Molecular Mechanism by which Prominent Human Gut Bacteroidetes Utilize Mixed-Linkage Beta-Glucans, Major Health-Promoting Cereal Polysaccharides

In Brief
Mixed-linkage $\beta(1,3)/\beta(1,4)$-glucan (MLG) is an important complex dietary polysaccharide (dietary fiber), the degradation of which in the human gut depends on the resident microbiota. Tamura et al. outline the molecular mechanism of MLG utilization by Bacteroides ovatus and reveal that the majority of surveyed humans possess MLG-utilizing Bacteroidetes.

Highlights
- The molecular mechanism of MLG utilization by Bacteroides ovatus is presented
- BoGH16$_{MLG}$ possesses structural features suited for efficient MLG degradation
- MLG utilization loci (MLGULs) serve as genetic markers for MLG catabolism
- Bacteroidetes MLGULs are ubiquitous in the gut microbiota of human populations

Data and Software Availability
5NBO
5NBP

Authors
Kazune Tamura, Glyn R. Hemsworth, Guillaume Déjean, ..., Gideon J. Davies, Eric C. Martens, Harry Brumer

Correspondence
brumer@msl.ubc.ca
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 Déjean,1 Theresa E. Rogers,4 Nicholas A. Pudlo,4 Karthik Urs,4 Namrata Jain,1,2 Gideon J. Davies,3 Eric C. Martens,4 and Harry Brumer1,2,5,6,7,*

1Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC V6T 1Z4, Canada
2Department of Biochemistry and Molecular Biology, University of British Columbia, 2350 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada
3York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5DD, UK
4Department of Microbiology and Immunology, University of Michigan Medical School, 1150 West Medical Center Drive, Ann Arbor, MI 48109, USA
5Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, BC V6T 1Z1, Canada
6Department of Botany, University of British Columbia, 3200 University Boulevard, Vancouver, BC V6T 1Z4, Canada
7Lead Contact
*Correspondence: brumer@msl.ubc.ca
https://doi.org/10.1016/j.celrep.2017.09.049

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 have 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 fiber”—that evade degradation by the limited set of human-genome-encoded glycosidase 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 therapeutics based on prebiotics, probiotics, and symbiotics (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 (Baågenholm et al., 2017), and rhamnogalacturonan II (Ndeh et al., 2017). Mixed-linkage β(1,3)/β(1,4)-glucans (MLGs) (Figure 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 USA Food and Drug Administration (FDA) as well as the UK 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

**Figure 1. Cereal MLG and 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 cellotriosyl to cellotetraosyl units (Lazaridou et al., 2004). Arrows indicate the specific site of hydrolysis by the vanguard endo-glucanase of the MLGUL, BoGH16MLG.

(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 as genomic reference points. See also Figure S1 and Table S1.
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 \textit{B. ovatus}.

A putative MLGUL was previously identified in \textit{B. ovatus} (Figure 1B) based on the presence of a tandem \textit{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 upregulated (125- to 298-fold) during growth on bMLG versus 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\textsubscript{MLG} (BACOVA_02741) is a member of a glycoside hydrolase family 16 (GH16) in the carbohydrate active enzymes (CAZy) classification (Cantarel et al., 2009). GH16 notably includes canonical bacterial MLG \textit{endo}-glucanases (\textit{endo}-MLGase) (Planas, 2000), along with a diversity of \textit{endo}-glucanases and \textit{endo}-galactanases (Eklöf and Hehemann, 2016). BoGH3\textsubscript{MLG} (BACOVA_02745) is classified into glycoside hydrolase family 3 (GH3), whose members include exo-β-glucosidases (Fincher et al., 2017). 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 versus glucose control; Table S1); (2) a corresponding gene is not found among syntenic loci (Figure 1B); and (3) the encoded enzyme was catalytically feeble compared to BoGH3\textsubscript{MLG} on β-glucosides relevant to MLG saccharification (vide infra).

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

Enzymology and Structural Biology of BoGH16\textsubscript{MLG}, the Vanguard MLGase.

Cellular Localization.

The GH family membership of BoGH16\textsubscript{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 et al., 2003) suggested that the protein is membrane anchored via lipidation of the N-terminal cysteine residue (Paetzel et al., 2002). To validate this prediction, \textit{B. ovatus \Delta tdk} was grown on minimal medium with either glucose or bMLG as a sole carbon source prior to immunolocalization of BoGH16\textsubscript{MLG}. As shown in Figure 2A, BoGH16\textsubscript{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 (Figures S2C and S2F). Further analysis of cellular fractions by western blotting revealed the presence of BoGH16\textsubscript{MLG} in the membrane fraction, corroborating its attachment to the outer membrane (Figure 2C).

Interestingly, BoGH16\textsubscript{MLG} was also detected in the lysate supernatant (soluble periplasmic or cytoplasmic proteins) and in the culture supernatant (secreted
protein; Figure 2C). Whereas 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 BoGH16MLG in MLG utilization and determine the specificity of the enzyme for individual β-glucans, recombinant BoGH16MLG produced in *E. coli* (recBoGH16MLG; Figures S3A and S3B) was screened for hydrolytic activity against a library of polysaccharides. No activity was observed on tamarind xyloglucan, beechwood xylan, wheat arabinoylan, carob galactomannan, konjac glucomannan, synthetic carboxymethylcellulose, synthetic hydroxyethylcellulose, *Xanthomonas campestris* xanthan gum, or *Ulva* sp. ulvan. In this initial screen, BoGH16MLG 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°C 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 BoGH16MLG 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 (Figure 1A) over laminarin, an all-β(1,3)-glucan with single β(1,6)-linked glucosyl branches (Figure 3A; Table S2). BoGH16MLG 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; Figure 3A; Table S2).

