A peer-reviewed version of this preprint was published in PeerJ on 3 August 2016.

View the peer-reviewed version (peerj.com/articles/2289), which is the preferred citable publication unless you specifically need to cite this preprint.

Morrison JM, Elshahed MS, Youssef N. 2016. A multifunctional GH39 glycoside hydrolase from the anaerobic gut fungus *Orpinomyces* sp. strain C1A. PeerJ 4:e2289 https://doi.org/10.7717/peerj.2289

A multifunctional GH39 glycoside hydrolase from the anaerobic gut fungus *Orpinomyces* sp. strain C1A

Jessica M Morrison, Mostafa S Elshahed, Noha Youssef

Background. The anaerobic gut fungi (phylum Neocallimastigomycota) represent a promising source of novel lignocellulolytic enzymes. Here, we report on the cloning, expression, and characterization of a glycoside hydrolase family 39 (GH39) enzyme (Bgxg1) that is highly transcribed by the anaerobic fungus *Orpinomyces* sp. strain C1A under different growth conditions. This represents the first study of a GH39-family enzyme from the anaerobic fungi. **Methods.** Using enzyme activity assays, we performed a biochemical characterization of Bgxg1 on a variety of substrates over a wide range of pH and temperature values to identify the optimal enzyme conditions and the specificity of the enzyme. In addition, substrate competition studies and comparative modeling efforts were completed. **Results.** Contrary to the narrow range of activities (β-xylosidase or α-L-iduronidase) observed in previously characterized GH39 enzymes, Bgxg1 is unique in that it is multifunctional, exhibiting strong β-xylosidase, β-glucosidase, β-galactosidase activities (11.5 ± 1.2, 73.4 ± 7.15, and 54.6 ± 2.26 U/mg, respectively) and a weak xylanase activity (10.8 ± 1.25 U/mg), strength determined as compared to previously characterized enzymes. Physiological characterization revealed that Bgxg1 is active over a wide range of pH (3-8, optimum 6) and temperatures (25-60°C, optimum 39°C), and possesses excellent temperature and thermal stability. Substrate competition assays suggest that all observed activities occur at a single active site. Using comparative modeling and bioinformatics approaches, we putatively identified ten amino acid differences between Bgxg1 and previously biochemically characterized GH39 β-xylosidases that we speculate could impact active site architecture, size, charge, and/or polarity. The putative contributions of these changes to the observed relaxed specificities in Bgxg1 are discussed. **Discussion.** Collectively, the unique capabilities and multi-functionality of Bgxg1 render it an excellent candidate for inclusion in enzyme cocktails mediating cellulose and hemicellulose saccharification from lignocellulosic biomass.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2076v1 | CC-BY 4.0 Open Access | rec: 26 May 2016, publ: 26 May 2016

A multifunctional GH39 glycoside hydrolase from the anaerobic gut fungus *Orpinomyces* sp. strain C1A

Jessica M. Morrison[1], Mostafa S. Elshahed[1], and Noha H. Youssef[1*]

[1]Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK, USA

Author Information:

Jessica M. Morrison, Ph.D.
1110 S. Innovation way, Stillwater, Oklahoma 74074 USA.
jessica.morrison@okstate.edu

Mostafa S. Elshahed, Ph.D.
1110 S. Innovation way, Stillwater, Oklahoma 74074 USA.
mostafa@okstate.edu

***Corresponding author**.
Noha H. Youssef, Ph.D.
1110 S. Innovation way, Stillwater, Oklahoma 74074 USA.
Phone: 1-405-744-1192, Fax. 1-405-744-1112
Noha@Okstate.edu

Keywords: Anaerobic gut fungi, GH39, β-glucosidase, β-xylosidase, β-galactosidase
Short Title: A GH39 Triple oligosaccharide hydrolase

Abstract

Background. The anaerobic gut fungi (phylum Neocallimastigomycota) represent a promising source of novel lignocellulolytic enzymes. Here, we report on the cloning, expression, and characterization of a glycoside hydrolase family 39 (GH39) enzyme (Bgxg1) that is highly transcribed by the anaerobic fungus *Orpinomyces* sp. strain C1A under different growth conditions. This represents the first study of a GH39-family enzyme from the anaerobic fungi.

Methods. Using enzyme activity assays, we performed a biochemical characterization of Bgxg1 on a variety of substrates over a wide range of pH and temperature values to identify the optimal enzyme conditions and the specificity of the enzyme. In addition, substrate competition studies and comparative modeling efforts were completed.

Results. Contrary to the narrow range of activities (β-xylosidase or α-L-iduronidase) observed in previously characterized GH39 enzymes, Bgxg1 is unique in that it is multifunctional, exhibiting strong β-xylosidase, β-glucosidase, β-galactosidase activities (11.5 ± 1.2, 73.4 ± 7.15, and 54.6 ± 2.26 U/mg, respectively) and a weak xylanase activity (10.8 ± 1.25 U/mg), strength determined as compared to previously characterized enzymes. Physiological characterization revealed that Bgxg1 is active over a wide range of pH (3-8, optimum 6) and temperatures (25-60°C, optimum 39°C), and possesses excellent temperature and thermal stability. Substrate competition assays suggest that all observed activities occur at a single active site. Using comparative modeling and bioinformatics approaches, we putatively identified ten amino acid differences between Bgxg1 and previously biochemically characterized GH39 β-xylosidases that we speculate could impact active site architecture, size, charge, and/or polarity. The putative contributions of these changes to the observed relaxed specificities in Bgxg1 are discussed.

62 **Discussion.** Collectively, the unique capabilities and multi-functionality of Bgxg1 render it an
63 excellent candidate for inclusion in enzyme cocktails mediating cellulose and hemicellulose
64 saccharification from lignocellulosic biomass.

65 **Introduction**

66 The production of biofuels from lignocellulosic biomass is a global priority, necessitated by the
67 continuous depletion of recoverable fossil fuel reserves, the deleterious impact of fossil fuels on
68 air quality, as well as their contribution to global climate change (Hill et al. 2006; National
69 Research Council 2011; Ragauskas et al. 2006). Lignocellulosic biomass represents a vastly
70 underutilized and largely untapped source of energy, and its mass utilization for biofuel
71 production is one of the goals enacted by the U.S. Congress-implemented Renewable Fuel
72 Standard (RFS), aiming to generate 16 billion gallons of biofuel from lignocellulosic sources by
73 2022 (National Research Council 2011).

74 The most frequently used method of biofuel production from lignocellulosic biomass is
75 the enzymatic conversion of cellulose and hemicellulose polymers into sugar monomers/
76 oligomers that could subsequently be converted into biofuels using dedicated sugar metabolizers
77 (Elshahed 2010; Hill et al. 2006; Kumar et al. 2008). Historically, enzymatic cocktails designed
78 for the breakdown of lignocellulosic biomass focused primarily on cellulose degradation, due to
79 its relative structural simplicity and uniformity across all types of plant biomass. Nevertheless,
80 the hemicellulose components in lignocellulosic biomass should not be ignored, as hemicellulose
81 represents 20-35% of the composition of lignocellulosic biomass (Liu et al. 2008). Unlike
82 cellulose, plant hemicelluloses are structurally more complex, with multiple types of major
83 hemicelluloses (arabinoxylans/ glucuronoarabinoxylans, glucomannans/galactoglucomannans,
84 mixed glucans, and xyloglucans) present in various plants (Scheller & Ulvskov 2010). The most
85 common type of hemicellulose are the arabinoxylans/ glucuronoarabinoxylans that possess a
86 structural backbone of β-1,4-linked xylose units (Scheller & Ulvskov 2010). Xylan degradation
87 requires the consorted action of the endo-acting-β-1,4-xylanases and the oligosaccharide

88 depolymerizing β-xylosidases, among other enzymes (Elshahed 2010; Scheller & Ulvskov
89 2010).

90 The identification and characterization of novel enzymes and enzyme cocktails with
91 superior lignocellulosic biomass saccharification properties (e.g. high substrate affinity and
92 specific activity, activity retention at a wide range of pH and temperatures, and thermal and pH
93 stability) signify essential thrusts in biofuel research. Members of the anaerobic gut fungi
94 (phylum Neocallimastigomycota) represent a promising, and largely untapped, source of
95 biomass-degrading enzymes (Ljungdahl 2008; Wang et al. 2013). Members of the
96 Neocallimastigomycota are restricted to the herbivorous gut, where they are responsible for the
97 initial colonization and degradation of plant materials ingested by their hosts (Ljungdahl 2008;
98 Wang et al. 2013). The anaerobic gut fungi are excellent biomass degraders, capable of fast,
99 efficient, and simultaneous degradation of the cellulolytic and hemicellulolytic fraction of
100 various plants, including most common lignocellulosic biomass substrates (e.g. Corn Stover,
101 Switchgrass, Sorghum, Energy Cane, and Alfalfa) (Borneman et al. 1989; Harhangi et al. 2003;
102 Liggenstoffer et al. 2014; Youssef et al. 2013). Nevertheless, in contrast to the extensive efforts
103 dedicated to bioprospecting novel cellulases and hemicellulases from aerobic fungi (such as
104 *Aspergillus* (Kumar & Ramon 1996; vanPeij et al. 1997), *Trichoderma* (Matsuo & Yasui 1984)),
105 anaerobic prokaryotes (such as *Clostridium* (Bronnenmeier & Staudenbauer 1988) and
106 *Thermoanaerobacterium* (Shao et al. 2011)) and metagenomic sequence data (Brennan et al.
107 2004; Hess et al. 2011), efforts to identify, express, and characterize such enzymes from
108 anaerobic fungi have been relatively sparse (Borneman et al. 1989; Harhangi et al. 2003).

109 We aim to explore the utility of the anaerobic gut fungus *Orpinomyces* sp. strain C1A
110 (henceforth referred to as C1A) as a novel source of lignocellulolytic enzymes. Our approach

111 depends on implementing a transcriptomics-guided strategy to identify carbohydrate-active
112 enzymes (CAZyme) transcripts that are highly expressed by C1A when grown on lignocellulosic
113 biomass substrates as candidates for cloning, expression, and characterization. Here, we describe
114 our efforts in cloning, expression, and characterization of one such enzyme: a GH39 transcript
115 bioinformatically annotated as a β-xylosidase, designated Bgxg1. This represents the first study
116 of a GH39-family enzyme from anaerobic fungi. Our results document the high affinity, high
117 specific activity, wide pH and temperature ranges, and high thermal and pH stability of this
118 enzyme. More importantly, we demonstrate that this protein possesses novel multiple activities:
119 β-glucosidase, β-galactosidase, and xylanase activities, in addition to the annotated β-xylosidase
120 activity. This is the first report of a CAZyme capable of triple β- xylo-, gluco-, and galactosidase
121 activities within the narrow-substrate-range GH39 family. Indications of structural features in
122 Bgxg1 that may be responsible for this observed novel relaxed substrate specificity are
123 identified, and the ecological significance and evolutionary considerations of this novel multiple
124 specificity are discussed.

