103.9**	**9.2 ± 1.0**	**298.3**
BACOVA_02745	**GH3$_{MLG}$**	**2988.3 ± 53.7**	**16.5 ± 0.4**	**180.7**
BACOVA_02746[b]	Transposase	2763.9 ± 111.5	11.9 ± 9.7	232.8
BACOVA_02747[b]	Helicase	2821.0 ± 91.9	16.8 ± 5.8	167.6
BACOVA_02748		54.5 ± 10.5	25.8 ± 9.1	2.1
BACOVA_02749		62.3 ± 6.0	30.1 ± 6.8	2.1

[a] Microarray results of upregulation under bMLG induction normalized to a glucose background are shown with MLGUL genes in bold type. Values reported as averages and standard deviations of two biological replicates. Data are from (Martens et al., 2011).

[b] Loci BACOVA_02746 and BACOVA_02747 are predicted to encoded proteins of only 63 and 44 amino acids, respectively, and were therefore likely to have been originally mis-annotated based on limited regional sequence similarity. Despite apparently high transcript levels, which may result from read-through downstream of highly active operons whether there is protein coding function or not, these loci are not considered to be part of the MLGUL.

Table S2, related to Figure 3. Kinetic parameters of BoGH16$_{MLG}$ on polysaccharide substrates.

Substrate	k_{cat} (s^{-1})	K_m (mg mL^{-1})	$kcat/K_m$ (s^{-1} mg^{-1} mL)	Assay
barley MLG	85.8 ± 4.6	0.364 ± 0.051	238	BCA
laminarin	18.1 ± 5.25	2.54 ± 0.97	7.12	BCA
yeast β-glucan	0.875 ± 0.101	0.541 ± 0.169	1.61	BCA
curdlan	ND	ND	0.042	BCA

Data is only presented for substrates on which BoGH16$_{MLG}$ showed activity (no detectable activity on tamarind xyloglucan, beechwood xylan, wheat arabinoxylan, carob galactomannan, konjac glucomannan, carboxymethyl cellulose, hydroxyethylcellulose, xanthan gum, and ulvan as determined by BCA and HPLC analyses). ND: not determined (in cases where Michealis-Menten curve fitting was not feasible, individual k_{cat} and K_m values are not reported and k_{cat}/K_m value was determined from linear curve fit to initial rate data in the [S] << $K_{m(apparent)}$ range). Data are represented as curve fit parameters ± standard deviation.

Table S3, related to Figure 3. Kinetic parameters of BoGH16$_{MLG}$ on chromogenic substrates.

Substrate	k_{cat} (s^{-1})	K_m (mM)	$kcat/K_m$ (s^{-1} mM^{-1})	Assay
G-*p*NP	NA	NA	NA	*p*NP
G-CNP	NA	NA	NA	CNP
G4G-CNP	NA	NA	NA	CNP
G4G4G-CNP	NA	NA	NA	CNP
G3G-*p*NP	ND	ND	0.0298	*p*NP
G3G-CNP	ND	ND	0.768	CNP
G4G3G-CNP	82.3 ± 3.8	0.134 ± 0.018	614	CNP
G4G4G3G-CNP	103.7 ± 3.4	0.0895 ± 0.0097	1160	CNP

NA: no detectable activity. ND: not determined (in cases where Michealis-Menten curve fitting was not feasible, individual k_{cat} and K_m values are not reported and k_{cat}/K_m value was determined from linear curve fit to initial rate data in the $[S] << K_{m(apparent)}$ range). Data are represented as curve fit parameters ± standard deviation.

Table S4, related to Figure 4. Data collection and refinement statistics for BoGH16$_{MLG}$ structures.

	Apo-BoGH16$_{MLG}$ (5NBO)	G4G4G3G-BoGH16$_{MLG}$
Data collection		
Space group	C2	C2
Cell dimensions		
a, *b*, *c* (Å)	167.5, 60.8, 49.4	167.3, 60.2 49.7
α, β, γ (°)	90.0, 94.5, 90.0	90.0, 93.5, 90.0
Wavelength (Å)	0.976	0.979
Resolution (Å)	49.23–1.80 (1.84–1.80)	49.57-1.80 (1.84-1.80)
R_{merge}	0.067 (0.446)	0.094 (0.735)
$I / \sigma I$	10.0 (2.0)	10.8 (1.7)
Completeness (%)	98.9 (99.9)	99.0 (99.0)
Redundancy	3.0 (3.0)	3.9 (3.9)
Half-set correlation CC(1/2)	0.996 (0.669)	0.994 (0.516)
$R_{p.i.m.}$	0.053 (0.365)	0.066 (0.499)
Refinement		
Resolution (Å)	49.23–1.80	49.57-1.80
No. reflections (Work/Free)	43,072/2,392	45,346/2,384
R_{work} / R_{free}	0.161/0.204	0.175/0.212
No. atoms		
Protein	3,766	3,767
Ligand/solvent/ion	8	93
Water	353	262
Average B-factors (Å^2)		
Protein	18.6	21.6
Ligand/ion	25.4	29.9
Water	29.2	28.6
R.m.s deviations		
Bond lengths (Å)	0.018	0.013
Bond angles (°)	1.83	1.62

Values in parentheses represent data in the highest resolution shell.

Table S5, related to Figure 4. Dali search results.