Detailed product analysis was employed to determine the mode of hydrolysis, *endo* versus *exo*, and linkage specificity of recBoGH16MLG to gain information on the nature of the MLG cleavage products at the *B. ovatus* cell surface. High-performance liquid chromatography (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 (Figure 3B). This product evolution indicates that BoGH16MLG operates through an *endo*-dissociative mode of action, in which the MLG polysaccharide is stochastically cleaved along the backbone.

Comparison with oligosaccharide standards (Figure 3B) and additional liquid chromatography-mass spectrometry (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, BoGH16MLG specifically hydrolyzes β(1,4)-linkages of glucosyl 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 BoGH16MLG substrate specificity, Michaelis-Menten kinetics was determined for a series of chromogenic glycosides (Figures S4A and S4B; Table S3). recBoGH16MLG had no activity on the ortho-chloro-para-nitrophenyl (CNP) β-glucosides of glucose (G-CNP), cellobiose (G4-CNPG), cellotriose (G4G4-CNPG), or para-nitrophenyl (pNP) β-glucoside (G-pNP). Weak activity was observed on the pNP and CNP β-glucosides 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, G4G3-CNPG and G4G4G-CNPG were specifically and efficiently hydrolyzed to release the aglycone, with no cleavage of the internal glycosidic bonds. The specificity constants (kcat/Km values) for CNP release from these mixed-linkage tri- and tetrasaccharides were 800- and 1,500- fold greater than that of G3G-CNPG, respectively, which indicate that potential −3 and −4 subsites contribute 17 kJ/mol and 1.6 kJ/mol to transition state stabilization (ΔG∆). 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).

BoGH16MLG Tertiary Structure

Three-dimensional structures of recBoGH16MLG were solved by X-ray crystallography to reveal the molecular basis for MLG recognition and hydrolysis. The apo structure of recBoGH16MLG was determined to a resolution of 1.8 Å by molecular replacement using the structure of Zobellia galactanivorans laminaranase ZgLamC[GH16-E142S] (PDB code 4CRO; 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 BoGH16MLG, consistent with the existence of three primary negative subsites, spanning subsites 1 to 4 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 BoGH16MLG consists of a β-jelly roll architecture typical of other GH16 members (Davies and Sinnott, 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 hints at the possibility of the formation of a dimer (Figure 4A). Size-exclusion chromatography, however, indicated that BoGH16MLG 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 ([18], pointing into the active-site cleft. Notably, these residues are contained in an EXDXXE consensus sequence that is typical of bacterial GH16 laminaranases ([1,3]-specific endo-glucanases). The insertion of an extra amino acid (underlined), typically methionine, results in a so-called “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 Rosenström, 2010) feature a β-bulge active-site motif (Table S5). Specifically, the top match (Z score = 29.3) was the structure of laminaranase “ZgLamC[GH16-E142S]” from Zobellia galactanivorans (PDB code 4CTE; Labourel et al., 2015), which has 38% amino acid identity and superimposed with BoGH16MLG 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 202 out of 212 Cα pairs, despite having only 22% amino acid identity with BoGH16MLG.

Soaking crystals of the wild-type enzyme with G4G4G3G yielded a product complex with 1.8 Å resolution (Table S4). The complete tetrasaccharide was modeled in electron density spanning subsites −1 to −4 in the active-site cleft of chain A, whereas 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 because the corresponding −4 Glc in chain A is fully solvent exposed, makes no contact with the enzyme, and has significantly higher B factors (Figure 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 chains 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 (Figure 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, whereas it is “swung out” in chain A, the sidechain from chain A stacking on top of the chain B sidechain (Figure 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 artifact of crystallization 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

Cell Reports 21, 417–430, October 10, 2017 421
Glucose is also positioned by a stacking interaction with Trp-125 and Trp-129 (Figure 4E), both of which are conserved across all GH16 laminarinases. At subsite C0, 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 C0 and C1 sugars. Another conserved residue, Trp-138, serves as a platform that stacks with the subsite C0 glucose. In subsite C0, the main interaction is stacking against Trp-58, which also forms a hydrogen bond to the glucosidic bond oxygen between the C0 and C1 sugars (Figure 4E). Together, these interactions in subsite C0 are responsible for 17 kJ/mol of transition-state stabilization (vide supra).

**Downstream Saccharification of Mixed-Linkage Oligosaccharides Produced by BoGH16MLG**

To elucidate the mechanism for the downstream conversion of the oligosaccharide products of BoGH16MLG to glucose for primary metabolism, we examined the biochemistry of the two predicted exo-β-glucosidases, BoGH3MLG and BACOVA_02738(GH3), associated with the MLGUL.

**Cellular Localization of BoGH3MLG and the BACOVA_02738(GH3) Gene Product**

BoGH3MLG and BACOVA_02738(GH3) were unambiguously predicted by SignalP 4.0 (Petersen et al., 2011) to contain a secretion signal peptide, whereas LipoP 1.0 (Juncker et al., 2003) additionally indicated a type II lipoprotein signal sequence (Paetzelt et al., 2002) in BoGH3MLG only. The same *B. ovatus* AdtK cultures used for BoGH16MLG localization, containing glucose or bMLG as the sole carbon source, were probed using polyclonal antibodies independently raised against recombinant BoGH3MLG 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 BoGH16MLG production (Figures 2B and S2A). BoGH3MLG induction by bMLG was confirmed by a western blot of cellular fractions, which also confirmed that this enzyme is membrane anchored (Figure 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 (Figure S2B). The lack of upregulation by bMLG is consistent with transcriptional analysis, which showed a limited change in expression in bMLG versus 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...
Table 1. Kinetic Parameters for the Hydrolysis of Various Substrates by BoGH3MLG and BACOVA_02738(GH3)

