Characterization of an α-glucosidase enzyme conserved in *Gardnerella* spp. isolated from the human vaginal microbiome

Pashupati Bhandari[a], Jeffrey P. Tingley[b,c], David R. J. Palmer[d], D. Wade Abbott[b,c] and Janet E. Hill[a,#]

[a]Department of Veterinary Microbiology, Western College of Veterinary Medicine, University of Saskatchewan, 52 Campus Drive, Saskatoon, Saskatchewan, S7N 5B4, Canada

[b]Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, 5401-1st Ave S, Lethbridge, Alberta, T1J 4B1, Canada

[c]Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Dr W, Lethbridge, AB T1K 3M4, Canada

[d]Department of Chemistry, University of Saskatchewan, 110 Science Place, Saskatoon, SK S7N 5C9, Canada

[#]To whom correspondence should be addressed

e-mail addresses:

PB: pab230@mail.usask.ca

JPT: jeffrey.tingley@uleth.ca

DRJP: dave.palmer@usask.ca

DWA: wade.abbott@canada.ca

JEH: Janet.Hill@usask.ca

Abstract

Gardnerella spp. in the vaginal microbiome are associated with bacterial vaginosis, a dysbiosis in which a lactobacilli dominant microbial community is replaced with mixed aerobic and anaerobic bacteria including *Gardnerella* species. The co-occurrence of multiple *Gardnerella* species in the vaginal environment is common, but different species are dominant in different women. Competition for nutrients, particularly glycogen present in the vaginal environment, could play an important role in determining the microbial community structure. Digestion of glycogen into products that can be taken up and further processed by bacteria requires the combined activities of several enzymes collectively known as amylases, which belong to glycoside hydrolase family 13 (GH13) within the CAZy classification system. GH13 is a large and diverse family of proteins, making prediction of their activities challenging. SACCHARIS annotation of the GH13 family in *Gardnerella* resulted in identification of protein domains belonging to eight subfamilies. Phylogenetic analysis of predicted amylase sequences from 26 *Gardnerella* genomes demonstrated that a putative α-glucosidase-encoding sequence, CG400_06090, was conserved in all species in the genus. The predicted α-glucosidase enzyme was expressed, purified and functionally characterized. The enzyme was active on a variety of maltooligosaccharides over a broad pH range (4.0 - 8) with maximum activity at pH 7. The K_m, k_{cat} and k_{cat}/K_m values for the substrate 4-nitrophenyl α-D-glucopyranoside were 8.3 μM, 0.96 min^{-1} and 0.11 $\mu\mathrm{M}^{-1}\mathrm{min}^{-1}$ respectively. Glucose was released from maltose, maltotriose, maltotetraose and maltopentaose, but no products were detected on thin layer chromatography when the enzyme was incubated with glycogen. Our findings show that *Gardnerella* spp. produce an α-glucosidase enzyme that may contribute to the complex and multistep process of glycogen metabolism by releasing glucose from maltooligosaccharides.

Keywords: glycogen; *Gardnerella;* vaginal microbiome; α-glucosidase; glycoside hydrolase

Introduction

Gardnerella spp. in the vaginal microbiome are hallmarks of bacterial vaginosis, a condition characterized by replacement of the lactobacilli dominant microbial community with mixed aerobic and anaerobic bacteria including *Gardnerella*. This dysbiosis is associated with increased vaginal pH, malodorous discharge and the presence of biofilm (1). In addition to troubling symptoms, the presence of abnormal vaginal microbiota is associated with increased risk of HIV transmission and infection with other sexually transmitted pathogens such as *Neisseria gonorrhoeae* and *Trichomonas* spp. (2, 3). Historically, *Gardnerella* has been considered a single species genus. Jayaprakash *et al*. used cpn60 barcode sequences to divide *Gardnerella* spp. into four subgroups (A-D) (4), and this framework was supported by whole genome sequence comparison (5, 6). More recently, Vaneechoutte *et al.* emended the classification of *Gardnerella* based on whole genome sequence comparison, biochemical properties and matrix-assisted laser desorption ionization time-of-flight mass spectrometry and proposed the addition of three novel species: *Gardnerella leopoldii*, *Gardnerella piotii* and *Gardnerella swidsinskii* (7).

Colonization with multiple *Gardnerella* species is common, and different species are dominant in different women (8). Understanding factors that contribute to differential abundance is important since the species may differ in virulence (6) and they are variably associated with clinical signs (8, 9). Several factors, including inter-specific competition, biofilm formation and resistance to antimicrobials could contribute to the differential abundance of the different species. Khan *et al.* showed that resource-based scramble competition is frequent among *Gardnerella*

70 subgroups (10). Glycogen is a significant nutrient for vaginal microbiota, but previous reports on
71 the growth of *Gardnerella* spp. on glycogen containing medium are inconsistent (11). Species may
72 differ in their ability to digest glycogen and utilize the breakdown products, which may in turn
73 contribute to determining microbial community structure.

74 Glycogen is an energy storage molecule, which consists of linear chains of approximately
75 13 glucose molecules covalently linked with α-1,4 glycosidic linkages, with branches attached
76 through α-1,6 glycosidic bonds (12, 13). A single glycogen molecule consists of approximately
77 55,000 glucose residues with a molecular mass of ~10^7 kDa (14). The size of glycogen particles
78 can vary with source; from 10-44 nm in human skeletal muscle to approximately 110-290 nm in
79 human liver (15). Glycogen is deposited into the vaginal lumen by epithelial cells under the
80 influence of estrogen (16), and the concentration of cell-free glycogen in vaginal fluid varies
81 greatly, ranging from 0.1 - 32 μg/μL (17). This long polymeric molecule must be digested into
82 smaller components that can be taken up by bacterial cells and metabolized further (18, 19).

83 Glycogen digestion is accomplished by the coordinated action of enzymes collectively
84 described as amylases (20). A large majority of these enzymes belong to family 13 within the
85 glycosyl hydrolase class (GH13) of carbohydrate active enzymes (CAZymes) (21). GH13 enzymes
86 are further classified into more than 43 subfamilies based on structure and activity (22). For
87 example, subfamily 23 is primarily composed of α-glucosidases, which are exo-acting enzymes
88 that act on α-1,4 glycosidic bonds from the non-reducing end to release glucose; subfamily 32 is
89 composed of α-amylases, endo-acting enzymes that cleave α-1,4 glycosidic bonds to produce
90 maltose and α-limit dextrin; and subfamily 14 contains debranching enzymes, such as pullulanase,
91 which target α-1,6 glycosidic bonds (15, 23). Limited information is available on glycogen
92 degradation mechanisms in the vaginal microbiome. Although it has been demonstrated that

93 vaginal secretions exhibit amylase activity that can degrade glycogen (19), the role of bacterial
94 amylase enzymes in this process is unknown and there is limited information about glycogen
95 metabolism by clinically important bacteria such as *Gardnerella* spp..
96 The objective of the current study was to annotate GH13 enzymes in *Gardnerella* spp. and
97 to characterize the activity of a putative α-glucosidase conserved among all *Gardnerella* spp..
98
99 **Methods**
100
101 ***Gardnerella* isolates**
102 Isolates of *Gardnerella* used in this study were from a previously described culture
103 collection (6) and included representatives of *cpn*60 subgroup A (*G. leopoldii* (n = 4), *G.*
104 *swidsinskii* (n = 6), other (n = 1)), subgroup B (*G. piotii* (n = 5), Genome sp. 3 (n = 2)), subgroup
105 C (*G. vaginalis* (n = 6)), and subgroup D (Genome sp. 8 (n = 1), Genome sp. 9 (n = 1)). Whole
106 genome sequences of these isolates had been determined previously (BioProject Accession
107 PRJNA394757) and annotated by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
108 (24).
109
110 **Identification and functional annotation of putative amylase sequences**
111 Putative amylase sequences were identified in proteomes of 26 *Gardnerella* isolates used
112 in this study based on the PGAP annotations. Multiple sequence alignments of predicted amylase
113 sequences were performed using CLUSTALw with results viewed and edited in AliView (version
114 1.18) (25) prior to phylogenetic tree building using PHYLIP (26). SignalP v 5.0 (27) and
115 SecretomeP 2.0 (28) were used to identify signal peptides.

