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

Pashupati Bhandari[1], Jeffrey P. Tingley[2], D. Wade Abbott[2] and Janet E. Hill[1,*]

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

[2]Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, Alberta, T1J 4B1, Canada

*To whom correspondence should be addressed

Janet.Hill@usask.ca

Abstract

Gardnerella spp. in the vaginal microbiome are associated with bacterial vaginosis, a dysbiosis in which 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 by bacteria requires the combined activities of several enzymes collectively known as "amylases". In the present study, glycogen degrading abilities of *Gardnerella* spp. were assessed. We found that *Gardnerella* spp. isolates and filtered culture supernatants had amylase activity. Phylogenetic analyses predicted conserved Glycoside Hydrolase family 13 (GH13) members among *Gardnerella* spp. including a putative α-glucosidase. The gene for this enzyme was cloned and expressed, and recombinant protein was purified and functionally characterized. The enzyme was active on a variety of maltooligosaccharides over a broad pH range (4 - 8) with an optimum activity at pH 6-7. Glucose was released from maltose, maltotriose and maltopentose, however, no products were detected on thin layer chromatography (TLC) when the enzyme was incubated with glycogen. Our findings show that *Gardnerella* spp. produce a secreted α-glucosidase enzyme that can contribute to the complex and multistep process of glycogen breakdown by degrading smaller oligosaccharides into glucose, contributing to the pool of nutrients available to the vaginal microbiota.

37

38 **Introduction**

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

53 Colonization with multiple *Gardnerella* species is common, and different species
54 are dominant in different women (8). Understanding factors that contribute to differential
55 abundance is important since the species may differ in virulence (6) and they are variably
56 associated with clinical signs (8, 9). Several factors, including inter-specific competition,
57 biofilm formation and resistance to antimicrobials could contribute to the differential
58 abundance of the different species. Khan *et al.* showed that resource-based scramble
59 competition is frequent among *Gardnerella* subgroups (10). Glycogen is a significant

60 nutrient for vaginal microbiota, but previous reports on the growth of *Gardnerella* spp. on
61 glycogen containing medium are inconsistent (11). Species may differ in their ability to
62 digest glycogen and utilize the breakdown products, which may in turn contribute to
63 determining microbial community structure.

64 Glycogen is an energy storage molecule, which consists of linear chains of
65 approximately 13 glucose molecules covalently linked with α-1,4 glycosidic linkages, with
66 branches attached through α-1,6 glycosidic bonds (12, 13). A single glycogen molecule
67 consists of approximately 55,000 glucose residues with a molecular mass of ~ 10^7 kDa
68 (14). The size of glycogen particles can vary with source; from 10-44 nm in human skeletal
69 muscle to approximately 110-290 nm in human liver (15). Glycogen is deposited into the
70 vaginal lumen by epithelial cells under the influence of estrogen (16), and the concentration
71 of cell-free glycogen in vaginal fluid varies greatly, ranging from 0.1 - 32 μg/μl (17).
72 Digestion of this long polymeric molecule into smaller components is essential before it
73 can be taken up by bacterial cells (18, 19).

74 Glycogen digestion is accomplished by the coordinated action of enzymes that are
75 collectively described as "amylases" (20). α-glucosidase is an exo-acting enzyme that acts
76 on α-1,4 glycosidic bond from the non-reducing end to release glucose, while α-amylase is
77 an endo-acting amylase that cleaves α-1,4 glycosidic bonds to produce α-limit dextrins and
78 maltose. Debranching enzymes like pullulanase and isoamylase break α-1,6 glycosidic
79 bonds (15, 21). Glycogen degradation mechanisms active in the vaginal microbiome are
80 not completely described. Although it has been demonstrated that vaginal secretions
81 exhibit amylase activity that can degrade glycogen and the generation of these products
82 enable growth of vaginal lactobacilli (19), the role of bacterial enzymes in this process is