125 **Materials and Methods**

126 *Transcriptomics-guided selection of a GH39 enzyme for cloning and characterization.*

127 As a part of an extensive transcriptomic analysis of lignocellulosic biomass degradation by the
128 anaerobic fungal isolate *Orpinomyces* sp. strain C1A (Couger et al. *Accepted*), the most highly
129 transcribed gene annotated as a β-xylosidase was selected for cloning and biochemical
130 characterization. The selected m.21910 transcript (GenBank accession number KT997999) was
131 annotated as member of the GH39 CAZyme family based on the presence of the conserved
132 protein domain pfam01229 (Glyco_hydro_39) family. When strain C1A was grown on different
133 substrates (glucose, Corn Stover, Energy Cane, Switchgrass, and Sorghum), m.21910 constituted

134 58-84% of the transcriptional activity (i.e. normalized FPKM values) of all GH39 transcripts
135 (n=9), and 5.7-18.2% of the transcriptional activities of all C1A genes putatively annotated as β-
136 xylosidases (members of GH39 and GH43, n=41) (Couger et al. *Accepted*). The gene encoding
137 for Bgxg1 protein was previously identified in the genome of strain C1A (GenBank contig
138 accession number ASRE01002650.1, range: 2346-3460). The ctg7180000059688.1 gene consists
139 of 1115 bp and no introns (refer to IMG gene ID 2518718918 for a visual representation of the
140 gene). The protein product is predicted to be extracellular and non-cellulosomal, based on the
141 presence of a signal peptide, and the absence of a CBM fungal dockerin domain, respectively.
142 *Bgxg1 sequence analysis and phylogeny.*
143 To determine the phylogenetic affiliation of Bgxg1 and the overall topology and global
144 phylogeny of GH39 CAZymes, GH39 β-xylosidase sequences available in CAZY database
145 (http://www.cazy.org/GH39_all.html) (n=1145 total GH39 sequences, retrieved October 28,
146 2015, edited to remove α-iduronidases and duplicates, resulting in n=200 β-xylosidases), in
147 addition to Bgxg1, were aligned using Clustal Omega (Sievers et al. 2011). The generated
148 alignment was used to construct a maximum likelihood tree in RAxML (Stamatakis 2014),
149 which was subsequently visualized and annotated using Mega6 (Sievers et al. 2011; Tamura et
150 al. 2013).
151 *Synthesis, cloning, expression, and purification of Bgxg1 protein.*
152 *bgxg1 gene synthesis and cloning.* A fraction (939 bp, positions 67-1035) of m.21910 transcript
153 was codon optimized for ideal expression in *E. coli*, and the entire recombinant pET28a(+)
154 plasmid containing the *bgxg1* insert was synthesized by a commercial provider (GenScript,
155 Piscataway, NJ). The plasmid, pET28a(+)-*bgxg1*, harbors kanamycin resistance (*kan*) and *NdeI*
156 and *XhoI* restriction sites for selection and cloning. The pET28a(+)-*bgxg1* plasmid was first

157 transformed into One-Shot Chemically Competent Top10 *E. coli c*ells (Invitrogen, Carlsbad,
158 CA), and the transformants were grown overnight on LB-kanamycin agar (15 µg/mL) for
159 selection. The purified plasmid was electroporated into a protease-deficient BL21(DE3)pLysS *E.*
160 *coli* strain (Novagen, EMD Millipore, Darmstadt, Germany), possessing an additional
161 chloramphenicol resistance (*cm*) marker, using a single pulse of 1.8 kV in 0.1 cm electrocuvettes.
162 Transformants were grown on LB agar using both kanamycin (15 µg/mL) and chloramphenicol
163 (34 µg/mL) for selection and screened for the presence of correctly sized inserts via colony PCR
164 using T7 forward and reverse primers.
165 *Bgxg1 expression and purification.* Ten milliliters of overnight cultures of BL21(DE3)pLysS *E.*
166 *coli* cells transformed with pET28a(+)-*bgxg1* were used to inoculate 1 L LB broth, containing
167 kanamycin (15 µg/mL) and chloramphenicol (34 µg/mL). The culture was incubated at 37°C
168 with shaking at 200 rpm until an OD_{600} = 0.6 was reached. Isopropyl-β-D-thiogalactopyranoside
169 (IPTG, 1mM final concentration) was then added to induce protein production, and the culture
170 was gently shaken at room temperature overnight. Cells were then pelleted by centrifugation
171 (6,000 x g, 10 minutes, 4°C) and the pellets were collected and stored at -20°C.
172 Preliminary small-scale experiments indicated that the protein is expressed in the
173 inclusion body fraction (data not shown). Inclusion body extraction was initiated by incubating
174 the cultures in B-Per Cell Lysis Reagent (Thermo Scientific, Grand Island, NY) (10 ml per 500
175 ml of culture) for 15 minutes at room temperature with gentle shaking to lyse the cells. The
176 homogenate was centrifuged (10,000 x g, 30 minutes, 4°C) and the inclusion body extraction
177 procedure (Grassick et al. 2004) was conducted on the cell pellet as follows: The pellet was
178 resuspended in a urea-based inclusion body extraction buffer (20% glycerol, 8 M urea, 50 mM
179 sodium monobasic phosphate, 500 mM sodium chloride, pH 8.0) for 30 minutes at room

180 temperature with gentle shaking. The homogenate was centrifuged (10,000 x g, 30 minutes, 4°C)
181 and the resultant supernatant containing target inclusion body proteins was subsequently utilized
182 for refolding and purification procedures.

183 Recombinant protein refolding was achieved using slow dialysis as previously described
184 (Grassick et al. 2004). In brief, inclusion body extract was incubated with EDTA (1 mM final
185 concentration) and β-mercaptoethanol (100 mM final concentration) for 2 hours at room
186 temperature with gentle shaking, transferred to dialysis tubing (NMWL: 12,000 – 14,000 Da),
187 and placed for 3 hours into inclusion body exchange buffer (20% glycerol, 8 M urea, 50 mM
188 sodium monobasic phosphate, 500 mM sodium chloride, 1 mM EDTA, pH 8.0) for removal of
189 the β-mercaptoethanol. The buffer was refreshed and dialyzed for an additional 3 hours. The
190 dialysis tubing was then placed into a low-urea refolding buffer (2 M urea, 50 mM sodium
191 monobasic phosphate, 500 mM sodium chloride, 1 mM EDTA, 3 mM reduced glutathione, 0.9
192 mM oxidized glutathione, pH 8.0) and dialyzed overnight, followed by a no-urea refolding buffer
193 (50 mM sodium monobasic phosphate, 500 mM sodium chloride, 1 mM EDTA, 3 mM reduced
194 glutathione, 0.9 mM oxidized glutathione, pH 8.0) for 36 hours.

195 Following dialysis, the contents of the tubing were centrifuged to remove insoluble,
196 precipitated proteins (15,000 x g, 15 minutes, 4°C). The supernatant, containing refolded soluble
197 protein, was then exposed to a nickel-nitriloacetic acid (Ni-NTA, 1:1 ratio) slurry (UBPBio,
198 Aurora, CO), packed in a glass frit column (25 x 200 mm, 98 mL volume Kimble-Chase Kontes
199 Flex Column, Vineland, NJ), and allowed to incubate at 4°C for 1 hour on an orbital shaker.
200 Protein purification followed as detailed previously (Morrison et al. 2012). Samples were
201 concentrated using Amicon Ultra-15 Centrifugal Filter Units (Millipore, NMWL 30 kDa) and
202 protein concentration was determined using a Qubit Fluorimeter (Thermo Scientific) in reference

203 to standard protein concentrations. Protein refolding was checked as activity against PNPX, as
204 described below. An SDS-PAGE gel was run to check protein size and purity, as previously
205 described (Laemmli 1970; Morrison et al. 2012).

206 *Biochemical characterization of Bgxg1 (Enzyme activity assays).*

207 *pH and temperature optima and stability.* The pH range and optimum for Bgxg1 was determined
208 by assaying its β-xylosidase activity (described below) at pH 3, 4, 5, 6, 7, 8, 9, and 10, using the
209 following buffer systems: sodium acetate buffer (pH 3.0-6.0), sodium phosphate buffer (pH 7.0-
210 8.0), and glycine buffer (pH 9.0-10). Similarly, the temperature range and optimum for Bgxg1
211 was determined by assaying its β-xylosidase activity at 25, 30, 39, 50, and 60°C. The stability of
212 Bgxg1 after exposure to pH extremes was determined by assaying its β-xylosidase activity
213 following a one-hour incubation at pH 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 13 at 4°C. The following
214 pH buffering systems were used for pH adjustment: sodium acetate buffer (pH 3.0-6.0), sodium
215 phosphate buffer (pH 7.0-8.0), glycine buffer (pH 9.0-10), sodium bicarbonate (pH 11.0), and
216 KCl-NaOH (pH 12-13). Similarly, the thermal stability of Bgxg1 was determined by assaying its
217 β-xylosidase activity following a one-hour incubation at 4, 25, 30, 37, 39, 50, 60, and 70ºC. In all
218 cases, 2.2 µg of pure Bgxg1 was used, since this concentration was determined to be optimal in
219 initial testing. All experiments were completed in triplicate, and relative specific activities in
220 relation to the best performing condition (100% activity) were reported.

221 *Enzyme activity assays.* All enzyme assays with Bgxg1 were conducted in pH 6.0 buffer and at
222 39°C, as these conditions were determined to be optimal for Bgxg1. All reagents were purchased
223 from Sigma Aldrich (St. Louis, MO) unless noted otherwise.

224 Endoglucanase, exoglucanase, xylanase, and mannanase activities were determined using
225 a DNS (3,5-dinitrosalicyclic acid)-based assay (Breuil & Saddler 1985), with carboxymethyl

226 cellulose sodium salt (CMC, 1.25% w/v), avicel microcrystalline cellulose (1.25% w/v),
227 beechwood xylan (1.25% w/v), and locust bean gum (0.5% w/v) as substrates, respectively.
228 Cellobiohydrolase, β-xylosidase, arabinosidase, mannosidase, β-glucosidase, β-
229 galactosidase, and acetyl xylan esterase activities were determined using (10mM) of the *p*-
230 nitrophenol-based (PNP) substrates: *p*-nitrophenyl-β-D-cellobioside (PNPC,), *p*-nitrophenyl-β-D-
231 xylopyranoside (PNPX), *p*-nitrophenyl-β-D-arabinofuranoside (PNPA) *p*-nitrophenyl-β-D-
232 mannoside (PNPM), *p*-nitrophenyl-β-D-glucopyranoside (PNPG), *p*-nitrophenyl-β-D-
233 galactopyranoside (PNPGal), and *p*-nitrophenyl-acetate (PNPAc), respectively (Dashtban et al.
234 2010; Kubicek 1982; Zhang et al. 2009). Assays were conducted in sodium acetate buffer with
235 sodium carbonate (1M) as a stop reagent. α-glucuronidase activity was assayed using the
236 Megazyme α-glucuronidase assay kit (Wicklow, Ireland).
237 All experiments were conducted in triplicate. One unit of enzymatic activity (U) was
238 defined as one µmol of products (reducing sugar equivalents in DNS assays, PNP released in
239 PNP substrate-based assays, and aldouronic acid in α-glucuronidase assay) released from the
240 substrate per minute. Specific activity was calculated by determining the units released per mg of
241 enzyme.
242 *Enzyme kinetics.* Standard procedures were used to determine the K_m, V_{max}, and specific activity
243 of Bgxg1 on all substrates described above (Lineweaver & Burk 1934). K_m and V_{max} values were
244 obtained using double-reciprocal Lineweaver-Burke plots were used to extrapolate from
245 experimentally-derived values using a constant protein concentration (2.2 µg) and variable PNP-
246 based substrate concentration (0.1 – 100 mM) (Lineweaver & Burk 1934). Given the extinction
247 coefficient of p-nitrophenol (PNP) is 17/mM/cm at 400 nm (Bessey & Love 1952), for a 1 cm
248 path length cuvette and absorbance minimum of 0.010, reliable K_m detection limits in such PNP-