116 Putative amylolytic enzymes in *G. vaginalis* ATCC 14018 and *G. swidsinskii* GV37
117 (genomes available in the CAZy database) were identified using dbCAN, and predicted sequences
118 were run through the SACCHARIS pipeline. SACCHARIS combines user sequences with CAZy
119 derived sequences and trims sequences to the catalytic domain using dbCAN2 (29). Domain
120 sequences were aligned with MUSCLE (30), and a best-fit model was generated with ProtTest
121 (31). Final trees were generated with FastTree (32) and visualized with iTOL (33). Alignment of
122 CG400_06090 from *G. leopoldii* NR017 to orthologues from *Halomonas* sp. (BAL49684.1),
123 *Xanthomonas campestris* (BAC87873.1), and *Culex quinquefasciatus* (ASO96882.1) was done
124 using CLUSTALw and visualized in ESPript (34). A predicted structure model of CG400_06090
125 was aligned with *Halomonas* sp. HaG (BAL49684.1, PDB accession 3WY1) (35) to identify
126 putative catalytic residues using PHYRE2 (36). BLASTp (37) alignment of CAZy derived
127 sequences from reference genomes to a database of the predicted proteomes of 26 study isolates
128 was conducted to identify orthologs.
129
130 **Expression and purification of the CG400_06090 gene product**
131 Genomic DNA from *G. leopoldii* NR017 was extracted using a modified salting out
132 procedure, and the CG400_06090 open reading frame (locus tag CG400_06090 in GenBank
133 accession NNRZ01000007, protein accession RFT33048) was PCR amplified with primers
134 JH0729-F (5'ATG CAT GCG CAT TAT ACG ATC ATG CTC-3') and JH0730-R (5'ATG GTA
135 CCT TAC ATT CCA AAC ACT GCA-3'). Underlined sequences indicate *Sph*I and *Kpn*I
136 restriction enzyme sites. PCR reaction contained 1 × PCR reaction buffer (0.2 M Tris-HCl pH 8.4,
137 0.5 M KCl), 200 μM each dNTPs, 2.5 mM $MgCl_2$, 400 nM each primer, 1 U/reaction of Platinum
138 Taq DNA polymerase high-fidelity (5 units/μL in 50% glycerol, 20 mM Tris-HCl, 40 mM NaCl,

139 0.1 mM EDTA, and stabilizers) (Life Technologies) and 2 μL of template DNA, in a final volume
140 of 50 μL. PCR was performed with following parameters: initial denaturation at 94 °C for 3
141 minutes followed by (denaturation at 94 °C, 15 seconds; annealing at 60 °C, 15 seconds and
142 extension at 72 °C, 2 minutes; 35 cycles), and final extension at 72 °C for 5 minutes. Purified PCR
143 products were digested with *Kpn*I and *Sph*I and ligated into expression vector pQE-80L (Qiagen,
144 Mississauga, ON) digested with same restriction endonucleases. The resulting recombinant
145 plasmid was used to transform One Shot TOP10 chemically competent *E. coli* cells (Invitrogen,
146 Carlsbad, California). Colony PCR was performed to identify transformants containing vector with
147 insert. Insertion of the putative amylase gene in-frame with the N-terminal 6×Histidine tag was
148 confirmed by sequencing of the purified plasmid.

149 *E. coli* cells containing the plasmid were grown overnight at 37 °C in LB medium with 100
150 μg/ml ampicillin. Overnight culture was diluted 1:60 in fresh medium to a final volume of 50 mL,
151 and the culture was incubated at 37 °C with shaking at 225 rpm until it reached an OD_{600} of 0.6.
152 At this point, expression was induced with 0.1 mM IPTG and a further 4 hours of incubation was
153 done at 37°C at 225 rpm. Cells were harvested by centrifugation at 10,000 × *g* for 30 mins and the
154 pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH
155 8.0). Lysozyme was added to 1 mg/mL and cells were lysed by sonication (8 mins total run time
156 with 15 sec on and 30 sec off). The lysate was clarified by centrifugation at 10,000 × *g* for 30 mins
157 at 4 °C and the supernatant was applied to an Ni-NTA affinity column according to the
158 manufacturer's instructions (Qiagen, Germany). Bound proteins were washed twice with wash
159 buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0) and eluted with elution buffer
160 (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). Eluted protein was buffer
161 exchanged with the same elution buffer but without imidazole using 30 kDa MWCO protein

162 concentrator (ThermoFisher Scientific). Final protein concentration was measured by
163 spectrophotometry.
164
165 **Enzyme assay and kinetics**
166 Purified protein was tested for α-glucosidase activity by measuring the release of 4-
167 nitrophenol from a chromogenic substrate 4-nitrophenyl α-D-glucopyranoside as described
168 elsewhere (38). Briefly, 200 μL of enzyme solution (50 nM) was added to 200 μL of 20 mM 4-
169 nitrophenyl α-D-glucopyranoside substrate and the reaction mixture was incubated at 37 °C. A 50-
170 μL aliquot of the reaction mixture was removed and added to 100 μL of 1 M Na_2CO_3 solution
171 every 90 seconds up to 7.5 mins (total 6 different time points). The absorbance of this solution was
172 measured at 420 nm, and the amount of 4-nitrophenol released was calculated from a standard
173 curve. To determine activity at different pH values, the substrate was prepared in 50 mM sodium
174 phosphate buffer (pH 4.0-8.0). Kinetic constants were determined for 4-nitrophenyl α-D-
175 glucopyranoside by measuring rate of reaction of enzyme (50 nM) with substrate concentrations
176 of 0.05, 1, 2, 5, 10, 20 and 25 mM at pH 7. Values were fit to the Michaelis-Menten equation, $v =$
177 k_{cat} $[E]_0[S]/(K_m + [S])$, where v is the observed rate of reaction, $[E]_0$ is the initial enzyme
178 concentration, [S] is the substrate concentration, K_m is the Michaelis constant, and k_{cat} is the
179 turnover number. Kinetics calculations were performed in GraphPad Prism 8.
180
181 **Enzymatic activity on different substrates**
182 The activity of the purified enzyme was examined with different substrates. Briefly, 750
183 μL of enzyme solution (50 nM) was added to 750 μL of 3 mM maltose, maltotriose, maltotetraose,
184 maltopentaose, or 0.1% (w/v) of maltodextrins (MD 4-7, MD 13-17 and MD 16.5-19.5) or bovine

185 liver glycogen in 50 mM sodium phosphate buffer, pH 6.0, and the reaction mixture was incubated
186 aerobically at 37 °C. Aliquots of the reaction mixture (250 µL) were removed at various time
187 intervals of incubation and were heat-inactivated at 93 °C for 1 minute. The reaction mixtures were
188 centrifuged at 10,000 × *g* for 1 min and supernatant was stored at -20 ºC and analyzed by thin layer
189 chromatography (TLC) and high performance anion exchange chromatography (HPAEC) with
190 pulsed amperometric detection (PAD). TLC was performed on silica plates in 1-butanol: acetic
191 acid: distilled water (2:1:1 v/v/v) mobile phase and stained using ethanol: sulfuric acid (70:3 v/v)
192 solution containing 1% (w/v) orcinol monohydrate (Sigma). HPAEC-PAD samples were diluted
193 to 50 µM and resolved on a Dionex PA20 column in a mobile phase of 30 mM NaOH and gradient
194 of 10 mM – 120 mM sodium acetate over 40 min at a flow rate of 0.5 mL min^{-1}. Data was analyzed
195 using Chromeleon v6.80 chromatography data system software. Maltooligosaccharide standards
196 ranging from maltose to maltopentaose (Carbosynth) were used for both TLC and HPAEC-PAD
197 analysis; whereas, isomaltotriose, D-panose, and 6-o-a-D-glucosyl-maltose (Carbosynth) were
198 used to identify minor peaks resolved by HPAEC-PAD.
199

200 **Results**

201 **Identification of putative amylase sequences**

202 A total of 60 proteins from the 26 study isolates were annotated by PGAP as amylases. After
203 removal of severely truncated sequences, a multiple sequence alignment of the remaining 42
204 sequences was trimmed to the uniform length of 290 amino acids. Phylogenetic analysis of these
205 predicted amylase sequences revealed five clusters, but only one of these clusters contained
206 orthologous sequences from all 26 isolates (Figure 1). Sequence identities within this cluster were
207 all 91-100% identical. This conserved sequence (corresponding to GenBank accession

208 WP_004104790) was annotated as an GH13 α-amylase (EC 3.2.1.1) in all genomes, suggesting it
209 was an endo-acting enzyme that releases maltooligosaccharides from glycogen.

210 The GH13 family in the CAZy database is a large, polyspecific family targeting α-glucosyl
211 linkages. The functional diversity of members in this family can make annotation of new sequences
212 challenging. In order to predict the activity of the conserved protein identified among the study
213 isolates, we employed SACCHARIS, which combines CAZyme family trees generated from
214 biochemically characterized proteins with related sequences of unknown function. Currently, two
215 *Gardnerella* reference genomes are available in CAZy: *G. vaginalis* ATCC 14018 and *G.*
216 *swidsinkii* GV37. GH13 family proteins encoded in these *Gardnerella* genomes were identified
217 and a family tree including all characterized GH13 sequences in the CAZy database was generated
218 (Figure S1). SACCHARIS annotation of the GH13 family in *Gardnerella* resulted in identification
219 of protein domains belonging to eight subfamilies (2, 9, 11, 14, 20, 23, 30, 31 and 32). Alignment
220 with characterized GH13 enzymes from the CAZy database (22) suggested that the conserved
221 amylase identified in the study isolates was an α-glucosidase (EC 3.2.1.20), closely related to
222 subfamilies 23 and 30 of GH13, and not an α-amylase as annotated in GenBank. Subsequent
223 examination of the α-glucosidase with SignalP and SecretomeP indicated that no signal peptide
224 was present, and that the protein was not likely secreted through a Sec-independent pathway.