83 unknown and there is limited information about glycogen degradation by clinically
84 important bacteria such as *Gardnerella* spp..
85 The objective of the current study was to assess amylase activity of *Gardnerella*
86 spp. and to characterize the activity of a putative amylase protein conserved in all
87 *Gardnerella* spp..
88
89 **Methods**
90
91 ***Gardnerella* isolates**
92 Isolates of *Gardnerella* used in this study were from a previously described culture
93 collection (6) and included representatives of *cpn*60 subgroup A (*G. leopoldii* (n = 4), *G.*
94 *swidsinskii* (n = 6), other (n = 1)), subgroup B (*G. piotii* (n = 5), Genome sp. 3 (n = 2)),
95 subgroup C (*G. vaginalis* (n = 6)), and subgroup D (Genome sp. 8 (n = 1), Genome sp. 9
96 (n = 1)). Whole genome sequences of these isolates had been determined previously
97 (BioProject Accessions PRJNA394757) and annotated by the NCBI Prokaryotic Genome
98 Annotation Pipeline (PGAP) (22).
99
100 **Amylase activity assay**
101 Amylase activity of *Gardnerella* isolates was assessed using a starch iodine test (23).
102 Isolates were revived from -80°C storage on Columbia blood agar plates with 5% sheep
103 blood, and incubated anaerobically using the GasPak system (BD, USA) at 37°C for 48
104 hours. Isolated colonies from blood agar plates were transferred in 0.85% saline using a
105 sterile swab and cell suspensions were adjusted to McFarland 1.0. The cell suspension (10

106 μL) was spotted on to BHI agar plates containing 1% (w/v) potato starch (Difco) or bovine
107 liver glycogen (Sigma-Aldrich), and plates were incubated anaerobically at 37°C for 24
108 hours. Lugol's iodine solution (1% w/v) was added to the plates to detect evidence of
109 amylase activity indicated by a clear halo around the bacterial colony.

110 To determine amylase activity of culture supernatant, overnight broth cultures were
111 centrifuged at 10,000 × *g* for 10 mins and the cell-free supernatant was passed through a
112 0.45 μm filter. Filtered supernatant (10 μL) was spotted on to a 1% (w/v) starch agar plate
113 and incubated anaerobically at 37°C for 24 hours. Plates were flooded with 1% (w/v)
114 Lugol's iodine solution to detect evidence of amylase activity. α-amylase from *Bacillus*
115 *licheniformis* (0.05 mg/ml) (Sigma-Aldrich, Oakville, ON, CAT No. A455) was used as a
116 positive control.

117

118 **Identification of putative secreted amylase sequences**

119 The predicted proteomes of the 26 *Gardnerella* isolates used in this study were examined
120 with Phobius v1.01 (24) to identify cytoplasmic and non-cytoplasmic proteins based on the
121 detection of signal peptides. Within the predicted secreted proteins, putative amylase
122 enzymes were identified by annotation terms using PGAP annotation. Multiple sequence
123 alignments of predicted secreted amylase sequences were performed using CLUSTALw
124 with results viewed and edited in AliView (version 1.18) (25) prior to phylogenetic tree
125 building using PHYLIP (26).

126

127 **Functional prediction of *Gardnerella* spp. amylases**

128 Putative amylolytic enzymes in *Gardnerella* spp. genomes were identified using dbCAN,
129 and predicted sequences were run through the SACCHARIS pipeline (27). SACCHARIS
130 combines user sequences with sequences from the CAZy database and trims sequences to
131 the catalytic domain using dbCAN2 (28). Domain sequences were aligned with MUSCLE
132 (29), and a best-fit model was generated with ProtTest (30). Final trees were generated
133 with FastTree (31) and visualized with iTOL (32). Alignment of CG400_06090 from *G.*
134 *leopoldii* NR017 to orthologues from *Halomonas* sp. (BAL49684.1), *Xanthomonas*
135 *campestris* (BAC87873.1), and *Culex quinquefasciatus* (ASO96882.1) was done using
136 CLUSTALw and visualized in ESPript (33). A predicted structure model of CG400_06090
137 was aligned with *Halomonas* sp. HaG (BAL49684.1, PDB accession 3WY1) (34) using
138 PHYRE2 (35). BLASTp (36) alignment of CAZy derived sequences from reference
139 genomes to a database of the predicted proteomes of 26 study isolates was conducted to
140 identify orthologs.
141
142 **Expression and purification of the CG400_06090 gene product**
143 Genomic DNA from *G. leopoldi* NR017 was extracted using a modified salting out
144 procedure, and the CG400_06090 open reading frame (locus tag CG400_06090 in
145 GenBank accession NNRZ01000007, protein accession RFT33048) was PCR amplified
146 with primers JH0729-F (5'ATG CAT GCG CAT TAT ACG ATC ATG CTC-3') and
147 JH0730-R (5'ATG GTA CCT TAC ATT CCA AAC ACT GCA-3'). Underlined sequences
148 indicate *Sph*I and *Kpn*I restriction enzyme sites. PCR reaction contained 1 × PCR reaction
149 buffer (0.2 M Tris-HCl pH 8.4, 0.5 M KCl), 200 µM each dNTPs, 2.5 mM $MgCl_2$, 400 nM
150 each primer, 1 U/reaction of Platinum Taq DNA polymerase high-fidelity (5 units/µL in