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}/K_m$ (s$^{-1}$ mM$^{-1}$)</th>
<th>Assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoGH3MLG</td>
<td>$\beta$-Glc-pNP</td>
<td>59.5 ± 1.46</td>
<td>2.95 ± 0.14</td>
<td>20.2</td>
<td>pNP</td>
</tr>
<tr>
<td></td>
<td>gentiobiose (G6G)</td>
<td>ND</td>
<td>ND</td>
<td>0.0571</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>cellobiose</td>
<td>5.52 ± 0.19</td>
<td>7.47 ± 0.48</td>
<td>0.739</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>cellotriose</td>
<td>22.1 ± 0.3</td>
<td>0.859 ± 0.033</td>
<td>25.7</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>celtetraose</td>
<td>17.3 ± 0.5</td>
<td>0.687 ± 0.044</td>
<td>25.2</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>cellopentaose</td>
<td>19.4 ± 0.8</td>
<td>0.777 ± 0.060</td>
<td>25.0</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>cellohexaose</td>
<td>17.4 ± 0.4</td>
<td>0.747 ± 0.041</td>
<td>23.3</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>laminaribiogene$^a$</td>
<td>28.0 ± 1.1</td>
<td>1.90 ± 0.12</td>
<td>14.7</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>laminarinriose</td>
<td>34.2 ± 1.0</td>
<td>0.911 ± 0.052</td>
<td>37.5</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>laminarintriase</td>
<td>31.3 ± 2.3</td>
<td>0.896 ± 0.135</td>
<td>34.9</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>laminarinpentaoase</td>
<td>39.5 ± 3.4</td>
<td>1.27 ± 0.20</td>
<td>31.1</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>MLGO$_3$ A (G3G4G)</td>
<td>61.6 ± 1.6</td>
<td>0.997 ± 0.040</td>
<td>61.8</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>MLGO$_3$ B (G4G3G)$^a$</td>
<td>24.7 ± 1.3</td>
<td>0.521 ± 0.064</td>
<td>47.4</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>MLGO$_4$ A (G3G4G4G)</td>
<td>55.7 ± 2.7</td>
<td>1.33 ± 0.12</td>
<td>41.9</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>MLGO$_4$ B (G4G3G4G)$^a$</td>
<td>30.8 ± 2.0</td>
<td>0.736 ± 0.106</td>
<td>41.8</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td></td>
<td>MLGO$_4$ C (G4G3G4G)$^a$</td>
<td>15.7 ± 0.3</td>
<td>0.601 ± 0.031</td>
<td>26.1</td>
<td>HK/G6PDH</td>
</tr>
<tr>
<td>BACOVA_02738(GH3)</td>
<td>$\beta$-Glc-pNP</td>
<td>0.212 ± 0.004</td>
<td>2.53 ± 0.13</td>
<td>0.0838</td>
<td>pNP</td>
</tr>
</tbody>
</table>

ND, not determined (in cases where Michaelis-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 ± SD. See also Figure S4.

$^a$Biologically relevant substrates that BoGH3MLG encounters in the periplasmic space.

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 (Figure S2B).

Substrate Product Specificity of BoGH3MLG and BACOVA_02738(GH3)

Initial activity screening on chromogenic pNP glycosides (see Experimental Procedures) revealed that both recBoGH3MLG and recBACOVA_02738(GH3) are specific exo-$\beta$-glucosidases (activity on other pNP glycosides was undetectable at micro-molar enzyme concentrations). However, recBACOVA_02738(GH3) is strikingly feeble compared to recBoGH3MLG on G-$\beta$-pNP ($k_{cat}/K_m$ values of 0.084 mM$^{-1}$ s$^{-1}$ versus 20 mM$^{-1}$ s$^{-1}$; Figures S4C and S4D; Table 1). Further, measuring Michaelis-Menten kinetic parameters on cello- and laminari-oligosaccharides was not feasible due to overall poor activity and low protein 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 BoGH3MLG, 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-$\beta$(1,4)-linked cello-oligosaccharides (d.p. 2–6), all-$\beta$(1,3)-linked laminari-oligosaccharides (d.p. 2–5), and mixed-linkage $\beta$(1,3)/$\beta$(1,4)-gluco-oligosaccharides (d.p. 3 to 4; 5 examples) with comparable efficiencies, according to classic Michaelis-Menten saturation kinetics (Figures S4E and S4F; Table 1). In this series, only cellobiose (G4G) was poorly hydrolyzed by BoGH3MLG vis-à-vis the $\beta$(1,3)-linked congener laminaribiogene (G3G) and all other gluco-oligosaccharides (e.g., G4G has a $k_{cat}/K_m$ value 20-fold lower than G3G; Table 1). The $\beta$(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 $\beta$(1,3)-linked glucosyl unit at the non-reducing end all have slightly higher $k_{cat}$ Values compared to those with a $\beta$(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. However, the magnitude of these differences, which are often less than 2-fold, indicate that BoGH3MLG is essentially agnostic to $\beta$(1,3) versus $\beta$(1,4) linkages. These results also suggest that, in addition to a single negative subsite (−1) commensurate with its exo activity, BoGH3MLG has only two positive subsites that contribute to catalysis: in each gluco-oligosaccharide series, tetrasaccharides and larger are hydrolyzed with identical $k_{cat}/K_m$ values to the corresponding trisaccharides.