249 based spectrophotometric assays is ≈ 500 nM. Therefore, K_m values < 500 nM are referred to as
250 BDL (below detection limit).
251 *Substrate competition assays.* Competitive inhibition experiments were conducted to determine
252 whether the observed multiple oligosaccharide hydrolase activities are catalyzed via a single or
253 multiple active sites. In such experiments, the effect of cellobiose (as a competitive inhibitor) on
254 the β-xylosidase activity of Bgxg1 was measured by conducting the β-xylosidase assay, using 10
255 mM of PNPX as the substrate, in the presence of different concentrations of cellobiose (0, 10,
256 and 20 mM) and evaluating the impact of cellobiose presence on the release of PNP. Conversely,
257 the effect of xylobiose (as a competitive inhibitor) on the β-glucosidase activity of Bgxg1 was
258 measured by conducting the β-glucosidase assay (using 10 mM of PNPG as the substrate) in the
259 presence of different concentrations of xylobiose (0, 10, and 20 mM), and evaluating the impact
260 of xylobiose presence on the release of PNP. In both experiments, the effect of inhibitor
261 concentration on K_m and V_{max} was evaluated using Lineweaver-Burke plots (Lineweaver & Burk
262 1934). All experiments were conducted in triplicate.
263 Substrate preferences of Bgxg1 were determined by conducting a substrate competition
264 assay, where Bgxg1 (2.2 μg of pure enzyme preparation) was challenged by a mixture of
265 xylobiose (10 mM) and cellobiose (10 mM). The kinetics of xylose and glucose release were
266 compared to the results obtained in control experiments were only one substrate (xylobiose or
267 cellobiose) was utilized. Samples were taken at 0, 1, 5, 10, 15, 30, and 60 minutes for the
268 determination of the glucose and xylose concentrations. Glucose was assayed using PGO
269 Enzyme Preparation Capsules (Sigma-Aldrich, St. Louis, MO) and xylose was assayed using
270 Megazyme Xylose Kit (Wicklow, Ireland). All experiments were conducted in triplicate.
271 *Bgxg1 modeling.*

272 Homology modeling by Iterative Threading ASSEmbly Refinement (I-TASSER) (Roy et al.
273 2010; Yang et al. 2015; Zhang 2008), was conducted to generate a three-dimensional model of
274 Bgxg1 using *Thermoanaerobacterium saccharolyticum* β-xylosidase (PBD entry 1UHV) as a
275 template. PyMOL was used to align the Bgxg1 structural prediction to that of
276 *Thermoanaerobacterium saccharolyticum* (PBD entry 1UHV) to examine and speculate the
277 impact of variations in amino acids residue on the enzyme's active site topology and putative
278 substrate binding capacities (PyMol).

279 **Results**

280 *Bgxg1 phylogenetic affiliation.*

281 Phylogenetic analysis grouped all GH39 sequences into 4 phylogenetically-resolved and
282 bootstrap-supported clades (Classes I-IV in Fig. 1). *Orpinomyces* sp. strain C1A Bgxg1 protein
283 belonged to Class III, forming a well-supported cluster with GH39 proteins from the anaerobic
284 fungus *Piromyces* sp. strain E2 (the only anaerobic fungal strain with a sequenced genome in
285 addition to strain C1A), as well as GH39 proteins from the bacterial genera *Clostridium* and
286 *Teredinibacter* (70-74% sequence identities) (Fig. 1). To our knowledge, none of the GH39
287 proteins within this specific cluster, or in the entire Class III GH39, has been biochemically
288 characterized.

289 *Physiological characterization.*

290 SDS-PAGE results show that the Bgxg1 protein is consistent with the predicted size of 42.7 kDa
291 (protein predicted molecular weight is 39.6 KDa + 0.996 kDa linker + 2.101 kDa double
292 histidine tag) (Fig. 2).

293 The thermal and pH stabilities of Bgxg1 were examined by conducting activity assays
294 post-stress (pH or thermal)-incubations as described above. Bgxg1 exhibited activity in a wide

295 range of pH (3-8) and temperatures (25-60°C), with optimal activity at pH 6 and 39°C (Fig. 3A,
296 3B). Bgxg1 retained more than 80% of its specific activity post-application of pH stress ranging
297 between 6 and 11 (Fig. 3C), and 60% of its specific activity post application of pH stress of 4, 5,
298 and 12 (Fig. 3C). Further, Bgxg1 retained ≥80% of its specific activity across the broad range of
299 temperature stressors applied (4 – 70°C) (Fig. 3D).

300 *Substrate specificities and kinetics.*

301 To date, all characterized GH39 enzymes exhibit a narrow substrate range (β-xylosidase or α-L-
302 iduronidase) (Table S1). Predictably, Bgxg1 exhibited a strong β-xylosidase activity (11.5 ± 1.2
303 U/mg, Table 1), compared to previously reported β-xylosidase activities (Table S1).
304 Interestingly, in addition to β-xylosidase activity, Bgxg1 also exhibited strong β-glucosidase
305 (73.4 ± 7.15 U/mg), β-galactosidase (54.6 ± 2.26 U/mg), and weak xylanase (10.8 ± 1.25 U/mg)
306 activities (Table 1), as compared to reported activities from previously characterized enzymes
307 (Tables S2-S4). Our extensive literature review identified 63 enzymes that have been
308 biochemically-characterized to have β-glucosidase activity and, of these, only seven have a
309 reported specific activity higher than that of Bgxg1 (Table S2). Similarly, we only identified
310 three β-galactosidase with a reported higher activity than Bgxg1 (Table S3). On the other hand,
311 the xylanase activity of Bgxg1 is relatively weak, with many previously reported xylanases
312 exhibiting a much higher specific activity (Table S4). Bgxg1 exhibited no detectable
313 exoglucanase, endoglucanase, mannanase, arabinosidase, acetyl xylan esterase,
314 cellobiohydrolase, mannosidase, or α-glucuronidase activities.

315 In addition to its high β-xylosidase, β-glucosidase, and β-galactosidase specific activities,
316 Bgxg1 exhibited remarkably high affinities towards all examined substrates, with K_m values in
317 the low nM range for PNPG and PNPGal, the low μM range for PNPX (Table 2, Table S1-S4).

318 *Substrate competition studies.*

319 Substrate competition studies were conducted using a variable concentration of an unlabeled
320 substrate (acting as an inhibitor) and a fixed concentration of a chromophore (PNP-based)
321 substrate (Table 3). The results strongly suggest the occurrence of cross-substrate competitive
322 inhibition between xylobiose and cellobiose (Table 3), since the presence of increasing
323 concentrations of a single substrate lowers the specific activity and increases the K_m of the
324 enzyme towards the other substrate, whilst not affecting its V_{max}. This pattern strongly indicates
325 that a single active site is responsible for the observed activities (Table 3), a conclusion that is in
326 agreement with the lack of identifiable additional domains other than pfam01229 in Bgxg1, as
327 well as with the structural modeling data described below.

328 In single substrate assays, Bgxg1 was capable of converting cellobiose to glucose and
329 xylobiose to xylose at a very fast rate (Fig. 4A, 4B). This reaction occurs more quickly for
330 xylobiose, as a stable maximal xylose concentration is reached after only 1 minute of incubation
331 (Fig. 4B), compared to 15 minutes for glucose release from cellobiose (Fig. 4A). However, the
332 extent of sugar release at the conclusion of the experiment was higher in cellobiose incubations
333 (Fig. 4A) than xylobiose incubations (Fig. 4B). Competition studies using equimolar
334 concentrations of both substrates revealed the preference of Bgxg1 for xylobiose, since a higher
335 proportion of xylose rather than glucose was detected within the first 15 minutes of the
336 incubation (Fig. 4C). Nevertheless, the final concentrations of sugars released after 60 minutes of
337 incubation did not differ when comparing single substrate versus competition experiments (Fig.
338 4A-4C). Similar to the patterns observed in single substrate assays, Bgxg1 reduced a larger
339 amount of cellobiose to glucose than xylobiose to xylose in competition experiments (Fig. 4C),

340 which is consistent with the higher affinity (lower K_m value) of Bgxg1 for PNPG (12.5 nM) over
341 PNPX (4.85 μM) (Table 2).

342 *Structure activity predictions.*

343 The Bgxg1 protein sequence was submitted to I-TASSER for structural prediction by Iterative
344 Threading ASSEmbly Refinement (Roy et al. 2010; Yang et al. 2015; Zhang 2008) utilizing the
345 β-xylosidase originating from *Thermoanaerobacterium saccharolyticum* (PBD entry 1UHV) as a
346 template for model creation (Roy et al. 2010; Yang et al. 2004; Yang et al. 2015; Zhang 2008).
347 Bgxg1 is predicted to have three distinct domains: a catalytic $(\alpha/\beta)_8$ barrel fold domain (position
348 26-307), a small α -helical domain (position 1-25), and a β sandwich domain (position 308-344)
349 (Fig. S1). Overall, the structure is predicted to contain 11 β-sheets (8 in $(\alpha/\beta)_8$-barrel, 3 in β-
350 sandwich), and 10 α-helices (8 in $(\alpha/\beta)_8$-barrel, 2 in α-domain). The catalytic $(\alpha/\beta)_8$-barrel fold
351 domain is predicted to consist of eight parallel β-sheets (β1-β8), and eight parallel α-helices (α1-
352 α8). Consistent with β-xylosidases of *Thermoanaerobacterium saccharolyticum (*1UHV) and
353 *Geobacillus stearothermophilus* (1PX8), the active site pocket of Bgxg1 is predicted to be
354 located on the upper side of the $(\alpha/\beta)_8$-barrel (Czjzek et al. 2005; Yang et al. 2004) (Fig. S1).
355 Alignment and structural predictions identified the conservation of the general acid-base active
356 site residue Glu127 in the C-terminal of β3, as part of the GH39-conserved Asn126-Glu127-
357 Pro128 motif as well as the nucleophilic residue Glu225 in β6 (Fig. 5) (Czjzek et al. 2005; Yang
358 et al. 2004).