151 50% glycerol, 20 mM Tris-HCl, 40 mM NaCl, 0.1 mM EDTA, and stabilizers) (Life-
152 technologies) and 2 µl of template DNA, in a final volume of 50 µl. PCR was performed
153 with following parameters: initial denaturation at 94°C for 3 minutes followed by
154 (denaturation at 94°C, 15 seconds; annealing at 60°C, 15 seconds and extension at 72°C, 2
155 minutes; 35 cycles), and final extension at 72°C for 5 minutes. Purified PCR products were
156 digested with *Kpn*I and *Sph*I and ligated into expression vector pQE-80 L (Qiagen,
157 Mississauga, ON) digested with same restriction endonucleases. The resulting recombinant
158 plasmid was used to transform One Shot TOP10 chemically competent *E. coli* cells
159 (Invitrogen, Carlsbad, California). Colony PCR was performed to identify transformants
160 containing vector with insert. Insertion of the putative amylase gene in-frame with the N-
161 terminal 6×Histidine tag was confirmed by sequencing of the purified plasmid.

162 *E. coli* containing the plasmid were grown overnight at 37°C in LB medium with
163 100 µg/ml ampicillin. Overnight culture was diluted 1:60 in fresh medium to a final volume
164 of 50 ml, and the culture was incubated at 37°C with shaking at 225 rpm until it reached
165 an OD_{600} of 0.6. At this point, expression was induced with 0.1 mM IPTG and a further 4
166 hours of incubation was done at 37°C at 225 rpm. Cells were harvested by centrifugation
167 at 10,000 × *g* for 30 mins and the pellet was resuspended in lysis buffer (50 mM NaH_2PO_4,
168 300 mM NaCl, 10 mM imidazole, pH 8.0). Lysozyme was added to 1 mg/ml and cells were
169 lysed by sonication (8 mins total run time with 15 sec on and 30 sec off). The lysate was
170 clarified by centrifugation at 10,000 × *g* for 30 mins at 4°C and the supernatant was applied
171 to an Ni-NTA affinity column according to the manufacturer's instructions (Qiagen,
172 Germany). Bound proteins were washed twice with wash buffer (50 mM NaH_2PO_4, 300
173 mM NaCl, 20 mM imidazole, pH 8.0) and eluted with elution buffer (50 mM NaH_2PO_4,

174 300 mM NaCl, 250 mM imidazole, pH 8.0). Eluted protein was buffer exchanged with the
175 same elution buffer but without imidazole using 30K MWCO protein concentrator
176 (ThermoFisher Scientific). Final protein concentration was measured by
177 spectrophotometry.
178
179 **Measuring enzyme activity**
180 Purified protein was tested for α-glucosidase activity by measuring the release of 4-
181 Nitrophenol from a chromogenic substrate 4-Nitrophenyl-α-D-glucopyranoside as
182 described elsewhere (37). Briefly, 25 µl of enzyme solution (50 µg/ml) was added to 25 µl
183 of 10 mM 4-Nitrophenylα-D-glucopyranoside substrate and the reaction mixture was
184 incubated at 37°C for 10 mins. Reactions were stopped by the addition of 100 µl of 1M
185 Na_2CO_3 solution and OD_{420} was measured. To determine activity at different pH values,
186 substrate was prepared in 50 mM citric acid-sodium citrate buffer (pH 3.0) or 50 mM
187 sodium phosphate buffer (pH 4.0-8.0).
188
189 **Enzymatic activity on different substrates**
190 The activity of the purified enzyme was examined with different substrates according to
191 Cockburn et *al.* (38). Briefly, 10 µl of enzyme solution (5.0 mg/ml) was added to 190 µl
192 of 0.1% (w/v) maltose, maltotriose, maltopentose, maltodextrins (MD 4-7, MD 13-17 and
193 MD 16.5-19.5) or glycogen in 50 mM sodium phosphate buffer, pH 6.0, and reaction
194 mixture was incubated at 37°C. Aliquots of the reaction mixture (10 µl) were taken out at
195 various time intervals of incubation and were frozen immediately until thin layer
196 chromatography (TLC) was performed. Standards and the reaction mixtures (0.5 µl) were