PDB ID	Organism	Protein name	Z score	RMSD (Å)	% ID	Active site	Predominant activity
4CTE	*Zobellia galactanivorans*	ZgLamC	29.3	2.0	38	β-bulge	laminarinase[a]
2HYK	*Nocardiopsis sp. F96*	BglF	28.8	1.9	32	β-bulge	MLGase[b]
3ATG	*Cellulosimicrobium cellulans*	BglII	28.3	1.9	34	β-bulge	laminarinase[c]
3AZY	*Thermotoga maritima*	Lam16	27.8	1.9	32	β-bulge	laminarinase[d]
2VY0	*Pyrococcus furiosus*	LamA	27.3	1.8	31	β-bulge	laminarinase[e]
4DFS	*Thermotoga petrophila*	TpLam	27.3	2.1	31	β-bulge	laminarinase[f]
4BOW	*Zobellia galactanivorans*	ZgLamA	26.9	2.0	30	β-bulge	laminarinase[g]
3ILN	*Rhodothermus marinus*	LamR	26.4	2.2	29	β-bulge	MLGase[h]
1MAC	*Paenibacillus macerans*	Bgi	25.1	2.0	22	regular β-strand	MLGase[i]
3DGT	*Streptomyces sioyaensis*	Curd1	25.0	2.4	28	β-bulge	laminarinase[j]
1GBG	*Bacillus licheniformis*	Bg1	25.0	2.1	23	regular β-strand	MLGase[k]
3O5S	*Bacillus subtilis*	BglS	25.0	2.0	24	regular β-strand	MLGase[l]

[a] (Labourel et al., 2014)
[b] (Masuda et al., 2006)
[c] (Ferrer, 2006)
[d] (Jeng et al., 2011)
[e] (Ilari et al., 2009)
[f] (Cota et al., 2011)
[g] (Labourel et al., 2015)
[h] (Krah et al., 1998)
[i] (Hahn et al., 1995a)
[j] (Hong et al., 2008)
[k] (Hahn et al., 1995b)
[l] (Furtado et al., 2011)

Supplemental Experimental Procedures

Microbiology

Bacteroidetes reverse genetics and growth analysis

Flat bottom 96-well plates (Costar) were loaded with 100 µL of sterilized carbohydrate stock at 2× concentration. A 24-hour culture was centrifuged to pellet bacteria, resuspended in 2× MM-no carbohydrate (MM-NC) and used to inoculate MM-NC at a ratio of 1:50. Each carbohydrate array was loaded with 100 µL of the inoculated 2× medium to produce 200 µL cultures at a final bacteria ratio of 1:100. Assay plates were sealed with an optically clear gas-permeable polyurethane membrane (Diversified Biotech, Boston, MA) in an anaerobic chamber (Coy manufacturing, Grass Lake, MI). Plates were loaded into a Biostack automated plate handling device coupled to a Powerwave HT absorbance reader (both devices from Biotek Instruments, Winooski, VT). Absorbance at 600 nm (A_{600}) was measured for each well at 10–15 minute intervals.

Enzyme localization

Immunofluorescence microscopy.

Fluorescence microscopy was performed on fixed *Bacteroides ovatus Δtdk* and *ΔMLGUL* cells. The cells were grown to mid-exponential phase (A_{600} = 0.5-0.6) in Minimal Media (MM) with bMLG or glucose (0.5% w/v) as the sole carbon source. The cultures were then pelleted, and washed with phosphate-buffered saline (PBS). The cells were then fixed by incubation in formalin (4.5% formaldehyde in PBS) for 1.5 h at room temperature, washed with PBS, and blocked for 16 hours at 4 °C in blocking solution (2% goat serum (Sigma-Aldrich), 0.02% NaN_3, PBS). The cells were then incubated with individual polyclonal antibodies raised against recBoGH16$_{MLG}$, recBoGH3$_{MLG}$, and recBACOVA_02738(GH3) (Cedarlane Laboratories, Burlington, ON) for 2 hours at room temperature (1:500 dilution of the antibody in blocking solution). For secondary labelling, cells were pelleted, washed three times in 1 mL of PBS and resuspended in 0.4 mL goat anti-rabbit IgG Alexa-Fluor 488 (Thermo Fisher Scientific), diluted 1:500 in blocking solution, and incubated 1 hour at room temperature in the dark. The cells were then washed three more times and resuspended in 0.05 mL of PBS containing ProLong Gold Antifade (Thermo Fisher Scientific). Cells were mounted on agarose pads on glass slides and capped with coverslips. Fluorescence was imaged on an Olympus IX70 inverted microscope (Olympus, Tokyo, Japan) at 100 X magnification.

Immunoblotting analysis.

Bacteroides ovatus Δtdk cells were grown as described above in MM on bMLG (0.5% w/v) or glucose (0.5% w/v) as a sole carbon source. The cells were then centrifuged at 10000g for 45 minutes, resuspended in Tris-buffered saline (TBS), and lysed. After cell disruption, the membranes and cell debris were harvested by centrifuga-

tion for 1 hour at 42000 rpm (TLA 100.3 Beckman) at 4 °C. To prepare the total membrane fraction, the pellet was resuspended in 60 mM of n-octyl β-D-glucopyranoside, agitated for 1 hour at room temperature and centrifuged at 35000 rpm for 45 minutes at 4 °C. The supernatant was then harvested for further analysis.

The appropriate dilution of the culture supernatant, the lysate supernatant, and the total membrane fraction were added to 4X Laemmli buffer, boiled for 10 minutes, and run on an SDS polyacrylamide gel (Mini-PROTEAN® TGX™ gels, Bio-Rad). Transfer to a western blot Polyvinylidene difluoride (PVDF) membrane (Immobilon®-P) was performed for 45 minutes at 20 volts using a semi-dry transfer cells (Trans-Blot SD, Bio-Rad). The membranes were then blocked for 1 hour at room temperature with blocking buffer (5% milk in TBST buffer (Tris-Buffered Saline (TBS) with 0.1% Tween20)). The membranes were then washed three times with TBST buffer and the proteins of interest were revealed by incubation with the primary antibodies generated for $BoGH16_{MLG}$, $BoGH3_{MLG}$, and BACOVA_02738(GH3), diluted in blocking buffer (1:15000, 1:20000, and 1:15000 dilution respectively). After three more washes, the membranes were incubated for 1 hour at room temperature with the secondary antibody goat anti-rabbit IgG H&L (Alkaline Phosphatase; Abcam), diluted 1:25000 in blocking buffer solution. The membranes were then washed another three times and the immunodetection of the alkaline phosphatase enzyme on the membrane was revealed with Novex® AP Chromogenic Substrate (ThermoFisher Scientific).