359 Using the predicted model we sought to infer structural differences potentially
360 responsible for the observed relaxed substrate specificities in Bgxg1 by investigating the amino
361 acid conservation patterns between Bgxg1 and all structurally and/or biochemically-
362 characterized β-xylosidases. These enzymes are: *Thermoanaerobacterium saccharolyticum* β-

363 xylosidase (Yang et al. 2004), *Geobacillus stearothermophilus* β-xylosidase (Bhalla et al. 2014;
364 Czjzek et al. 2005), and *Bacillus halodurans* C-125 protein BH1068 (Wagschal et al. 2008), all
365 of which belong to Class II (Fig. 1), as well as *Caulobacter crescentus* CcXynB2 (Correa et al.
366 2012), which belongs to Class I (Fig. 1). All of these enzymes have previously been reported to
367 solely possess β-xylosidase activity (Bhalla et al. 2014; Correa et al. 2012; Czjzek et al. 2005;
368 Wagschal et al. 2008; Yang et al. 2004). We focused on 25 amino acids in two groups: (i) those
369 previously shown to be important for β-xylosidase activity (Czjzek et al. 2005; Yang et al. 2004)
370 [this group includes (in addition to the conserved general acid-base and nucleophilic active sites
371 described above) amino acids providing the tight hydrogen bonding necessary to stabilize the
372 xylosyl-enzyme intermediate formed during the reaction, such as Arg52, His54, Asn159, His228,
373 Tyr230, Glu278, Trp315, Glu322, and Glu323 (locations refer to position in 1UHV)], as well as
374 (ii) those physically interacting with the active site as deduced by the predicted Bgxg1 model
375 (Fig. S1A) [this group includes Val46, Val81, Ile124, Trp125, Gly130, Thr131, Trp132, Phe139,
376 Pro162, Cys163, Tyr164, Ser165, Lys171, His192, Asn242, and Lys247 (locations refer to
377 position in Bgxg1)]. Of these 25 amino acids, 15 differed between Bgxg1 and the four other
378 proteins. Five of these 15 amino acids were not conserved amongst any of the five sequences
379 studied and so were not further investigated (Fig. 5). Therefore, 10 distinct differences (8
380 substitutions and 2 deletions) between Bgxg1 on one hand and the four biochemically-
381 characterized β-xylosidases on the other were identified (Table 4). These differences that are
382 predicted to exist in or around the active site of Bgxg1 would putatively impact the size, charge,
383 and/or polarity within the active site (Table 4, Fig. S1).
384 The expanded substrate specificity observed in this study could be a unique trait in
385 Bgxg1, or it could be specific to all GH39 CAZymes of anaerobic fungi (e.g. Class III-C), or to

386 the entire Class III β-xylosidases. Based on the above speculations about the amino acids
387 potentially responsible for Bgxg1 relaxed specificity, we further investigated the conservation of
388 these 10 amino acid changes (Table 4) within class III of GH39 proteins. Bgxg1 (as well as other
389 GH39 proteins encoded in C1A genome), all three GH39 proteins from the *Piromyces* genome
390 (accession numbers shown in Fig. 1), and all additional sequences from Class III-C belonging to
391 the genera *Clostridium* and *Teredinibacter* were found to encode 9 of the 10 observed amino
392 acid substitutions (Table 4). However, within the broader Class III, little similarity in key amino
393 acids was observed between Bgxg1 sequences and β-xylosidases belonging to Class III-A, III-B,
394 or III-D (Table 4). Collectively, these results putatively suggest that the observed relaxed
395 specificity in Bgxg1 could be exclusive to Class III-C β-xylosidases.

396 **Discussion**

397 In this study, we used a transcriptomics-guided approach to identify, clone, express, and
398 characterize a GH39 protein (Bgxg1) from the anaerobic gut fungus *Orpinomyces* sp. strain
399 C1A. Our results demonstrate that the expressed protein is multifunctional, possessing strong β –
400 xylosidase (11.5 U/mg), β –glucosidase (73.4 U/mg), and β -galactosidase (54.6 U/mg) activities,
401 as well as a weak xylanase activity (10.8 U/mg) (Table 1, 2), as compared to previously
402 characterized enzymes (Tables S1-S4). This novel multi-functionality has not been previously
403 encountered in GH39 enzymes, and therefore this work expands on the known activities of
404 GH39 CAZyme family. Further, Bgxg1 retains high levels of activity over a wide range of
405 temperatures (>80% of activity retained between 4-70°C) (Fig. 3D) and pH values (>80% of
406 activity retained between pH 6-11) (Fig. 3C).

407 In addition to its relaxed substrate specificity, the enzyme displays superior kinetic
408 properties (high specific activity and affinity) towards its multiple substrates. As a β-xylosidase,

409 Bgxg1 has one of the highest β-xylosidase specific activity among all reported ambient (<50^{0}C)
410 β-xylosidases and one of the highest specific activities amongst known GH39 β-xylosidases
411 (Tables 1, S1). Compared to other characterized β-glucosidases, Bgxg1 has the highest specific
412 activity for all ambient temperature β-glucosidases, and one of the highest reported specific
413 activities among all β-glucosidase (members of GH1, GH3, GH5, GH9, and GH30 (Cairns &
414 Esen 2010)), regardless of optimal temperature and GH affiliation (Tables 1, S2). Finally,
415 compared to other characterized β-galactosidases, Bgxg1 has the highest specific activity for all
416 ambient temperature β-galactosidases, and one of the highest reported specific activities among
417 all β-galactosidases (members of GH1, GH2, GH35, and GH42 (Skalova et al. 2005)) regardless
418 of optimal temperature and GH affiliation (Tables 1, 2, S3).
419 We reason that the observed kinetics and substrate specificity of Bgxg1 are beneficial for
420 strain C1A and are highly desirable for a saccharolytic enzyme acting within the highly
421 competitive rumen environment, where strain C1A originally existed (*Orpinomyces* sp. strain
422 C1A was isolated from the feces of an angus steer (Youssef et al. 2013)). The high specific
423 activity and high substrate affinity aid in fast and efficient scavenging of sugars from the
424 surrounding environment, where competition for sugars/oligosaccharide produced by
425 saccharolytic enzymes are intense, and where free sugar levels are permanently low (Garcia-
426 Vallve et al. 2000). We hence speculate that the survival in an anaerobic, eutrophic highly
427 competitive environment might be responsible for the acquisition, retention and directed
428 evolution of anaerobic fungal β-xylosidases towards superior kinetics and relaxed specificities.
429 Sequence analysis and structural modeling (Figs. 5 and S1), and substrate competition
430 experiments (Table 3) predict the presence of a single conserved active site within the $(\alpha/\beta)_8$-
431 barrel fold structure typically observed in GH39-family enzymes (Czjzek et al. 2005; Yang et al.

432 2004) (with the conserved catalytic nucleophile (Glu225) and general acid-base residue
433 (Glu127)) and potentially mediating all observed hydrolytic activities). To provide clues
434 regarding the structural basis of the observed multi-functionality, comparison of amino acid
435 conservation patterns putatively affecting the active site topology between Bgxg1 and
436 biochemically characterized GH39 xylosidases, all four of which display no additional activities
437 beyond β-xylosidase, was undertaken. We identified ten different distinct amino acid changes (8
438 substitutions and 2 deletions) (Table 4, Figs. S1 and 5) in Bgxg1 that putatively affect the
439 polarity (Tyr vs. Val46, Phe vs. Thr131, Tyr vs. Phe139, Ala vs. Cys163, Trp vs. Lys171, Tyr vs.
440 Leu194, and Ala vs. Arg242), constitute significant size changes (Tyr vs. Val46, Phe vs. Thr131,
441 Tyr vs. Leu194, and Ala vs. Arg242), result in the addition of charged moieties or unique
442 functional groups (Asn vs. Asp129, Ala vs. Cys163, Trp vs. Lys171, and Ala vs. Arg242), or
443 result in the deletion of a negatively charged residue, previously determined to be important
444 (Glu322-323 vs. deletion) to the active site (Czjzek et al. 2005). The impact of these speculated
445 changes is unclear, and it remains to be seen if any, all, or a combination of the above differences
446 is responsible for the observed relaxed specificity. However, while all these amino acid changes
447 are speculated to theoretically explain the relaxed specificity of Bgxg1, one such difference is
448 peculiar and deserves special scrutiny; deletions/gaps in the Bgxg1 sequence as opposed to
449 negatively charged glutamic acids in the other four sequences (Table 4, Fig. S1S). GH39
450 enzymes belong to the wider family of β-1,4-retaining hydrolases of clan GH-A e.g. GH1 β-
451 glucosidase and GH5 cellulases. Differences in structure between β 1,4-glucose cleaving
452 enzymes and β 1,4-xylose cleaving enzymes within clan GH-A have been extensively
453 investigated (Czjzek et al. 2005; Czjzek et al. 2001; Ducros et al. 1995; Hovel et al. 2003;
454 Verdoucq et al. 2004). Such studies have demonstrated that, within the active site of β 1,4-

455 glucose cleaving enzymes, a Gln residue (corresponding to position 39 in the enzyme dhurinase
456 of *Sorghum bicolor* (Czjzek et al. 2005; Ducros et al. 1995; Verdoucq et al. 2004)) interacts with
457 the substrate by forming a hydrogen bond with O3 and O4 of the glucose moiety (Czjzek et al.
458 2005; Ducros et al. 1995). On the other hand, β 1,4-xylosidases acting on C5 sugar dimers
459 contain a Glu residue in lieu of Gln (at position 322-323 in *Thermoanaerobacterium*
460 *saccharolyticum,* Fig. 5 and S1, Table 4) that binds to O3 and O4 of the xylose moiety (Czjzek et
461 al. 2005). Interestingly, these Glu residues are aligned with a gap in the sequence of the
462 multifunctional Bgxg1 (Fig. 5), with no apparent occurrence of either Glu or Gln amino acids
463 within the vicinity. Structurally predictive modeling suggests that in lieu of these Glu322-323
464 residues (1UHV numbering) Bgxg1 is predicted to possess Gly-Arg at an approximately
465 sterically-similar location near the active site (Figure S1R-S), representing a significant change
466 from two negatively-charged residues, to an uncharged and positively-charged pair of residues.
467 Since the Glu residues in biochemically characterized β-xylosidases are shown to be important
468 for stabilizing intermediates (Czjzek et al. 2005), the predicted absence of these residues in
469 Bgxg1 and their speculated replacement with Gly-Arg suggests that Bgxg1 might employ a
470 different mechanism for stabilizing its intermediates during the catalytic process; however, this
471 speculation will require further investigation.

472 The ecological relevance, global distribution, and evolutionary patterns of multi-
473 functionality within GH39 β-xylosidases remain to be conclusively determined. Phylogenetic
474 analysis demonstrated the occurrence of nine out of ten amino acids substitutions/deletions in all
475 sequenced members of Class III-C (Table 4). In addition to anaerobic fungal sequences, Class
476 III-C β-xylosidases contain sequences from the genera *Clostridium* and *Teredinibacter* (Fig. 1).
477 Since it has been previously demonstrated that the xylanolytic machinery in anaerobic fungi,

478 including β-xylosidases, has been acquired from bacteria via horizontal gene transfer (Youssef et
479 al. 2013), and assuming that some or all of the amino acids substitutions/deletions in members of
480 class III-C collectively account for the observed multi-functionality, therefore we reason that the
481 observed distribution pattern suggests the evolution of relaxed specificity in GH39 β-xylosidases
482 within the domain Bacteria, prior to the acquisition of GH39 β-xylosidases by the anaerobic
483 fungi and that the acquired capability is speculated to be retained in all anaerobic fungal GH39 β-
484 xylosidases.

485 **Conclusions**

486 In conclusion, we have characterized a novel β-xylosidase that represents the first GH39-family
487 enzyme cloned and expressed from anaerobic fungi. The enzyme is multi-functional, capable of
488 hydrolyzing cellobiose, xylobiose, as well as several PNP-glycosides. It also displays high
489 affinity towards various substrates, retains activity over a wide range of temperatures and pHs,
490 and possesses excellent temperature and thermal stability. Structurally predictive modeling
491 identified putative differences which potentially could account for the observed relaxed
492 specificity. Collectively, these capabilities render Bgxg1 an excellent candidate for inclusion in
493 enzyme cocktails mediating cellulose and hemicellulose saccharification from lignocellulosic
494 biomass (Morrison et al. 2016).

495 **Acknowledgements**

496 We thank Dr. Gilbert John (Oklahoma State University) for supplying the *E. coli*
497 BL21(DE3)pLysS cells used in this study. We also thank Dr. Robert Gruninger for helpful
498 discussions.