197 loaded on 0.25 mm silica gel plates (Sigma-Aldrich, Oakville, ON) and the chromatogram
198 was developed in a solvent system containing acetonitrile, ethyl acetate, 2-propanol and
199 water in (8.5:2:5:6 vol/vol) in a developing chamber until the mobile phase reached the top
200 of the plates. The plates were allowed to dry in a fume hood and products were resolved
201 by dipping the plates into the 0.3% solution of 1-napthyl ethylene diamine dihydrochloride
202 followed by drying and heating with the heat gun until spots became visible.
203
204 **Results**
205
206 **Amylase activity of *Gardnerella* spp.**
207 All 26 isolates tested showed amylase activity by degrading starch and glycogen as
208 indicated by the appearance of a clear zone around bacterial colonies after iodine staining.
209 A filter sterilized culture supernatant from *G. leopoldii* (NR017) was also able to digest the
210 starch and glycogen in the agar plate assay, indicating that amylase enzymes are secreted
211 (Figure 1).
212
213 **Identification of sequences of secreted amylases**
214 A total of 2,525 predicted proteins from the combined proteomes of the 26 study isolates
215 were determined to contain a signal peptide and predicted to be non-cytoplasmic. Of these
216 putative secreted proteins, 60 were annotated as amylases. After removal of severely
217 truncated sequences, a multiple sequence alignment of the remaining 42 sequences was
218 trimmed to the uniform length of 290 amino acids. Phylogenetic analysis of these predicted
219 extracellular amylase sequences revealed five clusters, but only one of these clusters

220 contained orthologous sequences from all 26 is1olates (Figure 2). Sequence identities
221 within this cluster were all 91-100%. This conserved sequence (corresponding to GenBank
222 accession WP_004104790) was annotated as an GH13 α-amylase (EC 3.2.1.1) in all
223 genomes, suggesting it was an endo-acting enzyme that releases maltooligosaccharides
224 from glycogen.

225 The GH13 family in the CAZy database is a large, polyspecific family targeting α-
226 glucose linkages. The functional range of members in this family make annotation of new
227 sequences challenging. In order to predict the activity of the conserved protein identified
228 among the study isolates, we employed SACCHARIS, which combines CAZyme family
229 trees generated from biochemically characterized proteins with related sequences of
230 unknown function. Currently, two *Gardnerella* reference genomes are available in CAZy:
231 *G. vaginalis* ATCC 14018 and *G. swidsinkii* GV37. GH13 family proteins encoded in these
232 *Gardnerella* genomes were identified and a family tree including all characterized GH13
233 sequences in the CAZy database was generated (Figure S1). SACCHARIS annotation of
234 the GH13 family in *Gardnerella* resulted in identification of protein domains belonging to
235 ten subfamilies. Alignment with characterized GH13 enzymes from the CAZy database
236 (39), suggested that the conserved amylase identified in the study isolates was an α-
237 glucosidase (EC 3.2.1.20), closely related to subfamilies 23 and 30 of GH13, and not an α-
238 amylase as annotated in GenBank.

239

240 **Relationship to other α-glucosidases**

241 A representative sequence of the predicted α-glucosidase was selected from *G.*
242 *leopoldii* (NR017). This sequence (CG400_06090) was combined with functionally

243 characterized members of the GH13 CAZy database using SACCHARIS (Figure 3A).
244 CG400_06090 partitioned with members of GH13 subfamilies 17, 23 and 30, of which the
245 prominent activity is α-glucosidase. This clade was expanded to better illustrate the
246 relationship of CG400_06090 with related characterized sequences as an apparently deep-
247 branching member of subfamily 23, although with low bootstrap support (Figure 3B).
248 When sequences of additional (not functionally characterized) members were included,
249 CG400_06090 clustered within subfamily 23 (Figure S2).