Cloning, expression, and purification of recombinant enzymes

Gene sequences were obtained from the *B. ovatus* ATCC 8483 draft genome available on the Integrated Microbial Genomes database from the Joint Genome Institute. PCR primers were synthesized by Integrated DNA technologies.

Cloning

Open reading frames encoding BACOVA_02738, BACOVA_02742, and BACOVA_02745 were amplified by PCR using Q5 high fidelity polymerase (NEB) with appropriate primers (see below) and genomic *B. ovatus* DNA as template. All primers were designed to amplify constructs truncated to exclude predicted signal peptides (prediction by SignalP 4.1) and N-terminal lipidation cysteine residues (prediction by LipoP 1.0). NdeI and XhoI restriction sites were included in the forward and reverse primers of BACOVA_02742 for subsequent digestion (NdeI and XhoI from NEB) and ligation (T4 ligase from Thermo Scientific) into the pET28 vector. pMCSG complementary sequences were included in the forward and reverse primers of BACOVA_02738 and BACOVA_02745 for subsequent ligation independent cloning into pMCSG53 plasmids as per Eschenfeldt *et al.* (Eschenfeldt, William H.,

Stols, Lucy, Millard, Cynthia Sanville, Joachimiak, Andrzej, Donnelly, 2009). All three constructs were designed to harbor an N-terminal His_6-tag fusion in the translated recombinant peptide. Successful cloning was confirmed by colony PCR (GoTaq polymerase from Promega) and sequencing (Genewiz).

List of primers used for cloning.

Primer Name	Primer sequence (5' → 3')	Vector
$BoGH16_{MLG}$_F	GACGACCATATGTCGGATTCTGTTGGAACG	pET28
$BoGH16_{MLG}$_R	GACGACCTCGAGCTATAATATTTTCACCCA	pET28
$BoGH3_{MLG}$_F	TACTTCCAATCCAATGCCATGGTTCCCACTGCCATTCCTGAA	pMCSG53
$BoGH3_{MLG}$_R	TTATCCACTTCCAATGTTATTACTTGCATATAATATTCAGTGTTTGA	pMCSG53
BACOVA_02738_F	TACTTCCAATCCAATGCCATGAACAACAAACCTACTGATAACA	pMCSG53
BACOVA_02738_R	TTATCCACTTCCAATGTTATTATTGGACCTCAAAACTCCCCT	pMCSG53

Restriction sites are underlined and pMCSG LIC vector complementary sequences are double underlined.

Expression

Plasmids harboring the gene of interest were transformed into chemically competent *E. coli* BL21 (DE3) and cultured in lysogeny broth (LB) containing 50 μ g/mL kanamycin for BACOVA_02742 or 100 μ g/mL ampicillin for BACOVA_02738 and BACOVA_02745. Cells were grown on a large scale at 37 ° C until mid-logarithmic growth phase was reached (A_{600} = 0.4-0.6) at which point protein expression was induced by addition of isopropyl β-D-thioglactopyranoside to a final concentration of 0.5 mM and temperature was lowered to 16 ° C. Induction of recombinant protein production continued overnight after which the cells were collected by centrifugation at 4000 g for 20 minutes.

Purification

The harvested cell pellet was resuspended in binding buffer (20 mM sodium phosphate pH 7.4, 500 mM sodium chloride, 20 mM imidazole) and lysed using a Sonic Dismembrator F550 Ultrasonic Homogenizer (Fisher Scientific). Cell debris was pelleted by centrifugation at 15000 rpm for 45 minutes and the supernatant was loaded onto a 2 × 1 mL HisTrap IMAC FF nickel-nitrilotriacetic acid column (GE Healthcare), using a BioLogic FPLC system (BioRad). After washing with 10 column volumes of binding buffer, His_6-tagged protein was eluted using a linear gradient of 0 - 100% elution buffer (20 mM sodium phosphate pH 7.4, 500 mM sodium chloride, 500mM imidazole) over 10 column volumes. Fractions were monitored by A_{280} and eluted protein fractions were pooled and buffer

exchanged into 50 mM sodium phosphate pH 7.0 using Vivaspin centrifugal filters (GE Healthcare). After concentrating, aliquots were flash frozen in liquid nitrogen and stored at -80 ° C. Protein purity was determined by SDS-PAGE analysis and mass was confirmed by intact protein mass spectrometry on a Waters Xevo Q-TOF with nanoACQUITY UPLC system, as described previously (Sundqvist et al., 2007). Protein concentrations were determined by spectrophotometry on an Epoch Microplate Spectrophotometer (BioTek) using the following molar extinction coefficients, which were calculated using ProtParam tool from the ExPASy Bioinformatics Resource Portal (Gasteiger et al., 2005): 108555 $M^{-1}cm^{-1}$ for BACOVA_02738, 54890 $M^{-1}cm^{-1}$ for BACOVA_02742, and 108180 $M^{-1}cm^{-1}$ for BACOVA_02745.

Typical yields after purification were 80 mg of recBoGH16$_{MLG}$ from 1 L of lysogeny broth (LB) culture, 70 mg of recBoGH3$_{MLG}$ from 1 L of LB culture, and 8 mg of recBACOVA_02738(GH3) from 1 L terrific broth culture (Fig. S3).