Product analysis following extended incubation of BoGH3MLG with G4G4G3G and G4G3G demonstrated that BoGH3MLG completely degrades the BoGH16MLG limit digest products to glucose. HPLC also revealed that laminaribiogene (G3G) is the only new intermediate formed during the course of hydrolysis (Figure 3B). This demonstrates that BoGH3MLG sequentially hydrolyzes one glucose unit at a time from the non-reducing end of MLG oligosaccharides, viz.: G4G4G3G → G + G4G3G (also present in the starting mixture) → G + G3G → G + G. Notably, the individual $k_{cat}$ and $K_m$ values for each step are nearly identical (Table 1).
**BoGH3<sub>MLG</sub> and BACOVA_02738(GH3) Primary Structures**

Despite extensive efforts, we were unable to crystallize the key β-glucosidase BoGH3<sub>MLG</sub> for experimental tertiary structure determination. However, BoGH3<sub>MLG</sub> has 63% sequence identity to a B. ovatus β-glucosidase (BoGH3B; PDB code 5JP0; Hems- worth et al., 2018) from the xyloglucan utilization locus (Figure S5A) and, as such, was amenable to tertiary structure homology modeling. Phyre2 (Kelley et al., 2015) utilized PDB code 5JP0 as the sole template, and 728 out of 742 residues (98% of the sequence, excluding the signal peptide) were modeled with 100% confidence. The model suggests that BoGH3<sub>MLG</sub> possesses a homologous three-domain architecture with the active site being formed at the interface of the (α/β)₈ triosephosphate isomerase (TIM) barrel and α/β sandwich domains (Figure S5B). The predicted catalytic nucleophile and acid/base residues were modeled on opposite sides of the entrance to the active site pocket (Figure S5D), forming a narrow “coin slot,” which may confer a preference toward β(1,3)- and β(1,4)-linked glucans and poor activity against β(1,6)-linked gentiobiose (Table 1). In contrast, 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<sub>MLG</sub> and similar GH3 members, despite having only 31% sequence identity to BoGH3<sub>MLG</sub> (Figure S5A). The most similar characterized GH3 member to BACOVA_02738(GH3) among ~300 members identified in the CAZy Database is a Chrysosporium lucknowense β-glucosidase with 39% sequence identity (Dotenko et al., 2012).

**Syntenic MLGULs 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; Rosgowski 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 (Figure 5). In particular, 33 of 33 B. ovatus strains (including the type strain ATCC 48438) grew well on bMLG, as did 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 penetration of the MLGUL across the genus clearly demonstrates nutrient-niche specialization among individual Bacteroidetes species.

Comparative analysis of available genomic sequences revealed that strains able to grow on bMLG as the sole carbon source harbor a syntenic MLGUL (Figure 1B). Previous transcriptional analysis demonstrated that the syntenic MLGUL found in B. cellobiosilyticus is also activated during growth on bMLG (McNulty et al., 2013). Concordance between the presence of a syntenic MLGUL and the ability to utilize MLG is further highlighted by the lack of a MLGUL in the B. uniformis ATCC 8492, one of only two strains of B. uniformis that could not grow on bMLG. Based on this analysis, we can also predict MLG utilization ability in two sequenced species of Prevotella, *Pr. copri* DSM 18205 and *Pr. multiformis* DSM16608, important members of the Bacteroidetes from the human gut and oral cavity, respectively (Figure 1B).

**DISCUSSION**

A Molecular Model for MLG Utilization by B. ovatus

Our current suite of data suggests a model by which the MLGUL gene products work in concert to enable the utilization of MLG (Figure 6), analogous to that of other PUL-encoded systems (Grondin et al., 2017). Thus, BoGH16<sub>MLG</sub> is anchored to the outer membrane, where it plays a leading role in fragmenting large MLG polysaccharide chains (typical d.p. 700–5,000, depending upon the plant species of origin; Lazaridou et al., 2004; Zheng et al., 2011) into oligosaccharides that can be imported into the periplasm via the TBDT. Notably, the specific limit digest products of BoGH16<sub>MLG</sub> endo-hydrolysis identified here, viz. the trisaccharide G4G3G and the tetrasaccharide G4G4G3G (Figure 3B), have been shown previously to bind the periplasmic sensor domain of the HTCS encoded by BACOVA_02740 (K<sub>0</sub> 300 µM and 400 µM, respectively), whereas the intact MLG polysaccharide does not (Martens et al., 2011). Monomeric glucose, all-β(1,4)-linked cello-oligosaccharides, and all-β(1,3)-linked laminari-oligosaccharides are also not bound by the HTCS (Martens et al., 2011), indicating that the unique linkages present in MLG are central to inducing the MLGUL system. Thus, there is an essential yet distant interplay between the outer-membrane-localized MLGase and the HTCS in specific nutrient sensing.

It is therefore likely that the BoGH16<sub>MLG</sub> limit digest products, or minimal repeats of these structures ((G4G4G3G)<sub>m</sub>(G4G3G)<sub>n</sub>), comprise the main products transported through the TBDT in vivo. Recent studies on inulin (β(2,1)-fructan) utilization suggest that some TBDTs are able to import longer polysaccharide chains (e.g., d.p. > 20; Rakoff-Nahoum et al., 2016). Regardless of length, the efficient exo-hydrolytic activity of BoGH3<sub>MLG</sub> in the periplasm is sufficient to completely saccharify all imported oligosaccharides to glucose (Figure 3B), to feed primary metabolism in the cytosol. In this process, the trisaccharide substrate of the HTCS, G4G3G, will always be generated, regardless of the imported saccharide chain length, ensuring continual production of the MLGUL upregulation signal until substrate is exhausted. Interestingly, BoGH3<sub>MLG</sub> will never encounter cellobiose (G4G), toward which it has relatively weak activity (Figure S4F; Table 1), in this process; the final step of saccharification of MLGOs is the hydrolysis of the competent substrate laminaribiose (G3G).
Structural Enzymology Reveals Complex Trajectories for the Evolution of MLG Activity in GH16

Previous phylogenetic analyses of GH16 have suggested that the evolution of the active-site β-bulge motif EXDXXE, which is widespread among clan GH-B (comprising GH16 and GH7), to a regular β strand motif EXDXE is a defining feature that delimits endo-β(1,3)-glucanases (laminarinases; EC 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), respectively (Barbeyron et al., 1998; Michel et al., 2001). In this context, the observation that BoGH16MLG is highly specific for MLG, despite having a β-bulge motif, was surprising.