250 To identify conserved catalytic residues between CG400_06090 and characterized
251 members of GH13, CG400_06090 was aligned with structurally characterized members of
252 subfamily 23 (BAL49684.1 and BAC87873.1) and subfamily 17 (ASO96882.1) (Figure
253 3C). Of the GH13 catalytic triad, the aspartate nucleophile and the aspartate that stabilizes
254 the transition state are conserved among the four sequences (40). The glutamate general
255 acid/base, however, does not appear to be sequence conserved between CG400_06090 and
256 the GH13 subfamily 23 members. Upon modelling CG400_06090 in PHYRE2 and
257 aligning with BAL49684.1 (PDBID 3WY1), CG400_06090 Glu256 did appear to be
258 spatially conserved with the general acid/base (not shown).

259

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

261 The *G. leopoldi* NR017 CG400_06090 open reading frame comprises 1701 base pairs
262 encoding a 567 amino acid protein, with a predicted mass of 62 kDa. The region of the
263 open reading frame encoding amino acids 129-418 was PCR amplified and ligated into
264 vector pQE-80L for expression in *E. coli* as an N-terminal 6xHis-tagged protein. A distinct
265 protein band in SDS-PAGE between 50 kDa and 75 kDa was obtained from IPTG induced

266 *E. coli* cells compared with non-induced cells, indicating the expression of the α-
267 glucosidase protein (Figure 4A). The recombinant protein was soluble, and was purified
268 using a Ni-NTA spin column (Figure 4B). From a 50 ml broth culture, approximately 700
269 µl of purified protein at 4.0 mg/ml was obtained.
270
271 **Effect of pH on enzyme activity**
272 α-glucosidase activity of the purified protein was demonstrated by release of 4-Nitrophenol
273 from chromogenic substrate 4-Nitrophenyl-α-D-glucopyranoside. This activity was
274 observed across a broad pH range (4.0 – 8.0), with maximum activity at pH 6.0 and 7.0
275 while no activity was seen at pH 3.0 (Figure 5).
276
277 **Analysis of substrate hydrolysis**
278 Production of glucose was detected when the purified α-glucosidase was incubated with
279 maltose, maltotriose and maltopentose at pH 6.0 (Figure 6). No appreciable activity,
280 however, was detected on maltodextrins (MD 4-7, MD 13-17 and MD 16.5-19.5) or
281 glycogen.
282
283 **Discussion**
284 Glycogen is a major available nutrient to vaginal microbiota and its utilization
285 likely plays an important role in the survival and success of *Gardnerella* spp. in the vaginal
286 microbiome. Glycogen is digested into smaller products, such as maltose and
287 maltodextrins, by a group of enzymes collectively known as "amylases". In this study, we

288 assessed the amylase activity of *Gardnerella* spp. and an α-glucosidase gene conserved in
289 *Gardnerella* spp. was expressed and its product was characterized.

290 Isolates of *G. leopoldii*, *G. swidsinskii*, *G. piotii* and *G. vaginalis* all showed
291 amylase activity by degrading starch and glycogen in an agar plate assay. Previously, Piot
292 *et al.* examined the starch fermentation ability of 175 *Gardnerella* isolates on Mueller
293 Hinton agar supplemented with horse serum and found that all *Gardnerella* strains were
294 capable of hydrolysing starch (41). They did not, however, assess glycogen degradation
295 and further, it is not clear if serum amylase contained in the media contributed to the
296 observed amylase activity. Robinson, in his review, reported that starch can be utilized by
297 *Gardnerella* (42) while Catlin in another review, reported that glycogen utilization is
298 inconsistent in *Gardnerella* (11). All *Gardnerella* spp. assessed in our study were amylase
299 positive, which suggests it is a conserved catalytic function. As these bacteria are found
300 almost exclusively in the glycogen-rich human vagina, it is not surprising that they produce
301 amylases to break down this substrate.