225

226 **Relationship to other α-glucosidases**

227 A representative sequence of the predicted α-glucosidase was selected from *G. leopoldii*
228 (NR017). This sequence (CG400_06090) was combined with functionally characterized members
229 of the GH13 CAZy database using SACCHARIS (Figure 2A). CG400_06090 partitioned with
230 members of GH13 subfamilies 17, 23 and 30, of which the prominent activity is α-glucosidase.

231 This clade was expanded to clarify the relationship of CG400_06090 with related characterized
232 sequences as an apparently deep-branching member of subfamily 23, although with low bootstrap
233 support (Figure 2B). When additional, uncharacterized sequences from the CAZy database were
234 included, CG400_06090 clustered within subfamily 23 (Figure S2).

235 To identify conserved catalytic residues between CG400_06090 and characterized
236 members of GH13, CG400_06090 was aligned with structurally characterized members of
237 subfamily 23 (BAL49684.1 and BAC87873.1) and subfamily 17 (ASO96882.1) (Figure 2C). Of
238 the GH13 catalytic triad, the aspartate nucleophile and the aspartate that stabilizes the transition
239 state are conserved among the four sequences (39). The glutamate general acid/base, however,
240 does not appear to be sequence conserved between CG400_06090 and the GH13 subfamily 23
241 members. Upon modelling CG400_06090 in PHYRE2 and aligning with BAL49684.1 (PDBID
242 3WY1), CG400_06090 Glu256 did appear to be spatially conserved with the general acid/base
243 (not shown).

244

245 **Expression and purification of α-glucosidase protein**

246 The *G. leopoldi* NR017 CG400_06090 open reading frame comprises 1701 base pairs
247 encoding a 567 amino acid protein, with a predicted mass of 62 kDa. The full-length open reading
248 frame encoding amino acids 2-567 was PCR-amplified and ligated into vector pQE-80L for
249 expression in *E. coli* as an N-terminal hexahistidine-tagged protein. A distinct protein band in
250 SDS-PAGE between 50 kDa and 75 kDa was obtained from IPTG-induced *E. coli* cells compared
251 with non-induced cells, indicating the expression of the α-glucosidase protein (Figure 3A). The
252 recombinant protein was soluble, and was purified using a Ni-NTA spin column (Figure 3B). From
253 a 50 mL broth culture, approximately 700 μL of purified protein at 4.0 mg/mL was obtained.

254

255 **Effect of pH on enzyme activity**

256 α-Glucosidase activity of the purified protein was demonstrated by release of 4-nitrophenol
257 from the chromogenic substrate 4-nitrophenyl α-D-glucopyranoside. A preliminary analysis of
258 enzyme activity over pH 3-8 showed that product was produced over a broad pH range from 4.0
259 to 8.0 (Figure S3). To more precisely examine effects of pH on activity a pH rate profile was
260 determined and the maximum rate was observed at pH 7.0 (Figure 4A). The dependence of rate on
261 substrate concentration fit the Michaelis-Menten equation (Figure 4B), with a Michaelis constant
262 (K_m) of 8.3 μM, and a V_{max} 96 μM/min, corresponding to a k_{cat} value of 0.96 (± 0.01) min^{-1} and
263 k_{cat}/K_m of 0.11 $\mu M^{-1}min^{-1}$.

264

265 **Analysis of substrate hydrolysis**

266 Production of glucose was detected when the purified α-glucosidase was incubated with
267 maltose (M2), maltotriose (M3), maltotetraose (M4) and maltopentaose (M5) (Figure 5). Samples
268 analyzed by HPAEC-PAD contained large peaks corresponding to maltooligosaccharides ranging
269 from maltose to maltopentaose; minor peaks, especially in the maltotetraose and maltopentaose
270 samples, resolved products not observed in the TLC analysis. These peaks did not align to
271 isomaltotriose, panose, or 6-o-a-D-glucosyl-maltose standards in HPAEC-PAD (not shown), and
272 likely represent larger, mixed linkage products. No appreciable activity was detected on
273 maltodextrins (MD 4-7, MD 13-17 and MD 16.5-19.5) or glycogen (Figure S4).

274

275 **Discussion**

276 Glycogen is a major nutrient available to vaginal microbiota and its utilization likely plays
277 an important role in the survival and success of *Gardnerella* spp. in the vaginal microbiome. The
278 ability of *Gardnerella* spp. to utilize glycogen has been reported in previous studies. Dunkelburg
279 *et al.* described the growth of four strains of what was then known as *Haemophilus vaginalis* in
280 buffered peptone water with 1% glycogen suggesting their ability to ferment glycogen, but no
281 additional details were provided (40). Similarly, Edmunds reported the glycogen fermenting ability
282 of 14 out of 15 *Haemophilus vaginalis* isolates, although serum was included in the media (41).
283 Piot *et al.* examined the starch fermentation ability of 175 *Gardnerella* isolates on Mueller Hinton
284 agar supplemented with horse serum and found that all *Gardnerella* strains were capable of
285 hydrolysing starch (42). They did not, however, assess glycogen degradation and further, it is not
286 clear if serum amylase contained in the media contributed to the observed amylase activity.
287 Robinson, in his review, reported that starch can be utilized by *Gardnerella* (43) while Catlin in
288 another review, reported that glycogen utilization is inconsistent in *Gardnerella* (11).
289 Glycogen is digested into smaller products, such as maltose and maltodextrins, by a group
290 of enzymes collectively known as amylases. Amylases that contribute to glycogen utilization in
291 the vaginal environment are not well studied. Spear *et al*. (19) suggested a role for a human
292 amylase enzyme in the breakdown of vaginal glycogen. They detected host-derived pancreatic α-
293 amylase and acidic α-glucosidase in genital fluid samples collected from women and showed that
294 genital fluid containing these enzymes can degrade the glycogen. The presence of these enzyme
295 activities alone in vaginal fluid, however, was not correlated with glycogen degradation,
296 suggesting contributions of activity from the resident microbiota. Many vaginal bacteria likely
297 produce amylases that contribute to this activity as various anaerobic taxa are known to encode
298 amylases and digest glycogen (44), however, the contributions of the vaginal microbiota to this

299 process and the importance of these processes to vaginal microbial community dynamics are not
300 yet well understood. Regardless of the source, amylases digest glycogen into maltose and or
301 maltooligosaccharides which are transported inside bacteria via ABC transporter systems for
302 further processing (45).

303 Protein sequences belonging to eight different GH13 subfamilies that could have roles in
304 glycogen degradation were identified in *G. vaginalis* ATCC 14018 and *G. swidsinskii* GV37
305 (Figure S1), and multiple amylase-like proteins were detected in the proteomes of additional
306 *Gardnerella* spp. (Figure 1), all of which were annotated as "alpha-amylase" in the GenBank
307 records. Our results demonstrate the value of using more nuanced analyses, including dbCAN and
308 SACCHARIS, to guide discovery of CAZymes based on comparison at a catalytic domain level
309 to functionally characterized enzymes. Further investigation will be required to demonstrate the
310 actual functions of these proteins, and how they contribute to carbohydrate-degrading activities of
311 *Gardnerella* spp..

312 In the phylogenetic analysis of putative amylase sequences, one cluster contained
313 sequences common to all 26 isolates (Figure 1). Although this protein sequence was annotated as
314 an α-amylase, suggesting it would produce maltooligosaccharides from glycogen, SACCHARIS
315 predicted it to be closely related to α-glucosidases, which cleave glucose from the non-reducing
316 end of its substrate. Automated annotation of sequences deposited in primary sequence databases
317 such as GenBank has resulted in significant problems with functional mis-annotation (46). This
318 problem is exacerbated by the increasing rate at which whole genome sequences are being
319 generated and deposited. The genome sequences of the study isolates were annotated using the
320 NCBI Prokaryotic Genome Annotation Pipeline, which identifies genes and annotates largely

321 based on sequence similarity to previously annotated sequences, thus potentially propagating
322 incorrect functional annotations.