Using the CAZy Database as a starting point (http://www.cazy.org/GH16_characterized.html), together with mining of the primary literature, we generated a contemporary maximum-likelihood phylogeny of all biochemically characterized GH16 members (Figure S6). This analysis indicates that, although canonical normal β strand MLGases do form a monophyletic group as previously observed, MLGase activity is in fact widespread among the historical “laminarinase” group, in which BoGH16MLG is itself positioned. Despite currently limited and disparate kinetic data for individual enzymes, it also appears that it is not possible to define further substrate-specific clades within this group based on 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 (Figures S7A and S7B). Structural superposition with the BoGH16MLG: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 ZgLamCGH16-E142S:thio-β-1,3-trisaccharide structure (Figure S7A; PDB code 4CTE; Labourel et al., 2015). Indeed, Ilari et al. (2009) observed that shortening the homologous loop in...
LamA from the archon Pyrococcus furiosus (Figure S7A; PDB code 2VY0) by 4 amino acids increased the activity toward MLG by 10-fold. Likewise, BglF from Nocardiopsis sp. F96 (Figure S7B; PDB code 2HYK) and LamR from Rhodothermus marinus (Figure 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 BoGH16MLG, in this position. The canonical, regular β strand MLGase from Paenibacillus macerans (Figure S7C; PDB code 1MAC) and Bacillus licheniformis (Figure 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 (Figure 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.

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 Figure 1. The cell-surface-localized endo-MLGase BoGH16MLG 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 BoGH3MLG and BACOVA_02738(GH3) act from the non-reducing ends to liberate individual glucose monomers, which are imported into the cell and metabolized.

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 (Figure 7). The *Pr. copri* MLGUL is more often the sole MLGUL of an individual than that of *B. cellulosilyticus* when only one is present (Figure 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 are diseased. MLGULs are ubiquitously detectable despite variability in sampling depth across different metagenomics sequencing projects (Figure 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 (Kurokawa 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.

**Conclusions**

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 BoGH16MLG, BoGH3MLG, 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 BoGH16MLG were screened and optimized by sitting drop vapor diffusion method. The structures of the apo- and G4G4G3G-BoGH16MLG were solved by molecular replacement. Details of crystallization, data collection, and structure solution are provided in the Supplemental Experimental Procedures.

Bioinformatics
Phylogenetic analysis 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 ± SD throughout the article. All growth experiment results are averages of two biological replicates.

DATA AND SOFTWARE AVAILABILITY
The accession numbers for the atomic coordinates and structure factors of apo- and G4G4G3G-complexed BoGH16MLG reported in this paper are PDB: 5NBO and PDB: 5NBP, respectively.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures, seven figures, and five tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2017.09.049.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS
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 (Project #30863) and the British Columbia Knowledge Development Fund. Work at the University of Michigan was supported by the U.S. NIH (GM124136). G.J.D. is supported by the Royal Society “Ken Murray” Research Professorship. We thank Diamond Light Source (Harwell, UK) for access to beamlines I02 and I03 (proposal mx9948) that contributed to the results presented here. We thank Constance M. Bahr (Koropatkin group; U. Michigan) for invaluable assistance with microscopy. We thank Adriana Cabrera (Brumer group) for preparing laminarin 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.

Received: June 2, 2017
Revised: August 30, 2017
Accepted: September 14, 2017
Published: October 10, 2017; corrected online November 7, 2017

REFERENCES


Supplemental Information

Molecular Mechanism by which Prominent Human Gut Bacteroidetes Utilize Mixed-Linkage Beta-Glucans, Major Health-Promoting Cereal Polysaccharides

Kazune Tamura, Glyn R. Hemsworth, Guillaume Déjean, Theresa E. Rogers, Nicholas A. Pudlo, Karthik Urs, Namrata Jain, Gideon J. Davies, Eric C. Martens, and Harry Brumer
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 microscopy 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. (C, D and E) 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). (F, G and H) 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 recBoGH16MLG, recBoGH3MLG, and recBACOVA_02738(GH3). (B) Reconstructed mass spectrum of intact recBoGH16MLG. 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 recBoGH3MLG. (D) Reconstructed mass spectrum of intact recBACOVA_02738(GH3).
Figure S4, related to Figure 3. Initial-rate kinetic analysis of MLGUL GHs.

(A) BoGH16<sub>MLG</sub> against G3G-β-CNP and G3G-β-pNP fitted to a linear equation. (B) BoGH16<sub>MLG</sub> against G4G3G-β-CNP and G4G4G3G-β-CNP fitted to the Michealis-Menten equation. (C) BoGH3<sub>MLG</sub> against glucose-β-pNP fitted to the Michealis-Menten equation. (D) BACOVA_02738(GH3) against glucose-β-pNP fitted to the Michealis-Menten equation. (E) BoGH3<sub>MLG</sub> against oligosaccharides with a β(1,3) bond at the non-reducing end. (F) BoGH3<sub>MLG</sub> 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.
Figure S5, related to Figure 3. Primary and tertiary structure analysis of BoGH3_{MLG}.