302 In the phylogenetic analysis of putative secreted amylase sequences, one cluster
303 contained sequences common to all 26 isolates. Although this protein sequence was
304 annotated as an α-amylase, suggesting it would produce maltooligosaccahrides from
305 glycogen, SACCHARIS predicted it to be closely related to α-glucosidases, which cleave
306 glucose from the non-reducing end of its substrate. SACCHARIS predicts enzyme activity
307 by comparing sequence domains to sequences deposited in the curated CAZy database for
308 which structural and / or biochemical information is available (27). Automated annotation
309 of sequences deposited in primary sequence databases, such as GenBank has resulted in
310 significant problems with functional mis-annotation (43). This problem is exacerbated by

311 the increasing rate at which whole genome sequences are being generated and deposited.
312 The genome sequences of the study isolates were annotated using the NCBI Prokaryotic
313 Genome Annotation Pipeline, which identifies genes and annotates largely based on
314 sequence similarity to previously annotated sequences, thus potentially propagating
315 incorrect functional annotations.

316 In the phylogenetic analysis of GH13 family enzymes, the conserved α-glucosidase
317 from *Gardnerella* spp. appears to be most closely related to GH13 subfamily 17, 23 and
318 30 characterized members, forming a monophyletic group. However, low bootstrap values
319 between CG400_06090 and other clade members make assignment of CG400_06090 to a
320 defined subfamily difficult based on comparison to sequences of functionally characterized
321 proteins (Figure 3B). Additional support for assignment to subfamily 23 was observed
322 when additional sequences of proteins lacking functional characterization from subfamilies
323 17, 23, and 30 were included (Figure S2). Regardless, the α-glucosidase activity observed
324 within this clade does extend to CG400_06090. Primary sequences of GH13 members of
325 amylases have seven highly conserved regions and three amino acids that form the catalytic
326 triad (21). Comparison of the primary sequence and three-dimensional model of
327 CG400_06090 with characterized α-glucosidases from GH13 subfamily 17 and 23 did
328 show sequence conservation of nucleophile and spatial conservation of the catalytic triad.

329 Our results demonstrate the value of using SACCHARIS to guide discovery of
330 CAZymes based on comparison at a catalytic domain level to functionally characterized
331 enzymes. All of the other proteins identified as putative secreted amylases were also
332 annotated as "alpha-amylase" in the GenBank records. Further investigation will be

333 required to demonstrate the actual functions of these proteins, and how they contribute to
334 carbohydrate degrading activities of *Gardnerella* spp..

335 Amylase enzymes that contribute to glycogen digestion in the vaginal environment
336 are not well studied. Spear *et al*. (19) suggested a role for a human amylase enzyme in the
337 breakdown of vaginal glycogen. They detected host-derived pancreatic α-amylase and
338 acidic α-glucosidase in genital fluid samples collected from women and showed that genital
339 fluid containing these enzymes can degrade the glycogen. The presence of these enzyme
340 activities alone in vaginal fluid, however, was not correlated with glycogen degradation,
341 suggesting contributions of activity from the resident microbiota. Many vaginal bacteria
342 likely produce amylases that contribute to this activity as various anaerobic taxa are known
343 to encode amylases and digest glycogen (44), however, the contributions of the vaginal
344 microbiota to this process and the importance of these processes to vaginal microbial
345 community dynamics are not yet well understood.

346 α-glucosidase enzymes are typically active over a broad pH range and exhibit
347 optimal activity at pH 4.5-6.0 (45). They have also been characterized in other bacteria
348 isolated from human body sites, particularly the intestinal tract. An extracellular α-
349 glucosidase form *Lactobacillus acidophilus* NCTC 1723 was reported to have activity at
350 pH 6.0-7.5 (46). Van den Broek *et al.* characterized an α-glucosidase (*algB*) in
351 *Bifidobacterium adolescentis* DSM20083 and reported its maximum activity at pH 6.8.
352 This protein had a molecular mass of 73 kDa and used maltose and maltotriose as substrates
353 (47). Purified α-glucosidase from *G. leopoldii* NR017 was active across a broad pH range
354 (4.0 – 8.0). By definition, the pH of the vaginal fluid in women with bacterial vaginosis is
355 > 4.5 (48), and has been reported as high as 7.0 - 7.5 (49), so it is not surprising for

356 *Gardnerella* to produce a secreted enzyme that functions optimally at pH 6-7. In fact, this
357 activity at high pH could help *Gardnerella* spp. liberate glucose at an increased rate in
358 dysbiotic microbiomes, thus increasing the nutrient pool available and contributing to
359 supporting the overgrowth of vaginosis-associated bacteria.