Enzyme kinetics and product analysis

Substrates and polysaccharides

Polysaccharides

Beta-glucan (barley) high viscosity (bMLG), yeast beta-glucan, curdlan, tamarind xyloglucan, konjac glucomannan, carob galactomannan, wheat arabinoxylan, beechwood xylan were purchased from Megazyme International (Bray, Ireland). Laminarin (from *Laminaria digitata*) was purchased from Sigma Aldrich (St. Louis, MO, USA). Carboxymethyl cellulose was purchased from Acros Organics (Morris Plains, NJ, USA). Hydroxyethyl cellulose was purchased from Amresco (Solon, OH, USA). Xanthan gum was purchased from Spectrum (New Brunswick, NJ, USA). Ulvan (from Ulva sp.) was purchased from Elicityl (Crolles, France). Laminarin was reduced to laminaritol as described previously (Abdel-Akher and Smith, 1951).

Oligosaccharides

Cellobiose (G4G) was purchased from Acros Organics. Cellotriose (G4G4G), cellotetraose (G4G4G4G), cellopentaose (G4G4G4G4G), cellohexaose (G4G4G4G4G4G), laminaribiose (G3G), laminaritriose (G3G3G), laminaritetraose (G3G3G3G), laminaripentaose (G3G3G3G3G), mixed-linkage glucotriose A (G3G4G), mixed-linkage glucotriose B (G4G3G), mixed-linkage glucotetraose A (G3G4G4G), mixed-linkage glucotetraose B (G4G4G3G), mixed-linkage glucotetraose C (G4G3G4G) were purchased from Megazyme. Gentiobiose (G6G) was purchased from Carbosynth (Compton, UK).

Chromogenic substrates

para-nitrophenyl (*p*NP) glycosides of β-glucoside (G-β-*p*NP), α-glucoside, β-galactoside, β-mannoside, and β-xyloside were purchased from Sigma Aldrich. *ortho*-chloro-*para*-nitrophenyl (CNP) glycosides of G4G3G (G4G3G-CNP) and G4G4G3G (G4G4G3G-CNP), and *p*NP β-laminaribioside (G3G-*p*NP) were purchased from Megazyme. CNP glycosides of cellobioside (G4G-CNP) and that of cellotriose (G4G4G-CNP) were purchased from Carbosynth. G3G-CNP was synthesized by glycosylation of the known α-laminaribiosyl bromide (Viladot et al., 1997) and the corresponding phenol under phase-transfer conditions (Ibatullin et al., 2008; Viladot et al., 1997), the details of which will be published elsewhere.

Enzyme kinetics

BCA endpoint assay

Polysaccharide hydrolysis was quantified using the bicinchoninic acid (BCA) reducing sugar assay. All reactions were carried out in 100 µL volumes in the optimum pH buffer (50 mM sodium citrate pH 6.5 for $BoGH16_{MLG}$) at 37 ° C unless otherwise specified. Reactions were initiated by adding 10 µL of enzyme solution to 90 µL of the remaining assay mixture, which had been pre-incubated at 37 ° C. Reactions were terminated by addition of equal volume (100 µL) of BCA reagent and developing the color by boiling at 80 ° C for 20 minutes. Absorbance at 563 nm (A_{563}) was measured in 96-well plates on an Epoch Microplate Spectrophotometer (BioTek). Blank absorbance readings were determined for each polysaccharide at each concentration by using inactivated enzyme (denatured by boiling at 100 ° C for 10 minutes). Reducing ends released were quantified with a glucose standard curve (25 – 150 µM). All kinetic assays were conducted in technical triplicates.

Activity on a library of polysaccharides was initially screened by incubating 10 µM $BoGH16_{MLG}$ with 1.0 mg/mL substrate for 24 hours. The polysaccharide was determined to be a substrate for $BoGH16_{MLG}$ if the A_{563} increased significantly compared to the blank.

The pH optimum of $BoGH16_{MLG}$ was determined by incubating 7.5 nM enzyme with 1.0 mg/mL bMLG for 10 minutes in different buffers at 50 mM: sodium citrate (pH 3.0 – 6.5), sodium phosphate (pH 6.5 – 8.5), glycylglycine (pH 7.5 – 9.0), glycine (pH 9.0 – 10.5). Released reducing ends were measured as described above.

The temperature optimum of $BoGH16_{MLG}$ was determined by incubating 7.5 nM enzyme with 1.0 mg/mL bMLG for 10 minutes at various temperatures ranging from 30 to 70 ° C. Released reducing ends were measured as described above.

For initial-rate saturation kinetics, the following concentrations of enzyme were used: 4.9 nM for bMLG, 48.6 nM for laminarin, 485.6 nM for yeast beta-glucan, and 4.9 μM for curdlan. These are the concentration that were optimized for the reaction to be in the initial, linear stage of the reaction (less than 10% conversion) after 12 minutes of hydrolysis. To determine Michaelis-Menten parameters, eight different concentrations of each substrate were hydrolyzed by appropriate concentration of enzyme for 10 minutes after which the reaction was quenched and reducing ends released were quantified as described above.

Chromogenic substrate assay

Reaction with *p*NP and CNP glycoside substrates was used to quantify the hydrolysis of chromophore from the aglycone. Enzyme concentrations used to maintain initial-rate conditions were 6.9 nM for $BoGH3_{MLG}$, 941 nM for BACOVA_02738(GH3), 9.4 nM for $BoGH16_{MLG}$ against the G4G3G-CNP and G4G4G3G-CNP, 446 nM for $BoGH16_{MLG}$ against G3G-CNP, and 942 nM for $BoGH16_{MLG}$ against G3G-*p*NP.