499

500 **References**

501 Bessey OA, and Love RH. 1952. Preparation and measurement of the purity of the phosphatase
502 reagent, disodium para-nitrophenyl phosphate. *J Biol Chem* 196:175-178.
503 Bhalla A, Bischoff KM, and Sani RK. 2014. Highly thermostable GH39 beta-xylosidase from a
504 *Geobacillus* sp strain WSUCF1. *BMC Biotechnol* 14:963-973. Artn 963 10.1186/S12896-
505 014-0106-8
506 Borneman WS, Akin DE, and Ljungdahl LG. 1989. Fermentation products and plant-cell wall-
507 degrading enzymes produced by monocentric and polycentric anaerobic ruminal fungi.
508 *Appl Environ Microbiol* 55:1066-1073.
509 Brennan Y, Callen WN, Christoffersen L, Dupree P, Goubet F, Healey S, Hernandez M, Keller
510 M, Li K, Palackal N, Sittenfeld A, Tamayo G, Wells S, Hazlewood GP, Mathur EJ, Short
511 JM, Robertson DE, and Steer BA. 2004. Unusual microbial xylanases from insect guts.
512 *Appl Environ Microbiol* 70:3609-3617. Doi 10.1128/Aem.70.6.3609-3617.2004
513 Breuil C, and Saddler JN. 1985. Comparison of the 3,5-dinitrosalicylic acid and Nelson-Somogyi
514 methods of assaying for reducing sugars and determining cellulase activity. *Enzyme*
515 *Microb Tech* 7:327-332. Doi 10.1016/0141-0229(85)90111-5
516 Bronnenmeier K, and Staudenbauer WL. 1988. Purification and properties of an extracellular
517 beta-glucosidase from the cellulolytic thermophile *Clostridium stercorarium*. *Appl*
518 *Microbiol Biot* 28:380-386.
519 Cairns JRK, and Esen A. 2010. Beta-glucosidases. *Cell Mol Life Sci* 67:3389-3405.
520 10.1007/s00018-010-0399-2
521 Correa JM, Graciano L, Abrahao J, Loth EA, Gandra RF, Kadowaki MK, Henn C, and Simao
522 RDG. 2012. Expression and characterization of a GH39 beta-xylosidase II from

523 *Caulobacter crescentus*. *Appl Biochem Biotech* 168:2218-2229. 10.1007/s12010-012-
524 9931-1
525 Couger MB, Youssef NH, Struchtemeyer CG, Liggenstoffer AS, and Elshahed MS. *Accepted*.
526 Transcriptomics analysis of lignocellulosic biomass degradation by the anaerobic fungal
527 isolate *Orpinomyces* sp. strain C1A. *Biotechnol Biofuels*.
528 Czjzek M, Ben David A, Braman T, Shoham G, Henrissat B, and Shoham Y. 2005. Enzyme-
529 substrate complex structures of a GH39 beta-xylosidase from *Geobacillus*
530 *stearothermophilus*. *J Mol Biol* 353:838-846. 10.1016/j.jmb.2005.09.003
531 Czjzek M, Cicek M, Zamboni V, Burmeister WP, Bevan DR, Henrissat B, and Esen A. 2001.
532 Crystal structure of a monocotyledon (maize ZMGlu1) beta-glucosidase and a model of
533 its complex with p-nitrophenyl beta-D-thioglucoside. *Biochem J* 354:37-46. Doi
534 10.1042/0264-6021:3540037
535 Dashtban M, Maki M, Leung KT, Mao CQ, and Qin WS. 2010. Cellulase activities in biomass
536 conversion: measurement methods and comparison. *Crit Rev Biotechnol* 30:302-309.
537 10.3109/07388551.2010.490938
538 Ducros V, Czjzek M, Belaich A, Gaudin C, Fierobe HP, Belaich LP, Davies GJ, and Haser R.
539 1995. Crystal structure of the catalytic domain of a bacterial cellulase belonging to family
540 5. *Structure* 3:939-949. Doi 10.1016/S0969-2126(01)00228-3
541 Elshahed MS. 2010. Microbiological aspects of biofuel production: current status and future
542 directions. *J Adv Res* 1:103-111.
543 Garcia-Vallve S, Romeu A, and Palau J. 2000. Horizontal gene transfer of glycosyl hydrolases of
544 the rumen fungi. *Mol Biol Evol* 17:352-361.

545 Grassick A, Murray PG, Thompson R, Collins CM, Byrnes L, Birrane G, Higgins TM, and
546 Tuohy MG. 2004. Three-dimensional structure of a thermostable native
547 cellobiohydrolase, CBHIB, and molecular characterization of the cel7 gene from the
548 filamentous fungus, *Talaromyces emersonii*. *Eur J Biochem* 271:4495-4506.
549 10.1111/j.1432-1033.2004.04409.x
550 Harhangi HR, Freelove AC, Ubhayasekera W, van Dinther M, Steenbakkers PJ, Akhmanova A,
551 van der Drift C, Jetten MS, Mowbray SL, Gilbert HJ, and Op den Camp HJ. 2003.
552 Cel6A, a major exoglucanase from the cellulosome of the anaerobic fungi *Piromyces* sp.
553 E2 and *Piromyces equi*. *Biochim Biophys Acta* 1628:30-39.
554 Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo SJ, Clark DS, Chen F,
555 Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, and Rubin
556 EM. 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow
557 rumen. *Science* 331:463-467. 10.1126/science.1200387
558 Hill J, Nelson E, Tilman D, Polasky S, and Tiffany D. 2006. Environmental, economic, and
559 energetic costs and benefits of biodiesel and ethanol biofuels. *P Natl Acad Sci USA*
560 103:11206-11210. 10.1073/pnas.0604600103
561 Hovel K, Shallom D, Niefind K, Belakhov V, Shoham G, Baasov T, Shoham Y, and Schomburg
562 D. 2003. Crystal structure and snapshots along the reaction pathway of a family 51 alpha-
563 L-arabinofuranosidase. *EMBO J* 22:4922-4932. Doi 10.1093/Emboj/Cdg494
564 Kubicek CP. 1982. Beta-glucosidase excretion by *Trichoderma pseudokoningii* - correlation with
565 cell-wall bound beta-1.3-glucanase activities. *Arch Microbiol* 132:349-354. Doi
566 10.1007/Bf00413388

567 Kumar R, Singh S, and Singh OV. 2008. Bioconversion of lignocellulosic biomass: biochemical
568 and molecular perspectives. *J Ind Microbiol Biot* 35:377-391. 10.1007/s10295-008-0327-
569 8
570 Kumar S, and Ramon D. 1996. Purification and regulation of the synthesis of a beta-xylosidase
571 from *Aspergillus nidulans*. *FEMS Microbiol Lett* 135:287-293. DOI 10.1111/j.1574-
572 6968.1996.tb08003.x
573 Laemmli UK. 1970. Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-
574 T4. *Nature* 227:680-&. Doi 10.1038/227680a0
575 Liggenstoffer AS, Youssef NH, Wilkins MR, and Elshahed MS. 2014. Evaluating the utility of
576 hydrothermolysis pretreatment approaches in enhancing lignocellulosic biomass
577 degradation by the anaerobic fungus *Orpinomyces* sp. strain C1A. *J Microbiol Meth*
578 104:43-48. 10.1016/j.mimet.2014.06.010
579 Lineweaver H, and Burk D. 1934. The determination of enzyme dissociation constants. *J Am*
580 *Chem Soc* 56:658-666. Doi 10.1021/Ja01318a036
581 Liu ZL, Saha BC, and Slininger PJ. 2008. Lignocellulosic biomass conversion to ethanol by
582 *Saccharomyces*. In: Harwood CS, Demain AL, and Wall JD, eds. *Bioenergy*. Washington
583 D.C.: ASM Press.
584 Ljungdahl LG. 2008. The cellulase/hemicellulase system of the anaerobic fungus *Orpinomyces*
585 PC-2 and aspects of its applied use. *Ann Ny Acad Sci* 1125:308-321.
586 10.1196/annals.1419.030
587 Lombard V, Ramulu HG, Drula E, Coutinho PM, and Henrissat B. 2014. The carbohydrate-
588 active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42:D490-D495.
589 10.1093/nar/gkt1178

590 Matsuo M, and Yasui T. 1984. Purification and some properties of beta-xylosidase from
591 *Trichoderma viride*. *Agr Biol Chem* 48:1845-1852.
592 Morrison JM, Elshahed MS, and Youssef NH. 2016. A defined enzyme cocktail from the
593 anaerobic fungus *Orpinomyces* sp. strain C1A effectively releases sugars from pretreated
594 corn stover and switchgrass. *Sci Rep* Submitted.
595 Morrison JM, Wright CM, and John GH. 2012. Identification, isolation and characterization of a
596 novel azoreductase from *Clostridium perfringens*. *Anaerobe* 18:229-234.
597 10.1016/j.anaerobe.2011.12.006
598 National Research Council N. 2011. Renewable Fuel Standard: Potential Economic and
599 Environmental Effects of U.S. Biofuel Policy. Washington, D.C.
600 PyMol. The PyMOL Molecular Graphics System. Version 1.7.4 ed: Schrodinger, LLC.
601 Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ,
602 Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, and Tschaplinski T.
603 2006. The path forward for biofuels and biomaterials. *Science* 311:484-489.
604 10.1126/science.1114736
605 Roy A, Kucukural A, and Zhang Y. 2010. I-TASSER: a unified platform for automated protein
606 structure and function prediction. *Nat Protoc* 5:725-738. 10.1038/nprot.2010.5
607 Scheller HV, and Ulvskov P. 2010. Hemicelluloses. *Annu Rev Plant Biol* 61:263-289.
608 10.1146/annurev-arplant-042809-112315
609 Shao WL, Xue YM, Wu AL, Kataeva I, Pei JJ, Wu HW, and Wiegel J. 2011. Characterization of
610 a bovel beta-xylosidase, XylC, from *Thermoanaerobacterium saccharolyticum* JW/SL-
611 YS485. *Appl Environ Microbiol* 77:719-726. 10.1128/Aem.01511-10

612 Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, Lopez R, McWilliam H, Remmert
613 M, Soding J, Thompson JD, and Higgins DG. 2011. Fast, scalable generation of high-
614 quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:1-6.
615 Artn 539 10.1038/Msb.2011.75
616 Skalova T, Dohnalek J, Spiwok V, Lipovova P, Vondrackova E, Petrokova H, Duskova J, Strnad
617 H, Kralova B, and Hasek J. 2005. Cold-active beta-galactosidase from *Arthrobacter* sp
618 C2-2 forms compact 660 kDa hexamers: crystal structure at 1.9 angstrom resolution. *J*
619 *Mol Biol* 353:282-294. 10.1016/j.jmb.2005.08.028
620 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
621 large phylogenies. *Bioinformatics* 30:1312-1313. 10.1093/bioinformatics/btu033
622 Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. 2013. MEGA6: molecular
623 evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725-2729.
624 10.1093/molbev/mst197
625 vanPeij NNME, Brinkmann J, Vrsanska M, Visser J, and deGraaff LH. 1997. Beta-xylosidase
626 activity, encoded by xlnD, is essential for complete hydrolysis of xylan by *Aspergillus*
627 *niger* but not for induction of the xylanolytic enzyme spectrum. *Eur J Biochem* 245:164-
628 173.
629 Verdoucq L, Moriniere J, Bevan DR, Esen A, Vasella A, Henrissat B, and Czjzek M. 2004.
630 Structural determinants of substrate specificity in family 1 beta-glucosidases - Novel
631 insights from the crystal structure of sorghum dhurrinase-1, a plant beta-glucosidase with
632 strict specificity, in complex with its natural substrate. *J Biol Chem* 279:31796-31803.
633 10.1074/jbc.M402918200