(A) Structure-based sequence alignment 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 and blue outlines indicate similar positions. Alignment illustration created with ESPript (Gouet et al., 2003). (B) Homology model of BoGH3_{MLG} generated by Phyre2. The (α/β)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.
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 $\beta$-strand from those with a bulged $\beta$-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)
Figure S7, related to Figure 4. BoGH16$_{\text{MLG}}$ loop comparison with laminarinases and MLGases.

BoGH16$_{\text{MLG}}$ (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$_{\text{MLG}}$ loop is colored blue and the loop of the compared structure is colored orange. The BoGH16$_{\text{MLG}}$ catalytic residues and mixed-linkage oligosaccharide G4G4G3G in complex with BoGH16$_{\text{MLG}}$ are also shown in cyan. The thio-β-1,3-trisaccharide in complex with ZgLamC$_{\text{GH16-E142S}}$ (PDB code 4CTE) is shown in green. (A) β-bulge-containing laminarinases. (B) β-bulge-containing MLGases. (C) regular β-stranded MLGases (canonical lichenases).
Table S1, related to Figure 1. Transcriptional expression of MLGUL and neighboring genes.a

<table>
<thead>
<tr>
<th>Locus tag</th>
<th>Putative protein ID</th>
<th>bMLG</th>
<th>MM-Glc</th>
<th>bMLG/MM-Glc</th>
</tr>
</thead>
<tbody>
<tr>
<td>BACOVA_02736</td>
<td></td>
<td>879.4 ± 29.4</td>
<td>511.6 ± 94.3</td>
<td>1.7</td>
</tr>
<tr>
<td>BACOVA_02737</td>
<td></td>
<td>1072.3 ± 23.0</td>
<td>590.4 ± 147.6</td>
<td>1.8</td>
</tr>
<tr>
<td>BACOVA_02738</td>
<td>GH3</td>
<td>355.7 ± 46.9</td>
<td>229.0 ± 55.5</td>
<td>1.6</td>
</tr>
<tr>
<td>BACOVA_02739</td>
<td>Sigma 70, region 4</td>
<td>22.7 ± 0.2</td>
<td>37.2 ± 8.9</td>
<td>0.6</td>
</tr>
<tr>
<td>BACOVA_02740</td>
<td>HTCS</td>
<td>376.8 ± 4.4</td>
<td>197.6 ± 0.1</td>
<td>1.9</td>
</tr>
<tr>
<td>BACOVA_02741</td>
<td>GH16_MLG</td>
<td>3657.6 ± 77.7</td>
<td>16.0 ± 16.5</td>
<td>228.8</td>
</tr>
<tr>
<td>BACOVA_02742</td>
<td>TBDT</td>
<td>3401.1 ± 375.5</td>
<td>13.1 ± 5.7</td>
<td>259.7</td>
</tr>
<tr>
<td>BACOVA_02743</td>
<td>SGBP-A</td>
<td>1837.5 ± 111.1</td>
<td>14.7 ± 6.3</td>
<td>124.9</td>
</tr>
<tr>
<td>BACOVA_02744</td>
<td>SGBP-B</td>
<td>2748.9 ± 103.9</td>
<td>9.2 ± 1.0</td>
<td>298.3</td>
</tr>
<tr>
<td>BACOVA_02745</td>
<td>GH3_MLG</td>
<td>2988.3 ± 53.7</td>
<td>16.5 ± 0.4</td>
<td>180.7</td>
</tr>
<tr>
<td>BACOVA_02746b</td>
<td>Transposase</td>
<td>2763.9 ± 111.5</td>
<td>11.9 ± 9.7</td>
<td>232.8</td>
</tr>
<tr>
<td>BACOVA_02747b</td>
<td>Helicase</td>
<td>2821.0 ± 91.9</td>
<td>16.8 ± 5.8</td>
<td>167.6</td>
</tr>
<tr>
<td>BACOVA_02748</td>
<td></td>
<td>54.5 ± 10.5</td>
<td>25.8 ± 9.1</td>
<td>2.1</td>
</tr>
<tr>
<td>BACOVA_02749</td>
<td></td>
<td>62.3 ± 6.0</td>
<td>30.1 ± 6.8</td>
<td>2.1</td>
</tr>
</tbody>
</table>

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\textsubscript{MLG} on polysaccharide substrates.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$k_{\text{cat}}$ (s\textsuperscript{-1})</th>
<th>$K_{\text{m}}$ (mg mL\textsuperscript{-1})</th>
<th>$k_{\text{cat}}/K_{\text{m}}$ (s\textsuperscript{-1} mg\textsuperscript{-1} mL)</th>
<th>Assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>barley MLG</td>
<td>85.8 ± 4.6</td>
<td>0.364 ± 0.051</td>
<td>238</td>
<td>BCA</td>
</tr>
<tr>
<td>laminarin</td>
<td>18.1 ± 5.25</td>
<td>2.54 ± 0.97</td>
<td>7.12</td>
<td>BCA</td>
</tr>
<tr>
<td>yeast β-glucan</td>
<td>0.875 ± 0.101</td>
<td>0.541 ± 0.169</td>
<td>1.61</td>
<td>BCA</td>
</tr>
<tr>
<td>curdlan</td>
<td>ND</td>
<td>ND</td>
<td>0.042</td>
<td>BCA</td>
</tr>
</tbody>
</table>