323 Phylogenetic analysis of GH13 family enzymes showed that the conserved α-glucosidase
324 from *Gardnerella* spp. appears to be most closely related to GH13 subfamily 23 (Figure S2).
325 Primary sequences of GH13 members of amylases have seven highly conserved regions and three
326 amino acids that form the catalytic triad (23). Comparison of the primary sequence and three-
327 dimensional model of CG400_06090 with characterized α-glucosidases from GH13 subfamily 17
328 and 23 did show sequence conservation of nucleophile and spatial conservation of the catalytic
329 triad.

330 Our findings suggest that, although α-glucosidase does not hydrolyze glycogen, it can
331 digest smaller maltooligosaccharides. The purified α-glucosidase enzyme was able to completely
332 digest maltose and maltotriose to glucose, but digestion of maltotetraose and maltopentaose was
333 incomplete, suggesting a preference for smaller oligosaccharides (Figure 5). Previously
334 characterized α-glucosidase (*algB*) from *Bifidobacterium adolescentis* DSM20083 also showed
335 higher activity against maltose and maltotriose while no activity was reported against
336 maltotetraose and maltopentaose (47). α-Glucosidases can have different substrate specificities
337 due to the variable affinity of the substrate binding site for particular substrates (48).

338 Most secreted proteins have a short N-terminal signal peptide to guide for extracellular
339 translocation of the newly synthesized protein (49), however, bacterial proteins can be also
340 secreted via non-classical secretion pathways in a Sec-independent manner, without having a
341 signal peptide (28). Analysis of the *Gardnerella* α-glucosidase sequence with SignalP and
342 SecretomeP showed that the protein is unlikely to be secreted by either route and thus likely acts
343 on intracellular substrates including products of extracellular glycogen degradation transported

344 into the cell. Many free-living and host-associated bacteria synthesize glycogen as a storage
345 molecule (50) but the extent to which this occurs in *Gardnerella* is not known. The possibility that
346 products of digestion of intracellular glycogen stores are substrates for the α-glucosidase
347 characterized in this study remains a question for future study. Purified α-glucosidase from *G.*
348 *leopoldii* NR017 was active across a broad pH range of 4.0 to 8.0 (Figure S3) with the highest rate
349 detected at pH 7.0 (Figure 4A). This is not unexpected for a cytoplasmic enzyme in a host-
350 associated bacterium, and consistent with at least one other report of an α-glucosidase from a
351 commensal bacterium, *Bifidobacterium adolescentis* (47).

352 Our findings show that *Gardnerella* spp. have an α-glucosidase enzyme that likely
353 contributes to the complex and multistep process of glycogen utilization by releasing glucose from
354 maltooligosaccharides. Identification and biochemical characterization of additional enzymes
355 involved in glycogen metabolism will provide insight into whether utilization of this abundant
356 carbon source is an important factor in population dynamics and competition among *Gardnerella*
357 spp.. The functional annotation strategy demonstrated here provides a powerful approach to guide
358 future experiments aimed at determining enzyme substrates and activities.

359

360 **Funding**

361 This research is supported by an NSERC Discovery grant to JEH, and Agriculture and Agri-Food
362 Canada project number: J-001589 (JPT and DWA). PB is supported by a Devolved Scholarship
363 from the University of Saskatchewan.

364

365 **Conflicts of interest**

366 The authors declare no competing interest.

367

368 **Acknowledgements**

369 The authors are grateful to Josseline Ramos-Figueroa, Douglas Fansher and Natasha Vetter
370 (Department of Chemistry, University of Saskatchewan) for guidance with thin layer
371 chromatography and Champika Fernando for technical support.

372

373 **References**

374 1. Nugent RP, Krohn MA, Hillier SL. 1991. Reliability of diagnosing bacterial vaginosis is
375 improved by a standardized method of gram stain interpretation. J Clin Microbiol 29:297–
376 301.

377 2. Martin HL, Richardson BA, Nyange PM, Lavreys L, Hillier SL, Chohan B, Mandaliya K,
378 Ndinya-Achola JO, Bwayo J, Kreiss J. 1999. Vaginal *Lactobacilli*, microbial flora, and risk
379 of Human Immunodeficiency virus Type 1 and sexually transmitted disease acquisition. J
380 Infect Dis 180:1863–1868.

381 3. Taha TE, Hoover DR, Dallabetta GA, Kumwenda NI, Mtimavalye LAR, Yang L-P, Liomba
382 GN, Broadhead RL, Chiphangwi JD, Miotti PG. 1998. Bacterial vaginosis and disturbances
383 of vaginal flora: association with increased acquisition of HIV. AIDS 12.

384 4. Paramel Jayaprakash T, Schellenberg JJ, Hill JE. 2012. Resolution and characterization of
385 distinct cpn60-based subgroups of *Gardnerella vaginalis* in the vaginal microbiota. PloS
386 One 7:e43009–e43009.

387 5. Ahmed A, Earl J, Retchless A, Hillier SL, Rabe LK, Cherpes TL, Powell E, Janto B, Eutsey
388 R, Hiller NL, Boissy R, Dahlgren ME, Hall BG, Costerton JW, Post JC, Hu FZ, Ehrlich
389 GD. 2012. Comparative genomic analyses of 17 clinical isolates of *Gardnerella vaginalis*
390 provide evidence of multiple genetically isolated clades consistent with subspeciation into
391 genovars. J Bacteriol 194:3922–3937.

392 6. Schellenberg JJ, Paramel Jayaprakash T, Withana Gamage N, Patterson MH, Vaneechoutte
393 M, Hill JE. 2016. *Gardnerella vaginalis* subgroups defined by cpn60 sequencing and
394 Sialidase activity in isolates from Canada, Belgium and Kenya. PloS One 11:e0146510.

395 7. Vaneechoutte M, Guschin A, Van Simaey L, Gansemans Y, Van Nieuwerburgh F, Cools P.
396 2019. Emended description of *Gardnerella vaginalis* and description of *Gardnerella*
397 *leopoldii* sp. nov., *Gardnerella piotii* sp. nov. and *Gardnerella swidsinskii* sp. nov., with
398 delineation of 13 genomic species within the genus *Gardnerella*. Int J Syst Evol Microbiol
399 69:679–687.

400 8. Hill JE, Albert AYK. 2019. Resolution and cooccurrence patterns of *Gardnerella leopoldii*,
401 *G. swidsinskii*, *G. piotii*, and *G. vaginalis* within the vaginal microbiome. Infect Immun
402 87:e00532-19.

403 9. Balashov SV, Mordechai E, Adelson ME, Gygax SE. 2014. Identification, quantification and
404 subtyping of *Gardnerella vaginalis* in noncultured clinical vaginal samples by quantitative
405 PCR. J Med Microbiol 63:162–175.

406 10. Khan S, Voordouw MJ, Hill JE. 2019. Competition among *Gardnerella* subgroups from the
407 human vaginal microbiome. Front Cell Infect Microbiol 9:374.

408 11. Catlin BW. 1992. *Gardnerella vaginalis*: characteristics, clinical considerations, and
409 controversies. Clin Microbiol Rev 5:213–237.

410 12. Meléndez-Hevia E, Waddell TG, Shelton ED. 1993. Optimization of molecular design in the
411 evolution of metabolism: the glycogen molecule. Biochem J 295 (Pt 2):477–483.

412 13. Deng B, Sullivan MA, Chen C, Li J, Powell PO, Hu Z, Gilbert RG. 2016. Molecular
413 structure of human liver glycogen. PloS One 11:e0150540–e0150540.

414 14. Roach PJ, Depaoli-Roach AA, Hurley TD, Tagliabracci VS. 2012. Glycogen and its
415 metabolism: some new developments and old themes. Biochem J 441:763–787.

416 15. Adeva-Andany MM, González-Lucán M, Donapetry-García C, Fernández-Fernández C,
417 Ameneiros-Rodríguez E. 2016. Glycogen metabolism in humans. BBA Clin 5:85–100.

418 16. Cruickshank R, Sharman A. 1934. The biology of the vagina in the human subject. BJOG Int
419 J Obstet Gynaecol 41:208–226.

420 17. Mirmonsef P, Hotton AL, Gilbert D, Gioia CJ, Maric D, Hope TJ, Landay AL, Spear GT.
421 2016. Glycogen levels in undiluted genital fluid and their relationship to vaginal pH,
422 estrogen, and progesterone. PloS One 11:e0153553.

423 18. Flint HJ, Scott KP, Duncan SH, Louis P, Forano E. 2012. Microbial degradation of complex
424 carbohydrates in the gut. Gut Microbes 3:289–306.

425 19. Spear GT, French AL, Gilbert D, Zariffard MR, Mirmonsef P, Sullivan TH, Spear WW,
426 Landay A, Micci S, Lee B-H, Hamaker BR. 2014. Human α-amylase present in lower-

427 genital-tract mucosal fluid processes glycogen to support vaginal colonization by
428 Lactobacillus. J Infect Dis 210:1019–1028.