634 Wagschal K, Franqui-Espiet D, Lee CC, Robertson GH, and Wong DWS. 2008. Cloning,
635 expression and characterization of a glycoside hydrolase family 39 xylosidase from
636 *Bacillus halodurans* C-125. *Appl Biochem Biotech* 146:69-78. 10.1007/s12010-007-
637 8055-5
638 Wang HC, Chen YC, Huang CT, and Hseu RS. 2013. Cloning and characterization of a
639 thermostable and pH-stable cellobiohydrolase from *Neocallimastix patriciarum* J11.
640 *Protein Expres Purif* 90:153-159. 10.1016/j.pep.2013.06.004
641 Yang JK, Yoon HJ, Ahn HJ, Lee BI, Pedelacq JD, Liong EC, Berendzen J, Laivenieks M, Vieille
642 C, Zeikus GJ, Vocadlo DJ, Withers SG, and Suh SW. 2004. Crystal structure of beta-D-
643 xylosidase from *Thermoanaerobacterium saccharolyticum*, a family 39 glycoside
644 hydrolase. *J Mol Biol* 335:155-165. 10.1016/j.jmb.2003.10.026
645 Yang JY, Yan RX, Roy A, Xu D, Poisson J, and Zhang Y. 2015. The I-TASSER Suite: protein
646 structure and function prediction. *Nat Methods* 12:7-8. 10.1038/nmeth.3213
647 Youssef NH, Couger MB, Struchtemeyer CG, Liggenstoffer AS, Prade RA, Najar FZ, Atiyeh
648 HK, Wilkins MR, and Elshahed MS. 2013. The genome of the anaerobic fungus
649 *Orpinomyces* sp strain C1A reveals the unique evolutionary history of a remarkable plant
650 biomass degrader. *Appl Environ Microbiol* 79:4620-4634. 10.1128/Aem.00821-13
651 Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40-
652 48. Artn 40 10.1186/1471-2105-9-40
653 Zhang YHP, Hong J, and Xinhao Y. 2009. Cellulase Assays. In: Mielenz JR, ed. *Biofuels:*
654 *Methods and Protocols*: Humana Press, 213-231.
655
656

657 **Figure Legends**

658 **Figure 1. Phylogenetic analysis of GH39 β-xylosidases, including Bgxg1.** Sequences
659 annotated as GH39 β-xylosidases (n=200 sequences, October 28, 2015) were retrieved from
660 CAZyme databases (Lombard et al. 2014). Genbank accession numbers are shown for reference
661 proteins (due to the unavailability of *Piromyces* proteins in Genbank, those proteins are shown as
662 JGI accession numbers). The Maximum Likelihood tree was generated in RAxML (Stamatakis
663 2014) using a BLOSUM62 substitution matrix and a GAMMA model of rate heterogeneity. The
664 model estimated an alpha parameter of 2.069. Bootstraps values (100 replicates) are shown for
665 nodes with >50 bootstrap support. The sequences were empirically classified into four classes
666 (Classes I-IV), and Class III, to which Bgxg1 is affiliated, is further classified into four distinct
667 lineages (III-A-III-D). The α-iduronidase sequence from *Mus musculus* was utilized as an
668 outgroup. β-xylosidases that were previously characterized biochemically were phylogenetically
669 affiliated with either Class II (*Bacillus halodurans* (BAB04787.1) and *Geobacillus*
670 *stearothermophilus* (ABI49941.1) in bottom Firmicutes wedge, and *Thermoanaerobacterium*
671 *saccharolyticum* (AAB68820.1) in middle Firmicutes wedge) or Class I (*Caulobacter crescentus*
672 (ACL95907.1), bottom α-Proteobacteria wedge). Bgxg1, from *Orpinomyces* sp. strain C1A, is
673 shown highlighted in yellow.

674

675 **Figure 2. SDS-PAGE analysis of Bgxg1.** A 12.5% SDS-PAGE analysis of recombinant Bgxg1
676 protein stained with Coomassie blue. *Lane A*, Pre-stained Protein Ladder (Caisson Labs,
677 Smithfield, Utah). *Lane B*, Purified Bgxg1.

678

679 **Figure 3. Effect of Temperature and pH on Bgxg1 activity A) Optimal pH, B) Optimal**
680 **Temperature, C) pH Stability, D) Thermal Stability.** All values are presented as relative
681 specific activities. Error bars represent standard deviation of triplicate (n=3) samples.
682
683 **Figure 4. Substrate competition and Bgxg1 preference.** Monosaccharides (glucose (■) or
684 xylose (ꕥ)) release was assayed when Bgxg1 was challenged with 10 mM cellobiose (**A**), 10
685 mM xylobiose (**B**), or an equimolar mixture of both substrates (**C**).
686
687 **Figure 5. Alignment of Bgxg1 and the four biochemically-characterized GH39-family**
688 **enzymes, highlighting structural predictions and conservation of or around the active site.**
689 Structural predictions for Bgxg1 sequence were obtained using I-TASSER three-dimensional
690 model (Fig. S1) (Roy et al. 2010; Yang et al. 2015; Zhang 2008). Bgxg1 sequence is compared
691 to those from *Caulobacter crescentus, Thermoanaerobacterium saccharolyticum, Geobacillus*
692 *stearothermophilus,* and *Bacillus halodurans.* α-helices in blue are those within the small α-
693 helical domain, α-helices and β-sheets in green are those within the $(\alpha/\beta)_8$ barrel, and β-sheets in
694 red are those within the β-sandwich. Red stars (*) represent catalytic residues within the active
695 site. Black stars (*) represent those residues close to the active site, as determined within the
696 Bgxg1 model. Blue stars (*) represent residues noted in the literature to be important for β-
697 xylosidase function (Czjzek et al. 2005; Yang et al. 2004).
698

Table 1 (on next page)

Table 1

Substrate Specificity and Specific Activity of Bgxg1

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2076v1 | CC-BY 4.0 Open Access | rec: 26 May 2016, publ: 26 May 2016

1 **Table 1. Substrate Specificity and Specific Activity of Bgxg1.**

Substrate[a]	Activity Tested	Specific Activity (U/mg ± SD)
PNPG	β-glucosidase	73.4 ± 7.15
Cellobiose	β-glucosidase	55.1 ± 5.36
PNPGal	β-galactosidase	54.6 ± 2.26
PNPX	β-xylosidase	11.5 ± 1.2
Xylobiose	β-xylosidase	10.9 ± 0.96
Beechwood Xylan	Xylanase	10.8 ± 1.25
Avicel	Exoglucanase	ND[b]
CMC	Endoglucanase	ND[b]
Locust Bean Gum	Mannanase	ND[b]
PNPA	Arabinosidase	ND[b]
PNPAc	Acetyl Xylan Esterase	ND[b]
PNPC	Cellobiohydrolase	ND[b]
PNPM	Mannosidase	ND[b]
Aldouronic acid	α-glucuronidase	ND[b]

2 **a** Abbreviations: PNPC - *p*-nitrophenyl-β-D-cellobioside, PNPX - *p*-nitrophenyl-β-D-xylopyranoside, PNPA - *p*-
3 nitrophenyl-β-D-arabinofuranoside, PNPM - *p*-nitrophenyl-β-D-mannoside, PNPG - *p*-nitrophenyl-β-D-
4 glucopyranoside, PNPGal - *p*-nitrophenyl-β-D-galactopyranoside, PNPAc - *p*-nitrophenyl-acetate.
5 b ND: Not detected.
6

Table 2 (on next page)

Table 2

Enzyme Kinetics for Bgxg1

1 **Table 2. Enzyme Kinetics for Bgxg1.**

Substrate[a]	Activity Tested	K_m[b]	V_{max} (U/mg)
PNPG	β-glucosidase	BDL[c]	769 ± 18
PNPGal	β-galactosidase	BDL[d]	769 ± 13
PNPX	β-xylosidase	0.00485 mM ± 0.00062	127 ± 8
Beechwood Xylan	Xylanase	0.038 mg/mL ± 0.0039	25.6 ± 10

2 a: Abbreviations: PNPG - *p*-nitrophenyl-β-D-glucopyranoside, PNPGal - *p*-nitrophenyl-β-D-galactopyranoside,
3 PNPX - *p*-nitrophenyl-β-D-xylopyranoside,.
4 b: K_m values are expressed in either mM or mg/mL, depending on the substrate tested. Values are shown ± standard
5 deviation of triplicate samples (n=3).
6 c: BDL: Below detection limit (500 nM). Extrapolated K_m value obtained using Lineweaver-Burke plot was
7 0.0000125 mM ± 0.0000096.
8 d: BDL: Below detection limit (500 nM). Extrapolated K_m value obtained using Lineweaver-Burke plot was
9 0.000214 mM ± 0.000016.
10

Table 3 (on next page)

Table 3

Substrate competition experiments

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2076v1 | CC-BY 4.0 Open Access | rec: 26 May 2016, publ: 26 May 2016

1

2 **Table 3. Substrate competition experiments.** "Activity tested" column refers to the colorimetric substrate tested
3 (PNPX for β-xylosidase, PNPG for β-glucosidase) in the presence of the active site inhibitor (cellobiose or
4 xylobiose, at listed Inhibitor concentrations). Specific activity, K_m, and V_{max} refer to the values calculated for the
5 colorimetric substrate in each experiment.

Activity Tested	Active Site Inhibitor	Inhibitor (mM)	Relative Specific Activity (%)	K_m (mM)	V_{max} (U/mg)
β-xylosidase	Cellobiose	0	100	0.00485	127
		10	78.9	1.438	118
		20	52.1	3.51	129
β-glucosidase	Xylobiose	0	100	0.0000125	769
		10	75.8	0.000235	763
		20	57.2	0.00349	752

6

7

Table 4(on next page)

Table 4

Comparison of key amino acids between Bgxg1 and all four biochemically characterized (BC) β-xylosidases from *Thermoanaerobacterium saccharolyticum, Bacillus halodurans, Geobacillus stearothermophilus, Caulobacter crescentus*, as well as in Classes III-A, III-B, and III-D.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2076v1 | CC-BY 4.0 Open Access | rec: 26 May 2016, publ: 26 May 2016

1 **Table 4. Comparison of key amino acids between Bgxg1 and all four biochemically characterized (BC) β-**
2 **xylosidases from *Thermoanaerobacterium saccharolyticum, Bacillus halodurans, Geobacillus***
3 ***stearothermophilus, Caulobacter crescentus*, as well as in Classes III-A, III-B, and III-D[a].**

Pos. [b]	AA in Bgxg1	AA in 4 BC[d]	Significance of Change	Importance of Residue	Class III-A	Class III-B	Class III-C	Class III-D
46	Val	Tyr	Small, nonpolar (Val) vs. Large, polar (Tyr)	Near active site	NC[e]	Ile	Val	NC
129	Asp	Asn	Negative charge (Asp) vs. Neutral charge (Asn)	H-bonding	Lys	Asp	Asp	Asp
131[c]	Thr/NC	Phe	Small, polar (Thr) vs. Large, nonpolar (Phe)	Near active site	NC	NC	NC	NC
139	Phe	Tyr	Large, nonpolar (Phe) vs. Large, polar (Tyr)	Near active site	Tyr	Tyr	Phe	Tyr
163	Cys	Ala	Polar, thiol (Cys) vs. Nonpolar (Ala)	Near active site	Tyr	Ala	Cys	Tyr
171	Lys	Trp	Positive charge (Lys) vs. Nonpolar (Trp)	Near active site	Trp	NC	Lys	Lys
194	Leu	Tyr	Small, nonpolar (Leu) vs. Large, polar (Tyr)	H-bonding	Ser	Ile/Glu	Leu	Tyr
242	Arg	Ala	Positive charge (Arg) vs. Small, nonpolar (Ala)	Near active site	NC	NC	Arg	NC
322-323	-gap-	Glu	Gap vs. Negative charge (Glu)	H-bonding	Arg/Thr/Lys	-gap-	-gap-	-gap-
322-323	-gap-	Glu	Gap vs. Negative charge (Glu)	H-bonding	Gly/-gap-	-gap-	-gap-	-gap-