360 Complete glycogen digestion is a complex and multistep process that requires the
361 action of several different enzymes at the different stages. Protein sequences belonging to
362 eight different GH13 subfamilies, which could have roles in other steps of the glycogen
363 degradation, were identified in *G. vaginalis* ATCC 14018 and *G. swidsinskii* GV37
364 (Supplemental Figure 1), and multiple secreted amylase-like proteins were detected in the
365 proteomes of additional *Gardnerella* spp. (Figure 2). Functional characterization of these
366 proteins will be required to understand their roles in glycogen degradation.

367 Our findings show that *Gardnerella* spp. have a secreted α-glucosidase enzyme that
368 likely contributes to the complex and multistep process of glycogen breakdown by
369 releasing glucose from oligosaccharides, contributing to the pool of nutrients available to
370 the vaginal microbiota. Identification and biochemical characterization of additional
371 enzymes involved in glycogen digestion will provide insight into whether utilization of this
372 abundant carbon source is an important factor in population dynamics and competition
373 among *Gardnerella* spp..

374

375 **Funding**

376 This research is supported by an NSERC Discovery grant to JEH, and Agriculture and
377 Agri-Food Canada project number: J-001589. PB is supported by a Devolved Scholarship
378 from the University of Saskatchewan.

Conflicts of interest

The authors declare no competing interest.

Acknowledgements

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

389 **References**

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

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

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

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

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

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

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

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

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

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

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

428 12. Meléndez-Hevia E, Waddell TG, Shelton ED. 1993. Optimization of molecular
429 design in the evolution of metabolism: the glycogen molecule. Biochem J 295 (Pt
430 2):477–483.

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

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

435 15. Adeva-Andany MM, González-Lucán M, Donapetry-García C, Fernández-Fernández
436 C, Ameneiros-Rodríguez E. 2016. Glycogen metabolism in humans. BBA Clin
437 5:85–100.

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

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

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

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

447 in lower-genital-tract mucosal fluid processes glycogen to support vaginal
448 colonization by Lactobacillus. J Infect Dis 210:1019–1028.

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

451 21. Mehta D, Satyanarayana T. 2016. Bacterial and archaeal α-amylases: Diversity and
452 amelioration of the desirable characteristics for industrial applications. Front
453 Microbiol 7.

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

457 23. Ryan SM, Fitzgerald GF, van Sinderen D. 2006. Screening for and identification of
458 Starch-, Amylopectin-, and Pullulan-degrading activities in Bifidobacterial Strains.
459 Appl Environ Microbiol 72:5289.

460 24. Lukas Käll, Anders Krogh, Erik L.L. Sonnhammer. 2004. A combined
461 transmembrane topology and signal peptide prediction method. J Mol Biol
462 338:1027–1036.

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

465 26. Felsenstein J. 1989. PHYLIP - phylogeny inference package (version 3.2). Cladistics
466 5:164–166.

467 27. Jones DR, Thomas D, Alger N, Ghavidel A, Inglis GD, Abbott DW. 2018.
468 SACCHARIS: an automated pipeline to streamline discovery of carbohydrate active
469 enzyme activities within polyspecific families and de novo sequence datasets.
470 Biotechnol Biofuels 11:27.

471 28. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y.
472 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme
473 annotation. Nucleic Acids Res 46:W95–W101.

474 29. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced
475 time and space complexity. BMC Bioinformatics 5:113.

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

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

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

482 33. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new
483 ENDscript server. Nucleic Acids Res2014/04/21. 42:W320–W324.

484 34. Shen X, Saburi W, Gai Z, Kato K, Ojima-Kato T, Yu J, Komoda K, Kido Y, Matsui
485 H, Mori H, Yao M. 2015. Structural analysis of the α-glucosidase HaG provides

486 new insights into substrate specificity and catalytic mechanism. Acta Crystallogr
487 Sect D 71:1382–1391.