Data is only presented for substrates on which BoGH16\textsubscript{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 $k_{\text{cat}}$ and $K_{\text{m}}$ values are not reported and $k_{\text{cat}}/K_{\text{m}}$ value was determined from linear curve fit to initial rate data in the $[S] << K_{\text{m(apparent)}}$ range). Data are represented as fitted parameters ± standard deviation.
Table S3, related to Figure 3. Kinetic parameters of BoGH16<sub>MLG</sub> on chromogenic substrates.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>$k_{cat}$ (s&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}/K_m$ (s&lt;sup&gt;-1&lt;/sup&gt; mM&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>Assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-&lt;i&gt;pNP&lt;/i&gt;</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>pNP</td>
</tr>
<tr>
<td>G-&lt;i&gt;CNP&lt;/i&gt;</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>CNP</td>
</tr>
<tr>
<td>G4G-&lt;i&gt;CNP&lt;/i&gt;</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>CNP</td>
</tr>
<tr>
<td>G4G4G-&lt;i&gt;CNP&lt;/i&gt;</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>CNP</td>
</tr>
<tr>
<td>G3G-&lt;i&gt;pNP&lt;/i&gt;</td>
<td>ND</td>
<td>ND</td>
<td>0.0298</td>
<td>pNP</td>
</tr>
<tr>
<td>G3G-&lt;i&gt;CNP&lt;/i&gt;</td>
<td>ND</td>
<td>ND</td>
<td>0.768</td>
<td>CNP</td>
</tr>
<tr>
<td>G4G3G-&lt;i&gt;CNP&lt;/i&gt;</td>
<td>82.3 ± 3.8</td>
<td>0.134 ± 0.018</td>
<td>614</td>
<td>CNP</td>
</tr>
<tr>
<td>G4G4G3G-&lt;i&gt;CNP&lt;/i&gt;</td>
<td>103.7 ± 3.4</td>
<td>0.0895 ± 0.0097</td>
<td>1160</td>
<td>CNP</td>
</tr>
</tbody>
</table>

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 fitted parameters ± standard deviation.
Table S4, related to Figure 4. Data collection and refinement statistics for BoGH16\textsubscript{MLG} structures.

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

\(^a\)Values in parentheses represent data in the highest resolution shell.
Table S5, related to Figure 4. Dali search results.

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>Organism</th>
<th>Protein name</th>
<th>Z score</th>
<th>RMSD (Å)</th>
<th>% ID</th>
<th>Active site</th>
<th>Predominant activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>4CTE</td>
<td><em>Zobellia galactanivorans</em></td>
<td>ZgLamC</td>
<td>29.3</td>
<td>2.0</td>
<td>38</td>
<td>β-bulge</td>
<td>laminarinase</td>
</tr>
<tr>
<td>2HYK</td>
<td><em>Nocardiopsis sp. F96</em></td>
<td>BglF</td>
<td>28.8</td>
<td>1.9</td>
<td>32</td>
<td>β-bulge</td>
<td>MLGase</td>
</tr>
<tr>
<td>3ATG</td>
<td><em>Cellulosimicrobium cellulans</em></td>
<td>BglII</td>
<td>28.3</td>
<td>1.9</td>
<td>34</td>
<td>β-bulge</td>
<td>laminarinase</td>
</tr>
<tr>
<td>3AZY</td>
<td><em>Thermotoga maritima</em></td>
<td>Lam16</td>
<td>27.8</td>
<td>1.9</td>
<td>32</td>
<td>β-bulge</td>
<td>laminarinase</td>
</tr>
<tr>
<td>2VY0</td>
<td><em>Pyrococcus furiosus</em></td>
<td>LamA</td>
<td>27.3</td>
<td>1.8</td>
<td>31</td>
<td>β-bulge</td>
<td>laminarinase</td>
</tr>
<tr>
<td>4DFS</td>
<td><em>Thermotoga petrophila</em></td>
<td>TpLam</td>
<td>27.3</td>
<td>2.1</td>
<td>31</td>
<td>β-bulge</td>
<td>laminarinase</td>
</tr>
<tr>
<td>4BOW</td>
<td><em>Zobellia galactanivorans</em></td>
<td>ZgLamA</td>
<td>26.9</td>
<td>2.0</td>
<td>30</td>
<td>β-bulge</td>
<td>laminarinase</td>
</tr>
<tr>
<td>3ILN</td>
<td><em>Rhodothermus marinus</em></td>
<td>LamR</td>
<td>26.4</td>
<td>2.2</td>
<td>29</td>
<td>β-bulge</td>
<td>MLGase</td>
</tr>
<tr>
<td>1MAC</td>
<td><em>Paenibacillus macerans</em></td>
<td>Bgi</td>
<td>25.1</td>
<td>2.0</td>
<td>22</td>
<td>regular β- strand</td>
<td>MLGase</td>
</tr>
<tr>
<td>3DGT</td>
<td><em>Streptomyces sioyaensis</em></td>
<td>Curd1</td>
<td>25.0</td>
<td>2.4</td>
<td>28</td>
<td>β-bulge</td>
<td>laminarinase</td>
</tr>
<tr>
<td>1GBG</td>
<td><em>Bacillus licheniformis</em></td>
<td>Bgl</td>
<td>25.0</td>
<td>2.1</td>
<td>23</td>
<td>regular β-strand</td>
<td>MLGase</td>
</tr>
<tr>
<td>3O5S</td>
<td><em>Bacillus subtilis</em></td>
<td>BglS</td>
<td>25.0</td>
<td>2.0</td>
<td>24</td>
<td>regular β-strand</td>
<td>MLGase</td>
</tr>
</tbody>
</table>