4
5 [a] No changes were identified in 10 different positions (Arg48, Ile124, Trp125, Asn126, Glu127, Pro128, Trp132,
6 Pro162, His192, Glu225), and 5 positions were variable across all sequences (Val81, Gly130, Tyr164, Ser165,
7 Lys247).
8 [b] Pos. (Positions) refer to the position of the amino acid in Bgxg1
9 [c] Bgxg1 and all proteins in Class III-C β-xylosidases have identical amino acid sequences in all key positions with
10 one exception (Thr131)
11 [d] Sequences identified using the alignment presented in Fig. 5.
12 [e] NC = Not conserved
13

14

Figure 1(on next page)

Phylogenetic analysis of GH39 β-xylosidases,

Figure 1. Phylogenetic analysis of GH39 β-xylosidases, including Bgxg1. Sequences annotated as GH39 β-xylosidases (n=200 sequences, October 28, 2015) were retrieved from CAZyme databases (Lombard et al. 2014) . Genbank accession numbers are shown for reference proteins (due to the unavailability of *Piromyces* proteins in Genbank, those proteins are shown as JGI accession numbers). The Maximum Likelihood tree was generated in RAxML (Stamatakis 2014) using a BLOSUM62 substitution matrix and a GAMMA model of rate heterogeneity. The model estimated an alpha parameter of 2.069. Bootstraps values (100 replicates) are shown for nodes with >50 bootstrap support. The sequences were empirically classified into four classes (Classes I-IV), and Class III, to which Bgxg1 is affiliated, is further classified into four distinct lineages (III-A-III-D). The α-iduronidase sequence from *Mus musculus* was utilized as an outgroup. β-xylosidases that were previously characterized biochemically were phylogenetically affiliated with either Class II (*Bacillus halodurans* (BAB04787.1) and *Geobacillus stearothermophilus* (ABI49941.1) in bottom Firmicutes wedge, and *Thermoanaerobacterium saccharolyticum* (AAB68820.1) in middle Firmicutes wedge) or Class I (*Caulobacter crescentus* (ACL95907.1), bottom α-Proteobacteria wedge). Bgxg1, from *Orpinomyces* sp. strain C1A, is shown highlighted in yellow.

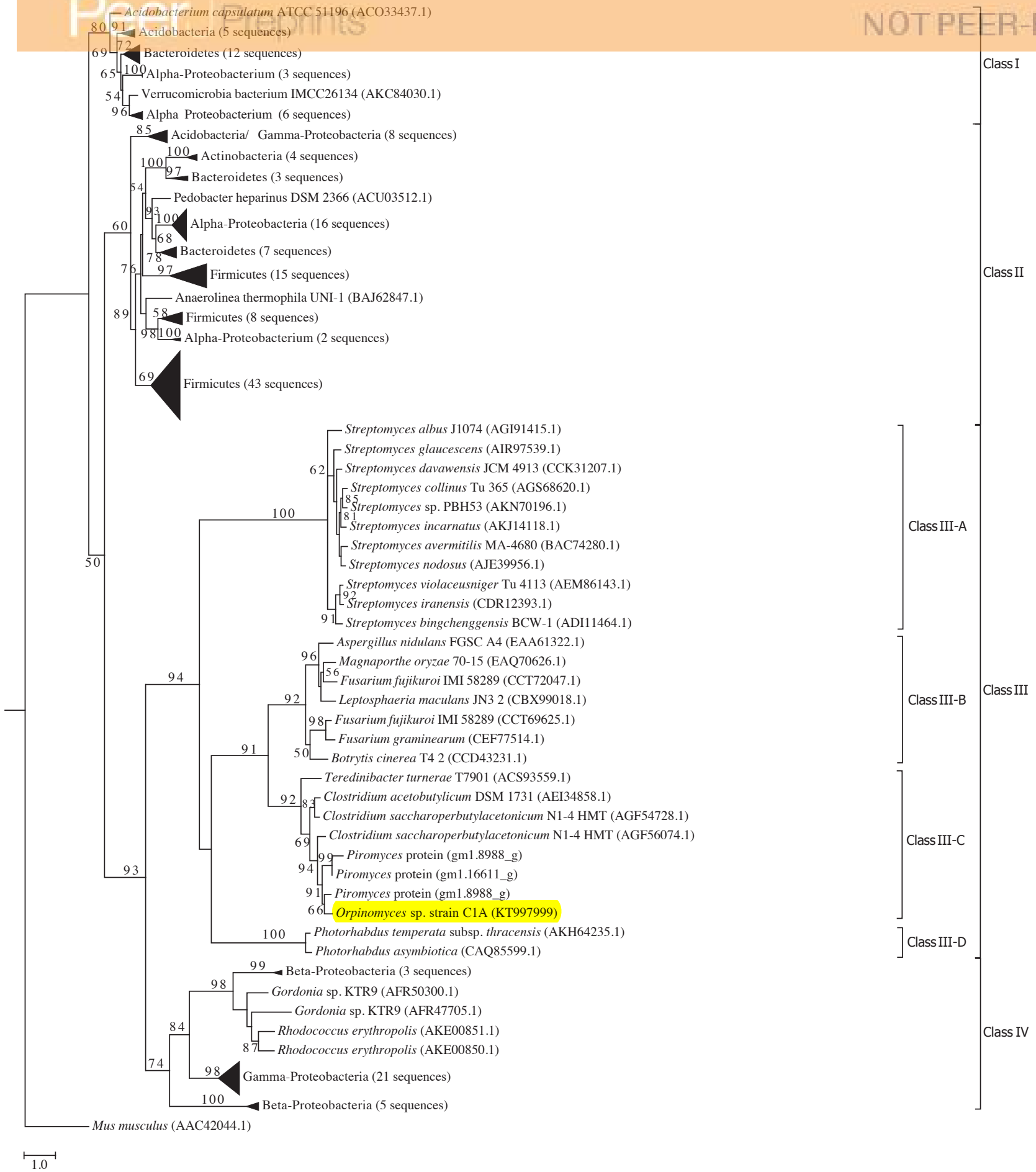

Figure 2(on next page)

SDS-PAGE analysis of Bgxg1.

A 12.5% SDS-PAGE analysis of recombinant Bgxg1 protein stained with Coomassie blue. Lane A , Pre-stained Protein Ladder (Caisson Labs, Smithfield, Utah). Lane B , Purified Bgxg1.

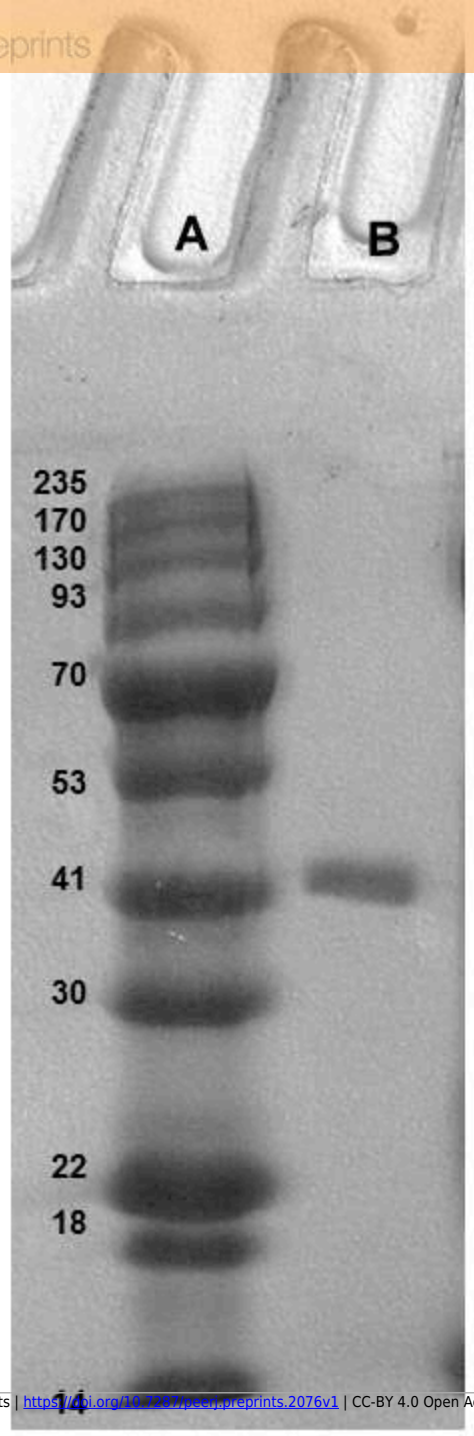

Figure 3(on next page)

Figure 3

Effect of Temperature and pH on Bgxg1 activity A) Optimal pH, B) Optimal Temperature, C) pH Stability, D) Thermal Stability.

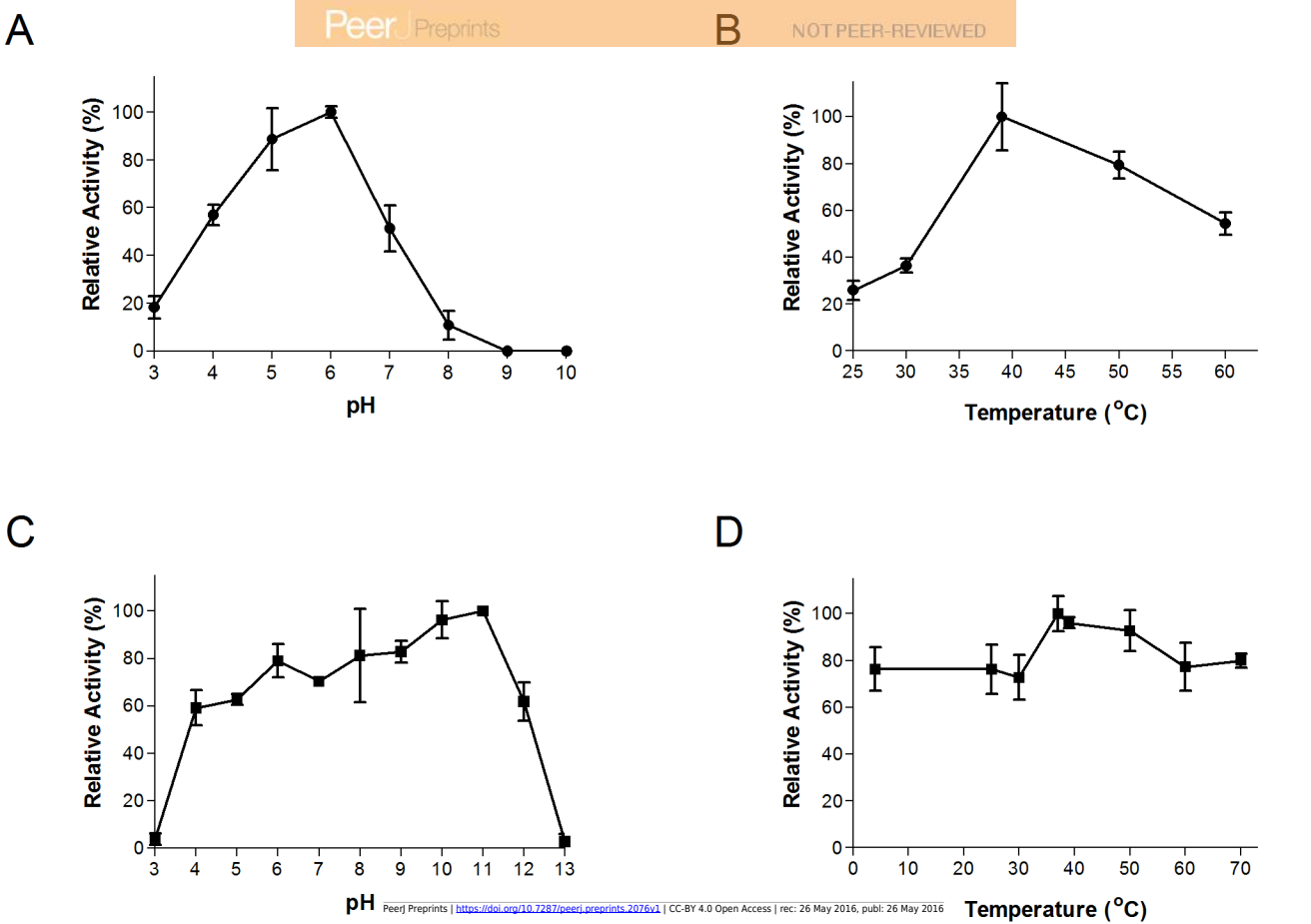

Figure 4(on next page)

Figure 4

Substrate competition and Bgxg1 preference.