429 20. Berg JM, Tymoczko JL, Stryer L. 2002. Glycogen breakdown requires the interplay of
430 several enzymes.Biochemistry, 5th ed. W H Freeman, New York.

431 21. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The
432 Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.
433 Nucleic Acids Res 37:D233–D238.

434 22. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The
435 carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495.

436 23. Mehta D, Satyanarayana T. 2016. Bacterial and archaeal α-amylases: Diversity and
437 amelioration of the desirable characteristics for industrial applications. Front Microbiol 7.

438 24. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze
439 A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation
440 pipeline. Nucleic Acids Res 44:6614–6624.

441 25. Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large
442 datasets. Bioinforma Oxf Engl 30:3276–3278.

443 26. Felsenstein J. 1989. PHYLIP - phylogeny inference package (version 3.2). Cladistics 5:164–
444 166.

445 27. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S,
446 von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep
447 neural networks. Nat Biotechnol 37:420–423.

448 28. Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. 2005. Non-classical protein secretion in
449 bacteria. BMC Microbiol 5:58.

450 29. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018.
451 dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic
452 Acids Res 46:W95–W101.

453 30. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and
454 space complexity. BMC Bioinformatics 5:113.

455 31. Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit
456 models of protein evolution. Bioinformatics 27:1164–1165.

457 32. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood
458 Trees for Large Alignments. PLOS ONE 5:e9490.

459 33. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
460 developments. Nucleic Acids Res 47:W256–W259.

461 34. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new
462 ENDscript server. Nucleic Acids Res, 2014/04/21 ed. 42:W320–W324.

463 35. Shen X, Saburi W, Gai Z, Kato K, Ojima-Kato T, Yu J, Komoda K, Kido Y, Matsui H, Mori
464 H, Yao M. 2015. Structural analysis of the α-glucosidase HaG provides new insights into
465 substrate specificity and catalytic mechanism. Acta Crystallogr Sect D 71:1382–1391.

466 36. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal
467 for protein modeling, prediction and analysis. Nat Protoc 10:845–858.

468 37. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search
469 tool. J Mol Biol 215:403–410.

470 38. Angelov A, Putyrski M, Liebl W. 2006. Molecular and biochemical characterization of
471 alpha-glucosidase and alpha-mannosidase and their clustered genes from the
472 thermoacidophilic archaeon *Picrophilus torridus*. J Bacteriol 188:7123–7131.

473 39. Uitdehaag JCM, Mosi R, Kalk KH, van der Veen BA, Dijkhuizen L, Withers SG, Dijkstra
474 BW. 1999. X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase
475 elucidate catalysis in the α-amylase family. Nat Struct Biol 6:432–436.

476 40. Dunkelberg WE, McVeigh I. 1969. Growth requirements of *Haemophilus vaginalis*. Antonie
477 Van Leeuwenhoek 35:129–145.

478 41. Edmunds PN. 1962. The biochemical, serological and haemagglutinating reactions of
479 *Haemophilus vaginalis*. J Pathol Bacteriol 83:411–422.

480 42. Piot P, Van Dyck E, Totten PA, Holmes KK. 1982. Identification of *Gardnerella*
481 (*Haemophilus*) *vaginalis*. J Clin Microbiol 15:19–24.

482 43. Taylor-Robinson D. 1984. The bacteriology of *Gardnerella vaginalis*. Scand J Urol Nephrol
483 Suppl 86:41.

484 44. Nunn KL, Forney LJ. 2016. Unraveling the dynamics of the human vaginal microbiome.
485 Yale J Biol Med 89:331–337.

486 45. Boos W, Shuman H. 1998. Maltose/Maltodextrin System of Escherichia coli: Transport,
487 Metabolism, and Regulation. Microbiol Mol Biol Rev 62:204–229.

488 46. Schnoes AM, Brown SD, Dodevski I, Babbitt PC. 2009. Annotation error in public
489 databases: misannotation of molecular function in enzyme superfamilies. PLoS Comput
490 Biol 5:e1000605–e1000605.

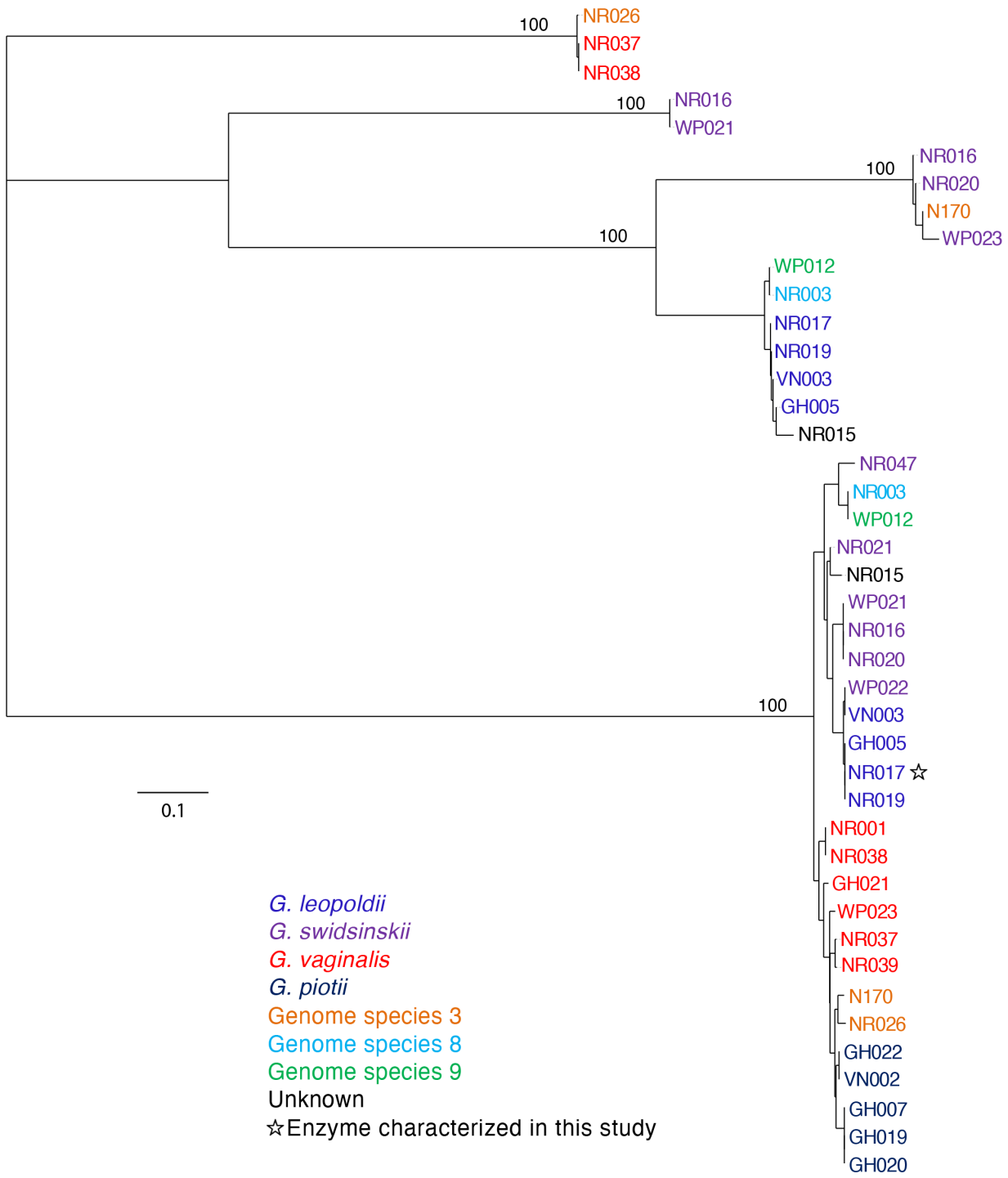
491 47. Van den Broek L, Struijs K, Verdoes J, Beldman G, Voragen A. 2003. Cloning and
492 characterization of two α-glucosidases from *Bifidobacterium adolescentis* DSM20083.
493 Appl Microbiol Biotechnol 61:55–60.

494 48. Okuyama M, Saburi W, Mori H, Kimura A. 2016. α-Glucosidases and α-1,4-glucan lyases:
495 structures, functions, and physiological actions. Cell Mol Life Sci 73:2727–2751.

496 49. Owji H, Nezafat N, Negahdaripour M, Hajiebrahimi A, Ghasemi Y. 2018. A comprehensive
497 review of signal peptides: Structure, roles, and applications. Eur J Cell Biol 97:422–441.

498 50. Preiss J. 1984. Bacterial glycogen synthesis and its regulation. Annu Rev Microbiol 38:419–
499 458.

500



501

502 **Figure 1.** Phylogenetic analysis of predicted extracellular amylases protein sequences from 26

503 *Gardnerella* genomes. Trees are consensus trees of 100 bootstrap iterations. Bootstrap values are

504 shown for major branch points. Species is indicated by label colour according to the legend.