a (Labourel et al., 2014)
b (Masuda et al., 2006)
c (Ferrer et al., 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, Washington, DC) 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 hours at room temperature, washed with PBS, and blocked for 16 hours at 4 °C in blocking solution (2 % goat serum (Sigma-Aldrich), 0.02% NaN₃, PBS). The cells were then incubated with individual polyclonal antibodies raised against recBoGH16_{MLG}, rec-BoGH3_{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× 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 10000× g 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® TGXTM 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 Sa-line (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 BoGH16MLG, BoGH3MLG, and BA-COVA_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. pMCGS complement-ary sequences were included in the forward and reverse primers of BACOVA_02738 and BACOVA_02745 for subsequent ligation independent cloning into pMCGS53 plasmids as per (Eschenfeldt et al., 2009). All three con-
structs 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.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer sequence (5' → 3')$^a$</th>
<th>Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoGH16$_{MLG}$ _F</td>
<td>GACGACCATATGTCGGATTCTGGACTGAACG</td>
<td>pET28</td>
</tr>
<tr>
<td>BoGH16$_{MLG}$ _R</td>
<td>GACGACCTCGAGCTATAATATTTTCACCCA</td>
<td>pET28</td>
</tr>
<tr>
<td>BoGH3$_{MLG}$ _F</td>
<td>TACTTCCAAATCCAATGCCATGGTTCCACTGCCATTCCTGAA</td>
<td>pMCSG53</td>
</tr>
<tr>
<td>BoGH3$_{MLG}$ _R</td>
<td>TTATCCACTTCCATCTTATTACTTGCATATAATATTCACTGTTGA</td>
<td>pMCSG53</td>
</tr>
<tr>
<td>BACOVA_02738 _F</td>
<td>TACTTCCAAATCCAATGCCATGGACAAAAAAACCTACTGATAACA</td>
<td>pMCSG53</td>
</tr>
<tr>
<td>BACOVA_02738 _R</td>
<td>TTATCCACTTCCATCTTATTATTTGGACCTCAAACCTCCCCCT</td>
<td>pMCSG53</td>
</tr>
</tbody>
</table>

$^a$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 μg/mL kanamycin for BACOVA_02742 or 100 μg/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-thiogalactopyranoside 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 24700× g 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, 500 mM 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 con-
centrating, 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 recBACOVA_02738(GH3), 54890 M⁻¹ cm⁻¹ for recBoGH16MLG, and 108180 M⁻¹ cm⁻¹ for recBoGH3MLG.

Typical yields after purification were 80 mg of recBoGH16MLG from 1 L of lysogeny broth (LB) culture, 70 mg of recBoGH3MLG 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 xylloglucan, 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**
Cellbiose (G4G) was purchased from Acros Organics. Cellobiose (G4G4G), cellotetraose (G4G4G4G4), cellopentaose (G4G4G4G4G4G), cellohexaose (G4G4G4G4G4G4G), laminaribiose (G3G), lamarintriose (G3G3G), laminaritetraose (G3G3G3G3G3G), mixed-linkage glucotriose A (G3G4G), mixed-linkage glucotriose B (G4G3G), mixed-linkage glucotetraose A (G3G4G4G4G), mixed-linkage glucotetraose B (G4G4G3G), mixed-linkage glucotetraose C (G4G3G4G4G) were purchased from Megazyme. Gentiobiose (G6G) was purchased from Carbosynth (Compton, UK).
**Chromogenic substrates**

*para*-nitrophenyl (pNP) glycosides of β-glucoside (G-β-pNP), 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 pNP β-laminaribioside (G3G-pNP) were purchased from Megazyme. CNP glycosides of cellobioside (G4-CNP) and that of cellotriose (G4G4G-CNP) were purchased from Carboynth. 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 BoGH16MLG) 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 pNP 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 BoGH3MLG, 941 nM for BACOVA_02738(GH3), 9.4 nM for BoGH16MLG against the G4G3G-CNP and G4G4G3G-CNP, 446 nM for BoGH16MLG against G3G-CNP, and 942 nM for BoGH16MLG against G3G-pNP.

Endpoint assays were used for pH and temperature optima of BoGH3MLG and BACOVA_02738(GH3). Enzyme, 1 mM G-pNP, 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 BoGH3MLG 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 Na₂CO₃ to raise the pH and absorbance at A₄₅₅ 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 pNP 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 pNP in sodium citrate pH 6.5, and 12511 M⁻¹ cm⁻¹ for pNP 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 triethyla-
mine pH 7.6 buffer provided in the kit. BoGH3\textsubscript{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\textsuperscript{+}, pre-incubated at 37 °C. The release of glucose monosaccharides corresponds stoichiometrically with the reduction of a molecule of NADP\textsuperscript{+} 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\textsuperscript{-1} cm\textsuperscript{-1} was used to convert to rate of hydrolysis.

**Enzyme limit digest assay**

To determine limit-digestion products of BoGH16\textsubscript{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\textsubscript{MLG} and BACOVA\textsubscript{02738}(GH3), 10 μM enzyme was incubated with the limit digest product of BoGH16\textsubscript{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\textsubscript{MLG} and 12 nM of BoGH3\textsubscript{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\textsubscript{4}OH. 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\textsubscript{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\textsubscript{GH16-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 collected 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 MLA complex were all confirmed as 4C1 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 MLA and BACOVA_02738 and aligned by T-Coffee Expresso.

Survey of metagenomic data sets
Human metagenomic sequence data sets (Huttenhower et al., 2012; Qin et al., 2012; Qin et al., 2010; Kurokawa et al., 2007) were searched by BLAST for the presence of MLGUL nucleotide sequences from B. ovatus (13.4kb), B. uniformis (14.4 kb), B. cellulosilyticus (14.1 kb), B. finegoldii (16.2 kb), and Pr. copri (13.9 kb). 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**


Kumagai, Y., and Ojima, T. (2009). Enzymatic properties and the primary structure of a β-1,3-glucanase from the