Endpoint assays were used for pH and temperature optima of $BoGH3_{MLG}$ and BACOVA_02738(GH3). Enzyme, 1 mM G-*p*NP, and 50mM of the same range of different pH buffers described above were mixed to a final volume of 100 μL. The reactions were also carried out in optimal pH buffer (50 mM sodium phosphate pH 7.5 for $BoGH3_{MLG}$ and 50 mM sodium citrate pH 6.5 for BACOVA_02738(GH3)) at various temperatures ranging from 30 to 70 ° C. Reactions were terminated after 10 minutes by addition of 100 μL of 1 M Na_2CO_3 to raise the pH and absorbance at A_{405} was measured in 96-well plates on an Epoch Microplate Spectrophotometer (BioTek). An extinction coefficient of 18,100 M^{-1} cm^{-1} was used for these assays.

Continuous assays were used for initial-rate saturation kinetics. Reactions, carried out in 250 μL volumes in the optimum pH buffer at 37 ° C, were initiated by adding 25 μL of enzyme solution to 225 μL of the remaining assay mixture, pre-incubated at 37 ° C. Release of *p*NP or CNP was monitored by following absorbance at 405 nm in quartz cuvettes using a Cary 60 UV-Vis spectrophotometer (Agilent Technologies). Eight different substrate concentrations were assayed and rate was calculated using an extinction coefficient of 15298 M^{-1} cm^{-1} for CNP in sodium citrate pH 6.5, 3311 M^{-1} cm^{-1} for *p*NP in sodium citrate pH 6.5, and 12511 M^{-1} cm^{-1} for *p*NP in sodium phosphate pH 7.5

HK/G6PDH coupled assay

Release of glucose monosaccharides was quantified using the D-Glucose HK Assay Kit from Megazyme, modified for use as a continuous assay. All reactions were carried out in 250 µL volumes at 37 ° C in the triethylamine pH 7.6 buffer provided in the kit. $BoGH3_{MLG}$ concentrations used to maintain initial-rate conditions were 9.2 nM for laminari-oligosaccharides and mixed-linkage oligosaccharides, 50.1 nM for cello-oligosaccharides, and 1.28 µM for gentiobiose. Reactions were initiated by adding 25 µL of enzyme solution to 225 µL of the remaining assay mixture containing hexokinase, glucose-6-phosphate dehydrogenase, ATP, and $NADP^+$, pre-incubated at 37 ° C. The release of glucose monosaccharides corresponds stoichiometrically with the reduction of a molecule of $NADP^+$ to NADPH, which was monitored by following absorbance at 340 nm on a Cary 60 UV-Vis spectrophotometer. An extinction coefficient of 6,220 M^{-1} cm^{-1} was used to convert to rate of hydrolysis.

Enzyme limit digest assay

To determine limit-digestion products of $BoGH16_{MLG}$, 10 µM enzyme was incubated with 1.0 mg/mL polysaccharide in 1 mL of 50 mM sodium citrate pH 6.5 for 24 hours at 37 ° C. To determine limit-digestion products of $BoGH3_{MLG}$ and BACOVA_02738(GH3), 10 µM enzyme was incubated with the limit digest product of $BoGH16_{MLG}$ hydrolysis of 1.0 mg/mL polysaccharide in 1 mL of the appropriate buffer for 24 hours at 37 ° C. 10 µL of the reaction was diluted into 1 mL of ultrapure water and analyzed on HPAEC-PAD and HILIC-MS as described below.

The same experiment was conducted to observe reaction progress, except 10 nM of $BoGH16_{MLG}$ and 12 nM of $BoGH3_{MLG}$ were used and reactions were stopped at various time points by taking 100 µL of the reaction mixture and adding to 100 µL of NH_4OH. 20 µL of the reaction was diluted into 1 mL of ultrapure water and analyzed on HPAEC-PAD as described below.

Carbohydrate analytical methods

HPAEC-PAD product analysis

HPAEC-PAD was performed on a Dionex ICS-5000 HPLC system operated by Chromelion software version 7. Samples were separated on a 3 × 250 mm Dionex Carbopac PA200 column (Thermo Scientific). Solvent A was ultrapure water, solvent B was 1 M sodium hydroxide, and solvent C was 1 M sodium acetate. Conditions used were 0 – 5 min, 10 % B (initial conditions); 5 – 12 min, 10 % B, linear gradient from 0 – 30 % C; 12.0 –12.1 min, linear gradient from 10 – 50 % B, linear gradient from 30 – 50 % C; 12.1 – 13.0 min, exponential gradient of B and C back to initial conditions; 13 – 17 min, initial conditions.

HILIC-MS product analysis

Samples were separated by hydrophilic interaction liquid chromatography on a TSKgel Amide-80 column (Tosoh Bioscience). Solvent A was ultrapure water and solvent B was 1 M acetonitrile. The mobile phase used was a linear gradient of 35 % A and 65 % B to 50 % A and 50 % B over 30 minutes. The eluent was split between an evaporative light scattering detector (ELSD) (Agilent Technologies) and the Bruker Esquire 3000 Plus ion trap mass spectrometer (Bruker Daltonics). The eluent was ionized in positive mode by electrospray ionization before detection by ion trap. The ELSD and total ion count chromatograms were identical to the HPAEC-PAD trace. Esquire HyStar software was used to process the mass spectrometry data (Bruker Daltonics).

X-ray crystallography

Pure $BoGH16_{MLG}$ at 23 mg/ml in 50 mM sodium phosphate pH 7, was used to set up initial sitting drop crystal screens using a Mosquito robot (TTP Labtech). An initial hit condition was identified in the PACT screen (Qiagen) condition B9: 100 mM MES pH 6, 200 mM LiCl, 20 % w/v PEG-6000. Crystals were readily reproduced by hand in larger sitting drops by screening around this condition, varying only the PEG-6000 concentration from 15 to 25%. The crystals obtained from these optimizations were used in all subsequent work.