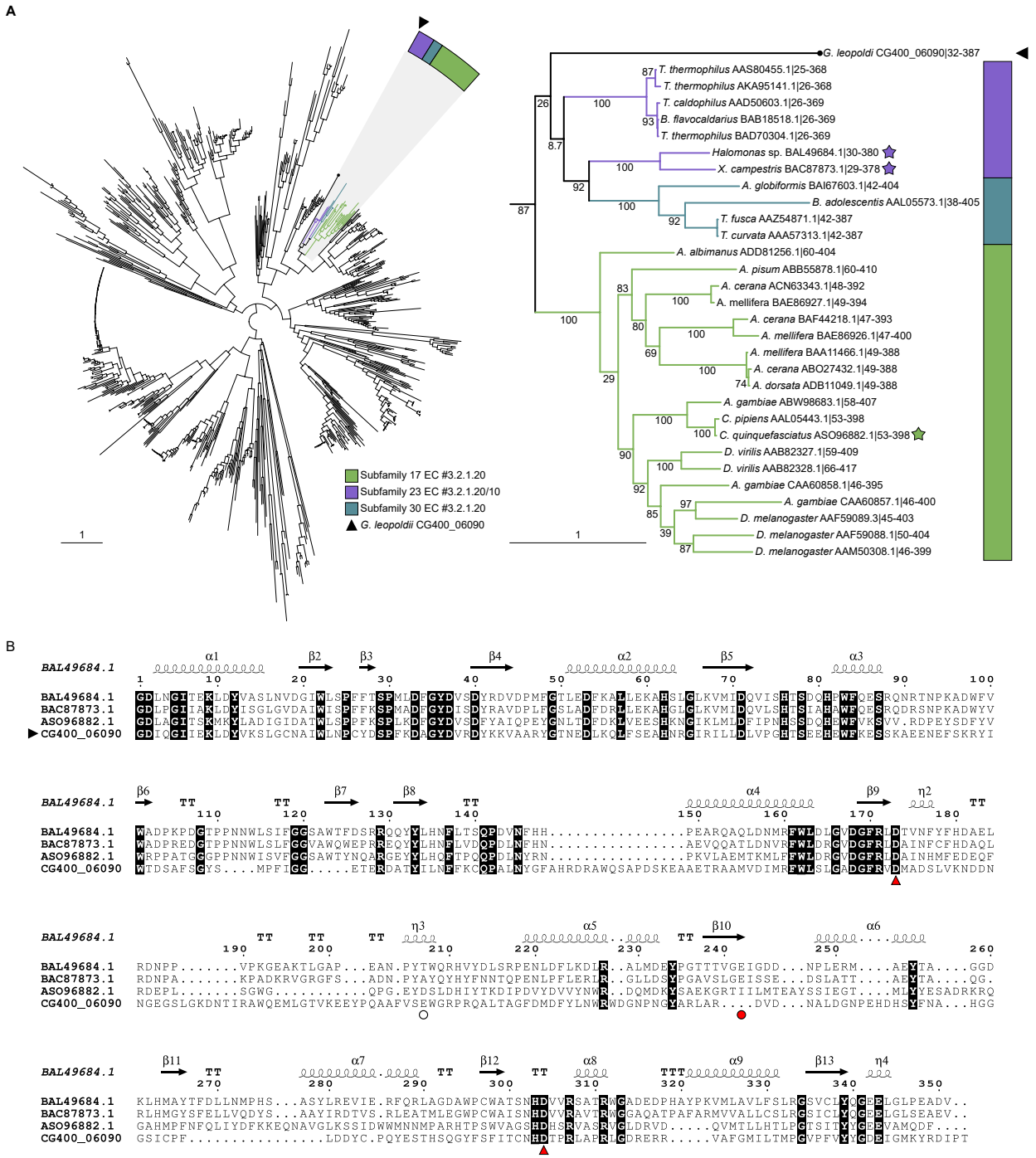
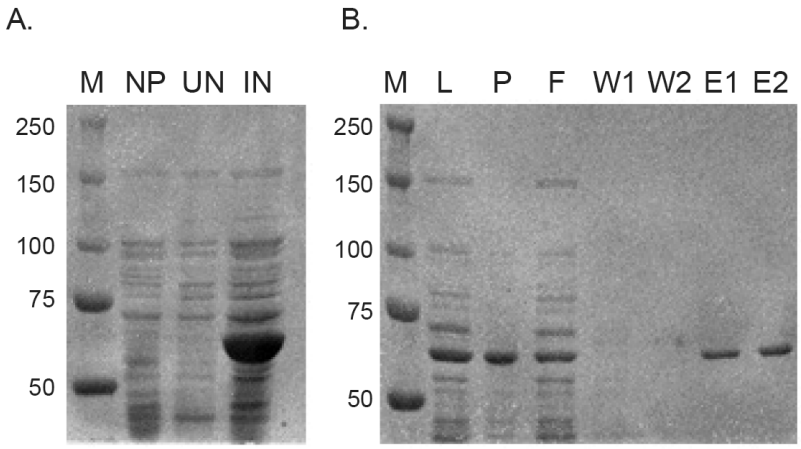


Figure 2. (A) Phylogenetic tree of characterized GH13 members in the CAZy database and predicted *G. leopoldii* NR017 GH13 sequence CG400_06090. (B) Expanded tree of subfamilies 17, 23 and 30. Stars indicate structurally characterized sequences and bootstrap values were included. (C) CLUSTAL alignment of the catalytic domain of CG400_06090 with catalytic

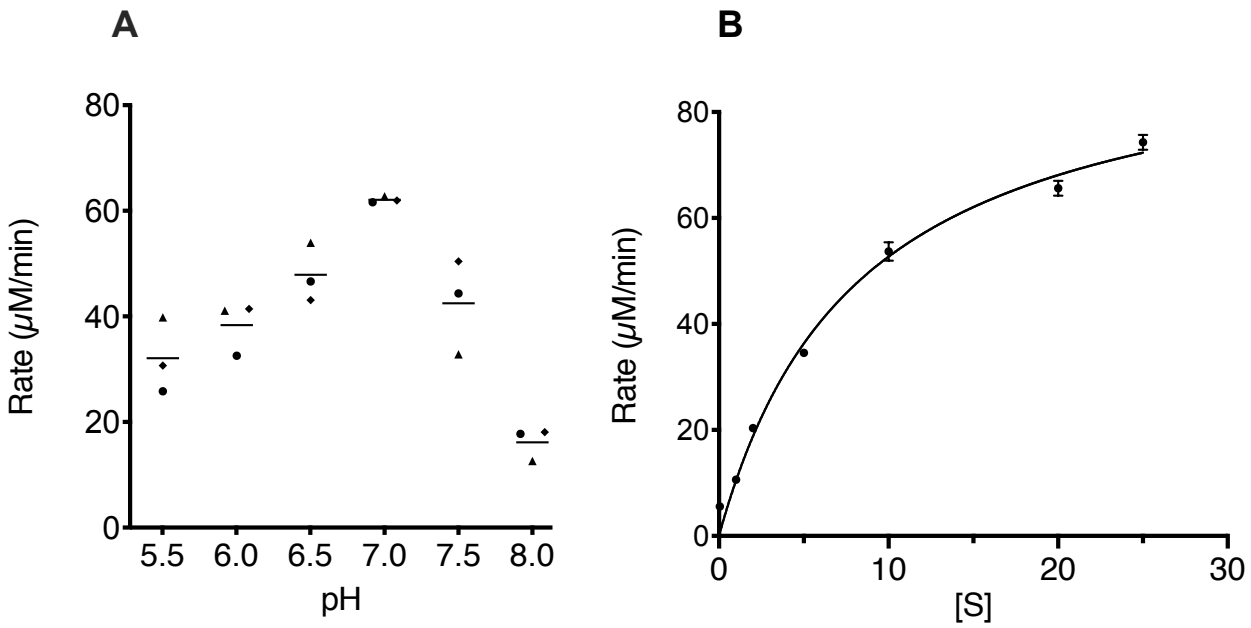
510 domains of GH13 subfamily 23 proteins from *Halomonas* sp. (BAL49684.1) and *Xanthomonas*
511 *campestris* (BAC87873.1), and subfamily 17 protein from *Culex quinquefasciatus* (ASO96882.1).
512 Red arrows indicate the conserved members of the catalytic triad, and a red circle indicates the
513 acid/base (Glu242) in BAL49684.1 and BAC87873.1. White circle indicates the predicted
514 acid/base in CG400_06090. Invariant sequences are highlighted in black.
515



516

517 **Figure 3.** Production and purification of α-glucosidase protein. (A) A protein of the predicted
518 mass of 62 kDa was observed in induced cultures. M: size marker, NP: *E. coli* with no vector, UN:
519 uninduced culture, IN: induced with IPTG. Numbers on the left indicate the marker size in kDa.
520 (B) Fractions from Ni-NTA affinity purification of His-tagged recombinant protein. M: size
521 marker, L: lysate, P: pellet, F: flow-through, W1: first wash, W2: second wash, E1: first elution,
522 E2: second elution.