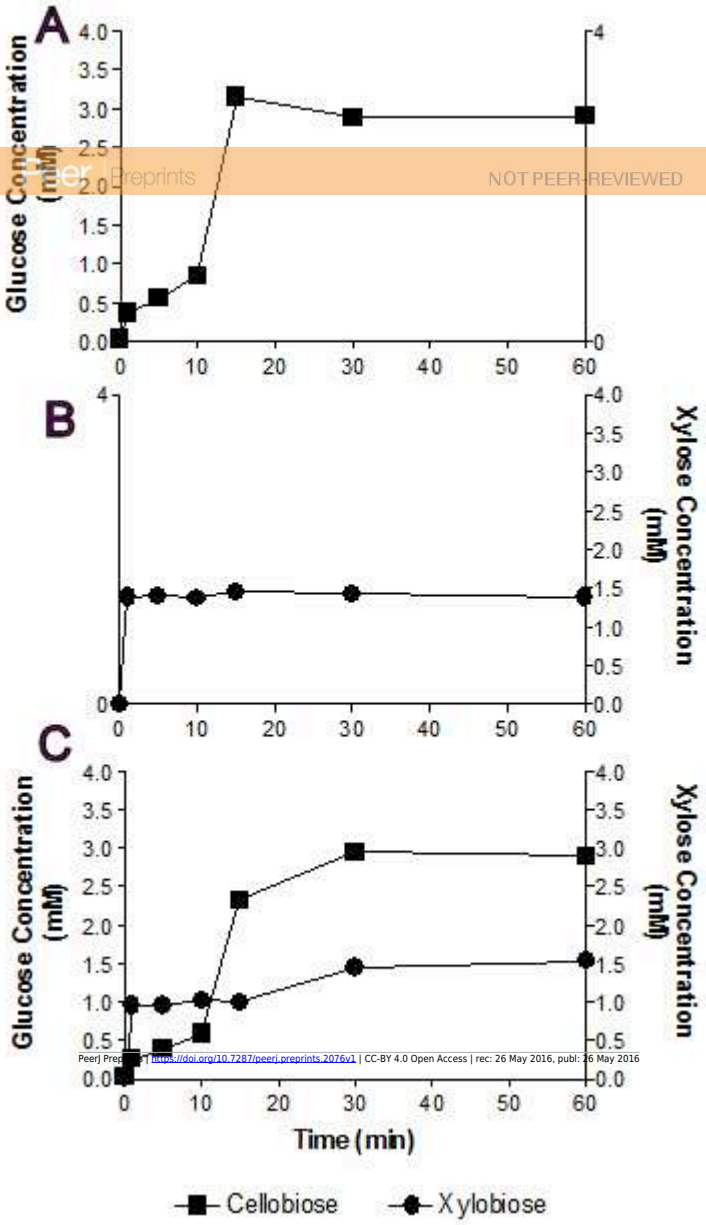

Figure 5(on next page)

Figure 5

Alignment of Bgxg1 and the four biochemically-characterized GH39-family enzymes, highlighting structural predictions and conservation of or around the active site.

α-1 α-2 α-3 β-1

```
Bgxg1                    1 MT......NVLTVE...CNNKL.RRATHCANGSLYGIT..ETTPRDYKNLVDPLHPFVMRN.............PARGGNG
Caulobacter              1 MANAGPGARVIDLDLRRAAGPVDRFFDLSIGSDYPGTLIREDSQAQLKTTVDELGFRYIRFHAIFHDVLGTVKVQ.....
Thermoanaerobacterium    1        MIKVRVP.DFSDKKFSDRWRYCVGTGRLGLALQKEYIETLKYVKENIDFKYIRGHGLLCDDVGIYRED.VVGD
Geobacillus              1        MKVVNVP.SNGREKFKKNWKFCVGTGRLGLALQKEYLDHLKLVQEKIGFRYIRGHGLLSDDVGIYREV.EIDG
Bacillus                 1        MKTVVVN.DRSFAYFPKKWKYCIGTGRLGLALQKEYVDHLARLQKELNFQYIRGHGLLHDDIGIYRERKRADG
```

α-4 β-2 α-5

```
Bgxg1                   57 NQHPYGDAI...............KVARRLADTPGALVSVDLPDMLPGWPYKWPG.......MQNWLNQVKSFI....KDK
Caulobacter             76 DGKIVYDWTKIDQLYDALLAKGIKPFIELGFTPEAMKTSDQ......TIFYWKGNTSHPK.LGPWRDLIDAFVHHLRARY
Thermoanaerobacterium   72 EVKPFYNFTYIDRIFDSFLEIGIRPFVEIGFMPKKLASGTQ......TVFYWEGNVTPPKDYEKWSDLVKAVLHHFISRY
Geobacillus             72 EMKPFYNFTYIDRIVDSYLALNIRPFIEFGFMPKALASGDQ......TVFYWKGNVTPPKDYNKWRDLIVAVVSHFIERY
Bacillus                73 TVEPFYNFTYIDRIFDTFLELNIRPFVEIGFMPKLLASGEQ......TIFDWQGNVTPPKDYDQWKQLIQAVISHFIDRY
```

α-5 β-3 α-6 β-4 α-7 β-5

```
Bgxg1                  112 KASGLKNWYGLEIWNEPD..GTWNNSN.GSFEEMWKQTYQAIRQADPNEKIIGPCYSWYTDDKLRNFLKYAKANNCLPDI
Caulobacter            149 GVEEVRTW.FFEVWNEPNLDGFWEKADQAAYFELYDVTARAIKAIDPSLRVGGPATAGA..AWVPEFLAHVKKSGSAVDF
Thermoanaerobacterium  146 GIEEVLKW.PFEIWNEPNLKEFWKDADEKEYFKLYKVTAKAIKEVNENLKVGGPAICGGADYWIEDFLNFCYEENVPVDF
Geobacillus            146 GIEEVRTW.LFEVWNEPNLVNFWKDANKQEYFKLYEVTARAVKSVDPHLQVGGPAICGGSDEWITDFLHFCAERRVPVDF
Bacillus               147 GVEEVTKW.PFEIWNEPNLINFWQHADKKEYFKLYKITARAIKEVHPYIQVGGPAICGGSDEWITDFLQFCHKEEVPVDF
```

β-5 α-8 β-6 α-9

```
Bgxg1                  189 ISWHELSGI...................DGVSSHLRSYREIEKSLGIPELPISINEYCDAKHELEGQPG..SSARFIG.KF
Caulobacter            226 VTTHTYGVDGGFLDEKGVQDTKLSPSPDAVVGDVRRVREQIEASAFPGLPLYFTEWSTSYTPRDSVHDSYVSAAYIVEKL
Thermoanaerobacterium  225 VSRHAYTSKQG.EYTPHLIYQEIMP.SEYMLNEFKTVREIIKNSHFPNLPFHITEYNTSYSPQNPVHDTPFNAAYIARIL
Geobacillus            225 VSRHAYTSKAPHKKTFEYYYQELEP.PEDMLEQFKTVRALIRQSPFPHLPLHITEYNTSYSPINPIHDTALNAAYIARIL
Bacillus               226 VSRHAYTSAKPHKVTPDYYYQELYE.NTHMLDELKSVKELIQQSPFPNLPFHITEYNTSYSPINPVHDTVLNAAYLARIL
```

α-9 β-7 α-10 β-8 β-9

```
Bgxg1                  248 ERYKV.DTAMITWWFV...........PLPGRLGSLLATDTQKGAGWYFYKWYGDMTGDMVYVKPPNDNSNLVDGAACVD
Caulobacter            306 RRVKGLVQAMSYWTYSDLFEEPGPPTAPFQGGFG.LMNPQGIRKPSWFAYKYLNALKGRELVCADDQVFAARDGDRVA..
Thermoanaerobacterium  303 SEGGDYVDSFSYWTFSDVFEERDVPRSQFHGGFG.LVALNMIPKPTFYTFKFFNAMGEEMLYRDEHMLVTRRDGDSVA..
Geobacillus            304 SEGGDYVDSFSYWTFSDVFEEMDVPKALFHGGFG.LVALHSIPKPTFHAFTFFNALGDELLYRDGEMIVTRRKDGSIA..
Bacillus               305 SEAGDIVDSFSYWTFSDVFEEAGVPTAPFHGGFG.LIALHGIAKPTYHLFSFFNQLGEQLLYRDSQMVVTKKQDGSIQ..
```

β-10 β-11

```
Bgxg1                  316 TNKEYISFIFGGPNDGTI.RASSIIFQALLDLLPML
Caulobacter            383 ....IVAYAWRQPDQKVSNRPFYTKLHPASDVEPLKVRLTSLKPGRYKLRVRRVGYRRNDAYSAYIDMGSPTTLTESQLQ
Thermoanaerobacterium  380 ....LI..AWNEVMDKTENPDEDYEVE.....IPVRF.......RDVFIKRQLIDEEHGNPWGTWIHMGRPRYPSKEQVN
Geobacillus            381 ....AV..LWNLVMEKGEGFTKEVQLV.....IPVSF.......SAVFIKRQIVNEQYGNAWRVWKQMGRPRFPSRQAVE
Bacillus               382 ....LV..VWNLIMEKGEGLEQTVQIE.....LPTQS.......DAVFIKRKTIDETNGNPWRVWKEMGRPRFPKKNEID

Bgxg1                  351
Caulobacter            459 SLQALTEDRPEIEKALKVSGETVVDLPMRANDVVLIELEPLA
Thermoanaerobacterium  442 TLREVAKPEIMTSQPVANDGYLNLKFKLGKNAVVLYELTERIDESSTYIGLDDSKINGY
Geobacillus            443 TLRQVAQPHVMTEQRRATDGVIHLSIVLSKNEVTLIEIEQVRDETSTYVGLDDGEMTSYSS
Bacillus               444 TLQEVAQPLIRTERRDGLSRTLKLELTLSKNEVSLVEVFPIIDETNTYPGLDDRLIPSY
```