Crystals of the apo protein were cryo-cooled for data collection by first soaking in a solution of mother liquor supplemented with 18% ethylene glycol for 30 seconds before plunging in liquid nitrogen. Diffraction data were collected from these crystals at Diamond Light Source, beamline I03 at a wavelength of 0.976 Å. Data were indexed and integrated using XDS (Kabsch, 2010) with all subsequent data processing performed using the CCP4 software suite (Winn et al., 2011). A search model for molecular replacement was prepared using a single subunit from the *Zobellia galactanivorans* laminarinase $ZgLamC_{GH16\text{-}E142S}$ (PDB code 4CRQ) (Labourel et al., 2015) and using CHAINSAW (Stein, 2008) to trim any sidechains in the model to the nearest common atom based on a sequence alignment. The structure was then determined using this model by molecular replacement in PHASER (McCoy et al., 2007). Following density modification in PARROT (Cowtan, 2010), BUCCANEER (Cowtan, 2007) was used to construct an initial model before further model building and refinement were performed in COOT (Emsley and Cowtan, 2004) and REFMAC5 (Murshudov et al., 2011) respectively.

To obtain the G4G4G3G complex structure, crystals were soaked for 30 minutes in cryo-protectant solution (100 mM MES pH 6, 200 mM LiCl, 25 % w/v PEG-6000, 18 % w/v ethylene glycol) in which the ligand had been dissolved at 50 mM. Crystals were then plunged in liquid nitrogen ready for data collection. X-ray data were collect-

ed from these crystals at Diamond Light Source, beamline I02 at a wavelength of 0.979 Å. Data were processed as above using XDS (Kabsch, 2010) for indexing and integration followed by subsequent processing in the CCP4 software suite (Winn et al., 2011). Since the crystals were isomorphous to the apo-structure, the apo model with waters and flexible loops removed was refined against these new data. The model was rebuilt and refined using COOT (Emsley and Cowtan, 2004) and REFMAC5 (Murshudov et al., 2011).

For both structures, the quality of the model was monitored throughout using MOLPROBITY (Davis et al., 2007) - the final models having no outliers, 98.5 % and 98.7 % of residues in the favored region of the Ramachandran plot for the apo- and G4G4G3G-complex respectively. Additionally, the sugar conformations in the G4G4G3G-BoGH16$_{MLG}$ complex were all confirmed as $^{4}C_{1}$ chairs using PRIVATEER (Agirre et al., 2015) and the generated restraints applied during structure refinement. Data processing and refinement statistics for both structures can be found in Table S4. The apo- and G4G4G3G-complex structures have been deposited in the Protein Data Bank with accession codes 5NBO and 5NBP respectively.

Bioinformatics

Phylogenetic analysis

Glycoside Hydrolase Family 16 sequences with EC number 3.2.1.6 and 3.2.1.39 were extracted from the CAZy Database (URL http://www.cazy.org) using the Extract Sequences tool (URL http://research.ahv.dk/cazy/extract). The sequences were initially aligned by MUSCLE (Edgar, 2004) in AliView (Larsson, 2014) and manually trimmed to remove amino acids outside of the GH16 catalytic domain. The resulting sequences were structurally aligned using T-Coffee Expresso (Armougom et al., 2006), then further manually refined in AliView, guided by available three-dimensional structures. A maximum-likelihood phylogenetic tree was constructed using MEGA6 v6.06 (Tamura et al., 2013) and reliability of the nodes was tested by bootstrap analysis using 100 resampling. Five cellulases from Glycoside Hydrolase Family 7 were used as an outgroup to root the tree.

All Glycoside Hydrolase Family 3 sequences listed as “characterized” as well as those that are structurally characterized with EC number 3.2.1.21 (β-glucosidases) were similarly extracted from the CAZy database. The roughly 300 characterized GH3 sequences were aligned by MUSCLE and a maximum-likelihood phylogenetic tree constructed using MEGA v6.06. The structurally characterized β-glucosidase sequences were combined with BoGH3$_{MLG}$ and BACOVA_02738 and aligned by T-Coffee Expresso.

Survey of metagenomic data sets

Human metagenomic sequence data sets (Huttenhower et al., 2012; Junjie Qin, Yingrui Li, Zhiming Cai, Shenghui Li, Jianfeng Zhu, Fan Zhang, Suisha Liang, Wenwei Zhang, Yuanlin Guan, Dongqian Shen, Yangqing Peng, Dongya Zhang, Zhuye Jie, Wenxian Wu, Youwen Qin, Wenbin Xue, Junhua Li, Lingchuan Han, Donghui Lu, Peixian W, 2012; Qin et al., 2010; Urokawa et al., 2007) were searched by BLAST for the presence of MLGUL nucleotide sequences from *B. ovatus* (13.4kb), *B. uniformis* (14.4kb), *B. cellulosilyticus* (14.1kb), *B. finegoldii* (16.2kb), and *Pr. copri* (13.9kb). Each BLAST probe was first searched against the NCBI Refseq genomes database to determine the background thresholds for BLAST hits and subsequently trimmed to remove any sequences that may return off-target hits. *B. ovatus* and *B. xylanisolvens* MLGULs could not be distinguished due to their very high nucleotide identity (97%). Otherwise, this analysis failed to reveal any off-target hits with length >75 bp, nucleotide identity >90%, and E value $<1^{-20}$. Thus, we considered a metagenome to be positive for a particular MLGUL probe if it returned two or more hits >100 bp in length with >90% identity and E value $<1^{-20}$, or one hit >1000 bp in length with the same identity and E value cut-offs.

Supplemental References

Abdel-Akher, M., and Smith, F. (1951). The Reduction of Sugars with Sodium Borohydride. J. Am. Chem. Soc. *73*, 4691–4692.

Agirre, J., Iglesias-fernández, J., Rovira, C., Davies, G.J., Wilson, K.S., and Cowtan, K.D. (2015). Privateer : software for the conformational validation of carbohydrate structures. Nat. Struct. Mol. Biol. *22*, 833–834.