523



524

525 **Figure 4.** (A) pH-rate profile of α-glucosidase. Each point represents the average of two technical
526 replicates for each of three independent experiments as indicated by different shapes. Horizontal
527 lines indicate the mean. (B) Michaelis-Menten plot. Results shown are the average of three
528 independent experiments, with error bars indicating standard deviation. The line represents the fit
529 to the Michaelis-Menten equation.

530

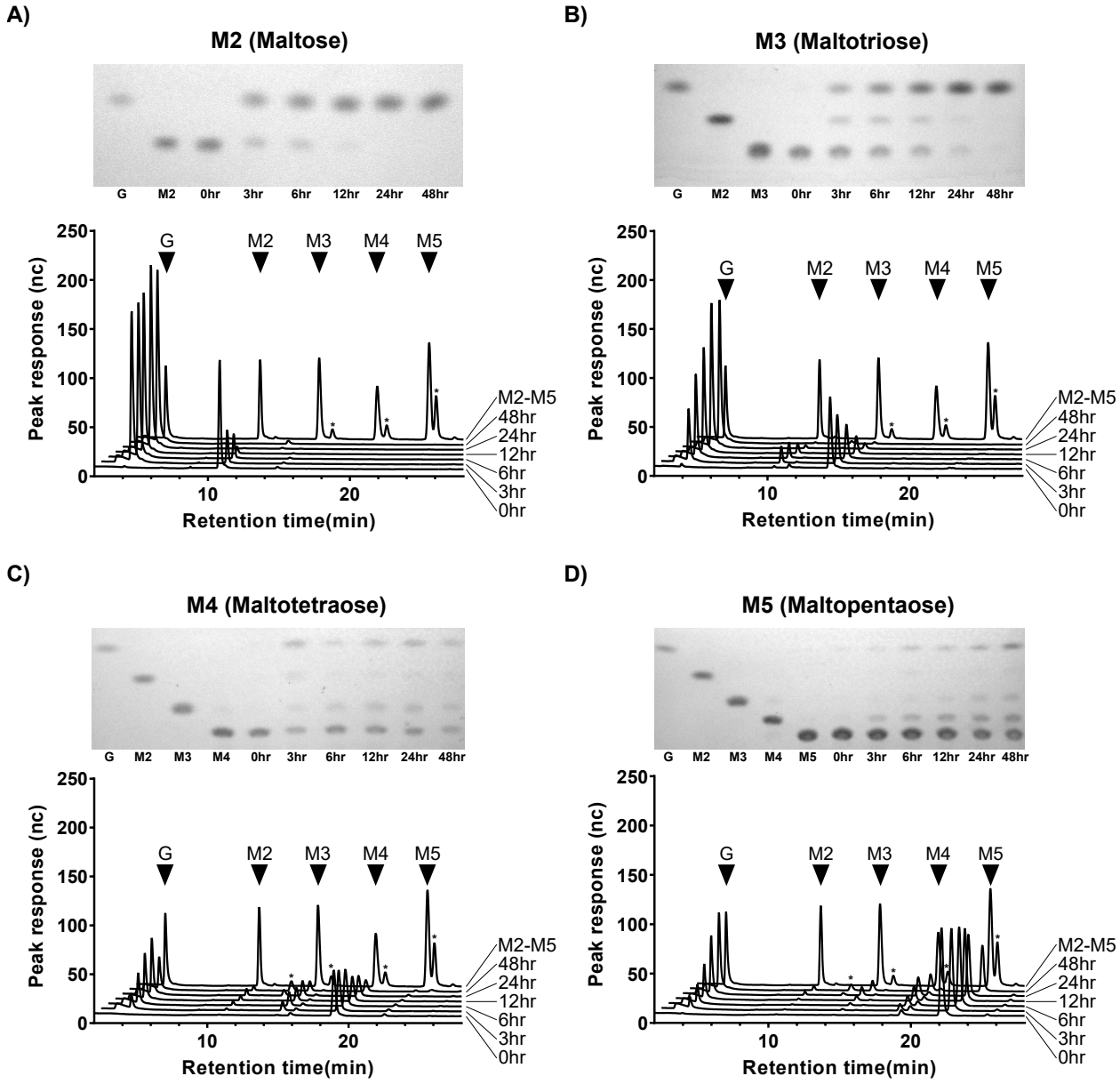


Figure 5. α-glucosidase CG400_06090 digests of maltose to maltopentaose (A-D). Each panel consists of a TLC plate and HPAEC-PAD trace of CG400_06090 digests of M2-M5 between 0hr and 48hr. Major peaks from M2-M5 standards are denoted with black triangles while stars represent unique, undefined oligosaccharides.

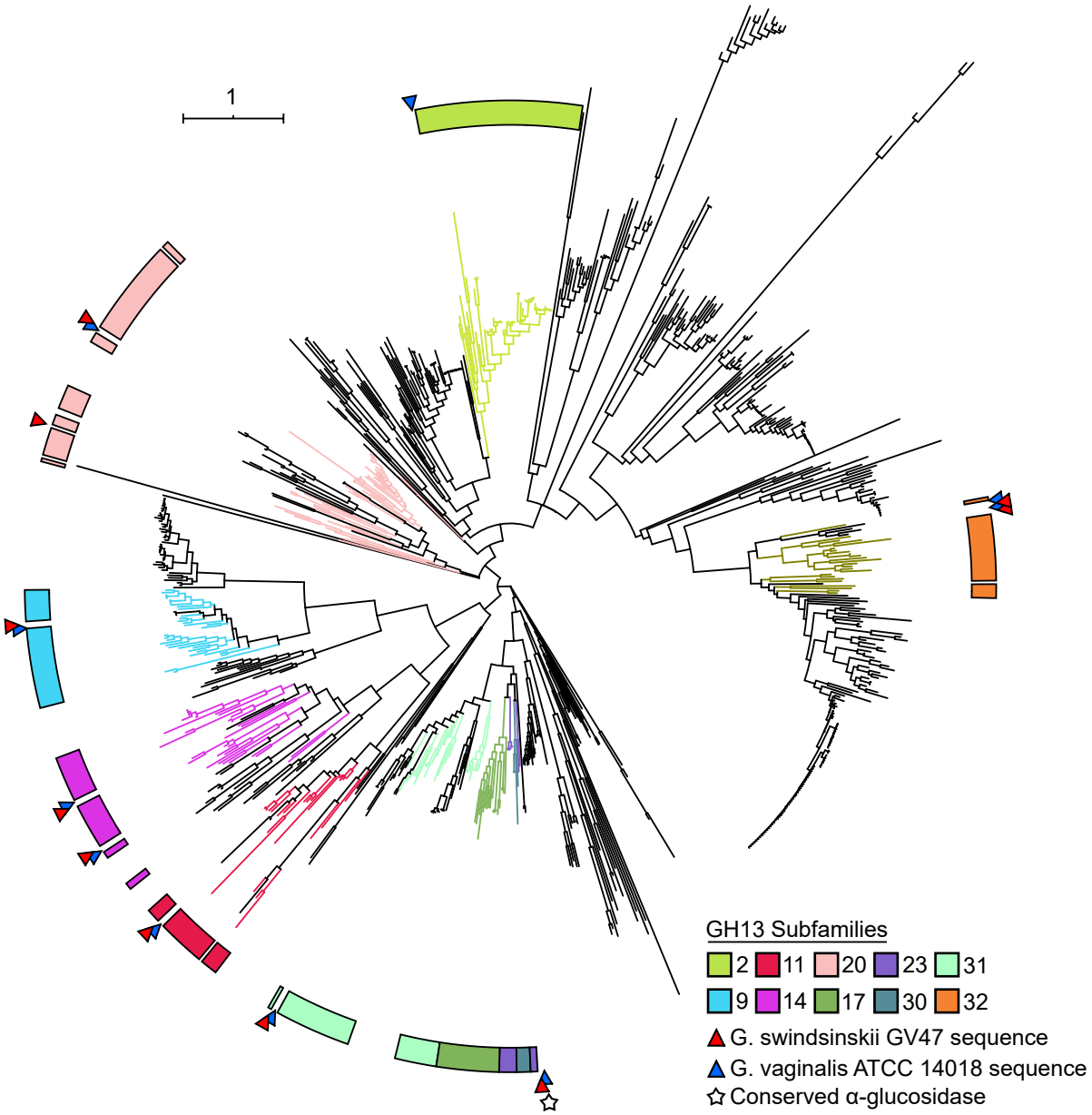


Figure S1. Phylogenetic tree of GH13 functionally characterized members with predicted GH13 domains from *G. swindsinskii* GV37 (red arrow) and *G. vaginalis* ATCC 14018 (blue arrow). Tree branches are coloured based on GH13 subfamily. The conserved α-glucosidase between the 26 proteomes is denoted by a white star. Trees were generated using SACCHARIS and viewed in iTOL**.**