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., and Notredame, C. (2006). Expresso: Automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucleic Acids Res. *34*, 604–608.

Cota, J., Alvarez, T.M., Citadini, A.P., Santos, C.R., de Oliveira Neto, M., Oliveira, R.R., Pastore, G.M., Ruller, R., Prade, R.A., Murakami, M.T., et al. (2011). Mode of operation and low-resolution structure of a multi-domain and hyperthermophilic endo-β-1,3-glucanase from Thermotoga petrophila. Biochem. Biophys. Res. Commun. *406*, 590–594.

Cowtan, K. (2007). Fitting molecular fragments into electron density. Acta Crystallogr. Sect. D Biol. Crystallogr. *64*, 83–89.

Cowtan, K. (2010). Recent developments in classical density modification. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 470–478.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., et al. (2007). MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. *35*, 375–383.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Emsley, P., and Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. *60*, 2126–2132.

Eschenfeldt, William H., Stols, Lucy, Millard, Cynthia Sanville, Joachimiak, Andrzej, Donnelly, M.I. (2009). A Family of LIC Vectors for High-Throughput Cloning and Purificaiton of Proteins. Methods Mol. Biol. High Throughput Protein Expr. Purif. *498*, 105–115.

Ferrer, P. (2006). Revisiting the Cellulosimicrobium cellulans yeast-lytic β-1,3-glucanases toolbox: a review. Microb. Cell Fact. *5*, 10.

Fuchs, K.P., Zverlov, V. V., Velikodvorskaya, G.A., Lottspeich, F., and Schwarz, W.H. (2003). Lic16A of Clostridium thermocellum, a non-cellulosomal, highly complex endo-β-1,3-glucanase bound to the outer cell surface. Microbiology *149*, 1021–1031.

Furtado, G.P., Ribeiro, L.F., Santos, C.R., Tonoli, C.C., De Souza, A.R., Oliveira, R.R., Murakami, M.T., and Ward, R.J. (2011). Biochemical and structural characterization of a β-1,3-1,4-glucanase from Bacillus subtilis 168. Process Biochem. *46*, 1202–1206.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. Proteomics Protoc. Handb. 571–607.

Genta, F.A., Bragatto, I., Terra, W.R., and Ferreira, C. (2009). Purification, characterization and sequencing of the major β-1,3-glucanase from the midgut of Tenebrio molitor larvae. Insect Biochem. Mol. Biol. *39*, 861–874.

Geoghegan, K.F., Dixon, H.B., Rosner, P.J., Hoth, L.R., Lanzetti, a J., Borzilleri, K. a, Marr, E.S., Pezzullo, L.H., Martin, L.B., LeMotte, P.K., et al. (1999). Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. Anal. Biochem. *267*, 169–184.

Gorlach, J.M., Van Der Knaap, E., and Walton, J.D. (1998). Cloning and targeted disruption of MLG1, a gene en-

coding two of three extracellular mixed-linked glucanases of Cochliobolus carbonum. Appl. Environ. Microbiol. *64*, 385–391.

Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. *31*, 3320–3323.

Hahn, M., Olsen, O., Politz, O., Borriss, R., and Heinemann, U. (1995a). Crystal structure and site-directed mutagenesis of Bacillus macerans endo- 1,3-1,4-β-glucanase. J. Biol. Chem. *270*, 3081–3088.

Hahn, M., Pons, J., Planas, a, Querol, E., and Heinemann, U. (1995b). Crystal structure of Bacillus licheniformis 1,3 -1,4-beta-D-glucan 4-glucanohydrolase at 1.8 A resolution. FEBS Lett. *374*, 221–224.

Hong, T.-Y., and Meng, M. (2003). Biochemical characterization and antifungal activity of an endo-1,3-beta-glucanase of Paenibacillus sp. isolated from garden soil. Appl. Microbiol. Biotechnol. *61*, 472–478.

Hong, T.Y., Hsiao, Y.Y., Meng, M., and Li, T.T. (2008). The 1.5 A˚ structure of endo-1,3-β-glucanase from Streptomyces sioyaensis: Evolution of the active-site structure for 1,3-β-glucan-binding specificity and hydrolysis. Acta Crystallogr. Sect. D Biol. Crystallogr. *64*, 964–970.

Huttenhower, C., Fah Sathirapongsasuti, J., Segata, N., Gevers, D., Earl, A.M., Fitzgerald, M.G., Young, S.K., Zeng, Q., Alm, E.J., Alvarado, L., et al. (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Ibatullin, F.M., Baumann, M.J., Greffe, L., and Brumer, H. (2008). Kinetic Analyses of Retaining endo-(Xylo) glucanases from Plant and Microbial Sources Using New Chromogenic Xylogluco-Oligosaccharide Aryl Glycosides. Society *74*, 7762–7769.

Ilari, A., Fiorillo, A., Angelaccio, S., Florio, R., Chiaraluce, R., Van Der Oost, J., and Consalvi, V. (2009). Crystal structure of a family 16 endoglucanase from the hyperthermophile Pyrococcus furiosus - Structural basis of substrate recognition. FEBS J. *276*, 1048–1058.

Jeng, W.Y., Wang, N.C., Lin, C.T., Shyur, L.F., and Wang, A.H.J. (2011). Crystal structures of the laminarinase catalytic domain from Thermotoga maritima MSB8 in complex with inhibitors: Essential residues for β-1,3- and β-1,4-glucan selection. J. Biol. Chem. *286*, 45030–45040.

Junjie Qin, Yingrui Li, Zhiming Cai, Shenghui Li, Jianfeng Zhu, Fan Zhang, Suisha Liang, Wenwei Zhang, Yuanlin Guan, Dongqian Shen, Yangqing Peng, Dongya Zhang, Zhuye Jie, Wenxian Wu, Youwen Qin, Wenbin Xue, Junhua Li, Lingchuan Han, Donghui Lu, Peixian W, S.Z. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55–60.