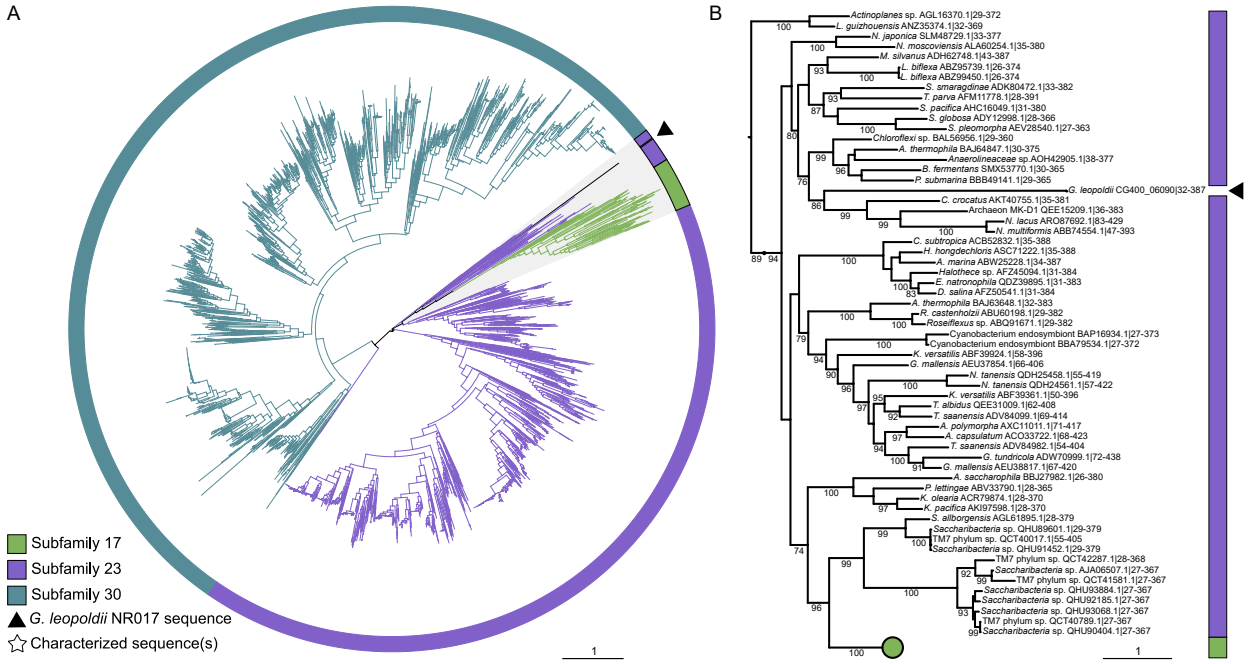
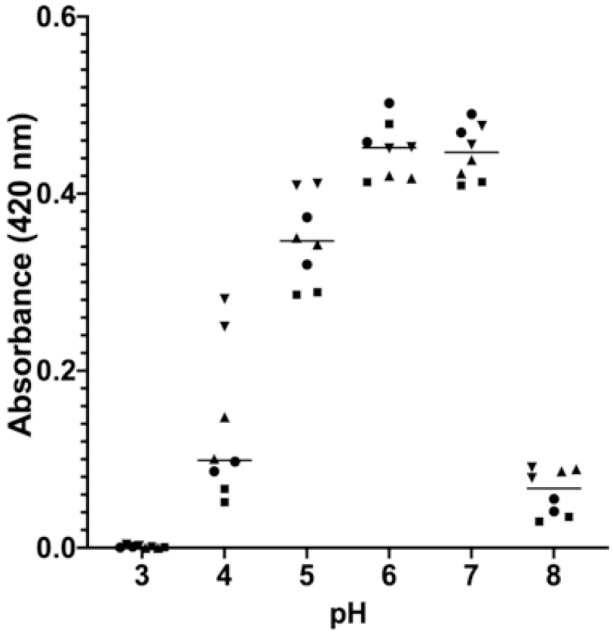


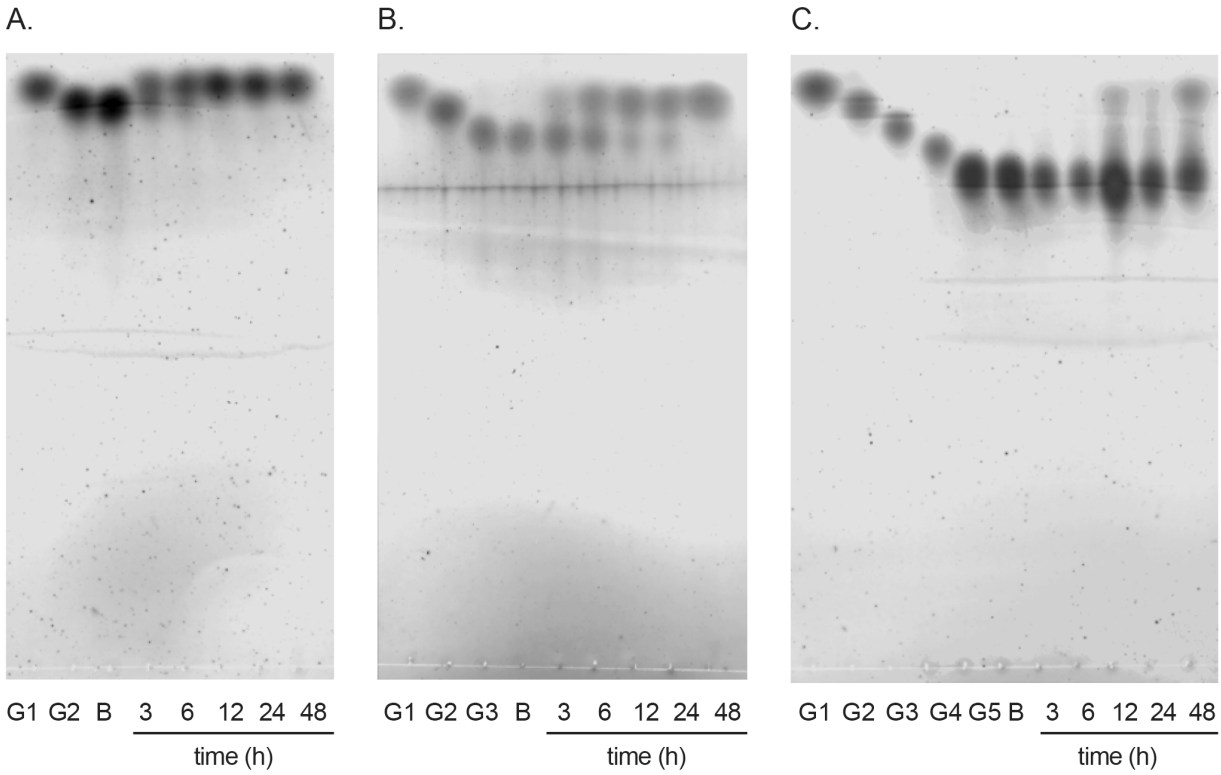
Figure S2. (A) Phylogenetic tree of all GH13 subfamily 17, 23 and 30 members in the CAZy database and *G. leopoldii* CG400_06090. Branch colour is based on GH13 subfamily and functionally characterized proteins are indicated with stars. The clade highlighted in grey including *G. leopoldii* CG400_06090 and its closest relatives is expanded in (B).



551

552 **Figure S3.** Release of 4-nitrophenol from chromogenic substrate 4-nitrophenyl-α-D-
553 glucopyranoside at pH 3-8. Results from four independent experiments each with two technical
554 replicates are shown. The α-glucosidase (0.8 mM) was incubated with 10 mM substrate at different
555 pH and amount of 4-nitrophenol released in 10 minutes was measured.

556



557

558 **Figure S4.** (A) TLC of products of glycogen hydrolysis by α-glucosidase enzyme. Reaction
559 mixtures were assessed at 3 h, 6 h, 12 h, 24 h and 48 h. N = Substrate with no enzyme. (B) TLC
560 of products hydrolysis by α-glucosidase enzyme of maltodextrins MD 4-7 (left panel), MD 13-17
561 (middle panel) and MD 16.5- 19.5 (right panel). Reaction mixtures were assessed at 3 h, 6 h, 12 h,
562 24 h and 48 h. N = Substrate with no enzyme.