Kabsch, W. (2010). Xds. Acta Crystallogr. Sect. D Biol. Crystallogr. *66*, 125–132.

Kawai, R., Igarashi, K., Yoshida, M., Kitaoka, M., and Samejima, M. (2006). Hydrolysis of β-1,3/1,6-glucan by glycoside hydrolase family 16 endo-1,3(4)-β-glucanase from the basidiomycete Phanerochaete chrysosporium. Appl. Microbiol. Biotechnol. *71*, 898–906.

Kitamura, E., Myouga, H., and Kamei, Y. (2002). Polysaccharolytic activities of bacterial enzymes that degrade the cell walls of Pythium porphyrae, a causative fungus of red rot disease in Porphyra yezoensis. Fish. Sci. *68*, 436–445.

Kobayashi, T., Uchimura, K., Kubota, T., Nunoura, T., and Deguchi, S. (2016). Biochemical and genetic characterization of β-1,3 glucanase from a deep subseafloor Laceyella putida. Appl. Microbiol. Biotechnol. *100*, 203 –214.

Kovalchuk, S.N., Sundukova, E. V., Kusaykin, M.I., Guzev, K. V., Anastiuk, S.D., Likhatskaya, G.N., Trifonov, E. V., Nurminski, E.A., Kozhemyako, V.B., Zvyagintseva, T.N., et al. (2006). Purification, cDNA cloning and homology modeling of endo-1,3-β-D- glucanase from scallop Mizuhopecten yessoensis. Comp. Biochem. Physiol. Part B *143*, 473–485.

Krah, M., Misselwitz, R., Politz, O., Thomsen, K.K., Welfle, H., and Borriss, R. (1998). The laminarinase from thermophilic eubacterium Rhodothermus marinus--conformation, stability, and identification of active site carboxylic residues by site-directed mutagenesis. Eur. J. Biochem. *257*, 101–111.

Kumagai, Y., and Ojima, T. (2009). Enzymatic properties and the primary structure of a β-1,3-glucanase from the digestive fluid of the Pacific abalone Haliotis discus hannai. Comp. Biochem. Physiol. Part B *154*, 113–120.

Labourel, A., Jam, M., Jeudy, A., Hehemann, J.H., Czjzek, M., and Michel, G. (2014). The β-glucanase ZgLamA from zobellia galactanivorans evolved a bent active site adapted for efficient degradation of algal laminarin. J. Biol. Chem. *289*, 2027–2042.

Labourel, A., Jam, M., Legentil, L., Sylla, B., Hehemann, J.H., Ferrières, V., Czjzek, M., and Michel, G. (2015). Structural and biochemical characterization of the laminarinase ZgLamCGH16 from Zobellia galactanivorans suggests preferred recognition of branched laminarin. Acta Crystallogr. Sect. D Biol. Crystallogr. *71*, 173–184.

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. Bioinformatics *30*, 3276–3278.

Li, J., Xu, X., Shi, P., Liu, B., Zhang, Y., and Zhang, W. (2015). Overexpression and characterization of a novel endo-β-1,3(4)-glucanase from thermophilic fungus Humicola insolens Y1. Protein Expr. Purif. *138*, 63–68.

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N., et al. (2011). Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. PLoS Biol. *9*, 1–16.

Masuda, S., Endo, K., Koizumi, N., Hayami, T., Fukazawa, T., Yatsunami, R., Fukui, T., and Nakamura, S. (2006). Molecular identification of a novel beta-1,3-glucanase from alkaliphilic Nocardiopsis sp strain F96. Extremophiles *10*, 251–255.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. Sect. D Biol. Crystallogr. *67*, 355–367.

Nakajima, M., Yamashita, T., Takahashi, M., Nakano, Y., and Takeda, T. (2012). A novel glycosylphosphatidylinositol-anchored glycoside hydrolase from Ustilago esculenta functions in β-1,3-glucan degradation. Appl. Environ. Microbiol. *78*, 5682–5689.

Pauchet, Y., Freitak, D., Heidel-Fischer, H.M., Heckel, D.G., and Vogel, H. (2009). Immunity or digestion: Glucanase activity in a glucan-binding protein family from lepidoptera. J. Biol. Chem. *284*, 2214–2224.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing: Commentary. Nature *464*, 59–67.

Stein, N. (2008). CHAINSAW: A program for mutating pdb files used as templates in molecular replacement. J. Appl. Crystallogr. *41*, 641–643.

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

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. *30*, 2725–2729.

Urokawa, K.K., Toh, T.I., Uwahara, T.K., Shima, K.O., Oh, H.T., Oyoda, A.T., Ori, H.M., Gura, Y.O., Hrlich,

D.S.E., Toh, K.I., et al. (2007). Comparative Metagenomics Revealed Commonly Enriched Gene Sets in Human Gut Microbiomes. DNA Res. *14*, 169–181.

Viladot, J.L., Moreau, V., Planas, A., and Driguez, H. (1997). Transglycosylation activity of Bacillus 1,3-1,4-β-D-glucan 4-glucanohydrolases. Enzymic snthesis of alternate 1,3-1,4-β-D-glucooligosaccharides. J. Chem. Soc. Perkin Trans. 1 *338*, 2383–2387.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. Sect. D Biol. Crystallogr. *67*, 235–242.

Zverlov, V. V, Volkov, I.Y., Velikodvorskayal, T. V, and Schwarz, W.H. (1997). Highly thermostable endo-1,3-β-glucanase (laminarinase) LamA from Thermotoga neapolitana: nucleotide sequence of the gene and characterization of the recombinant gene product. Microbiology *143*, 1701–1708.