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

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

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

495 38. Cockburn D, Koropatkin N. 2015. Product analysis of starch active enzymes by TLC.
496 Bioprotocol 5:1–8.

497 39. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The
498 carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res
499 42:D490–D495.

500 40. Uitdehaag JCM, Mosi R, Kalk KH, van der Veen BA, Dijkhuizen L, Withers SG,
501 Dijkstra BW. 1999. X-ray structures along the reaction pathway of cyclodextrin
502 glycosyltransferase elucidate catalysis in the α-amylase family. Nat Struct Biol
503 6:432–436.

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

506 42. Taylor-Robinson D. 1984. The bacteriology of *Gardnerella vaginalis*. Scand J Urol
507 Nephrol Suppl 86:41.

508 43. Schnoes AM, Brown SD, Dodevski I, Babbitt PC. 2009. Annotation error in public
509 databases: misannotation of molecular function in enzyme superfamilies. PLoS
510 Comput Biol 5:e1000605–e1000605.

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

513 45. Del Moral S, Barradas-Dermitz DM, Aguilar-Uscanga MG. 2018. Production and
514 biochemical characterization of α-glucosidase from *Aspergillus niger* ITV-01
515 isolated from sugar cane bagasse. 3 Biotech 8:7–7.

516 46. Chan K, Li KB. 1981. Properties of alpha-glucosidase from *Lactobacillus*
517 *acidophilus* NCTC 1723. Microbios 32:47–61.

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

521 48. Amsel R, Totten PA, Spiegel CA, Chen KC, Eschenbach D, Holmes KK. 1983.
522 Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic
523 associations. Am J Med 74:14–22.

524 49. O'Hanlon DE, Moench TR, Cone RA. 2013. Vaginal pH and microbicidal lactic acid
525 when lactobacilli dominate the microbiota. PloS One 8:e80074–e80074.

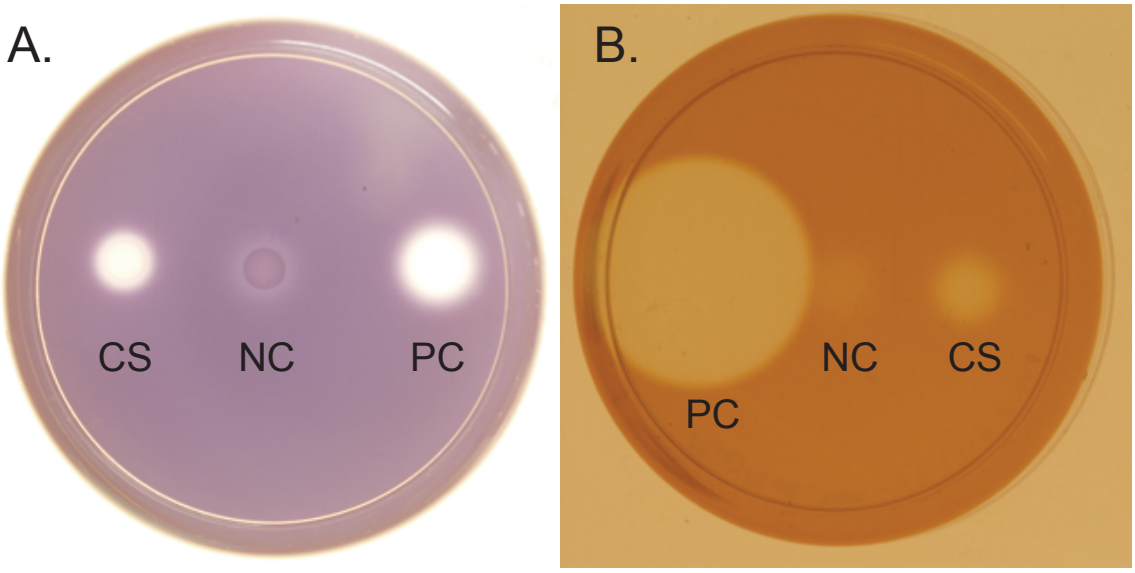


Figure 1. Amylase activity of filtered culture supernatant (CS) from *G. leopoldii* NR017 on starch (A) and glycogen (B) agar. PC: positive control (amylase enzyme from *B. licheniformis*, 0.05 mg/ml) and NC: negative control, BHI broth.

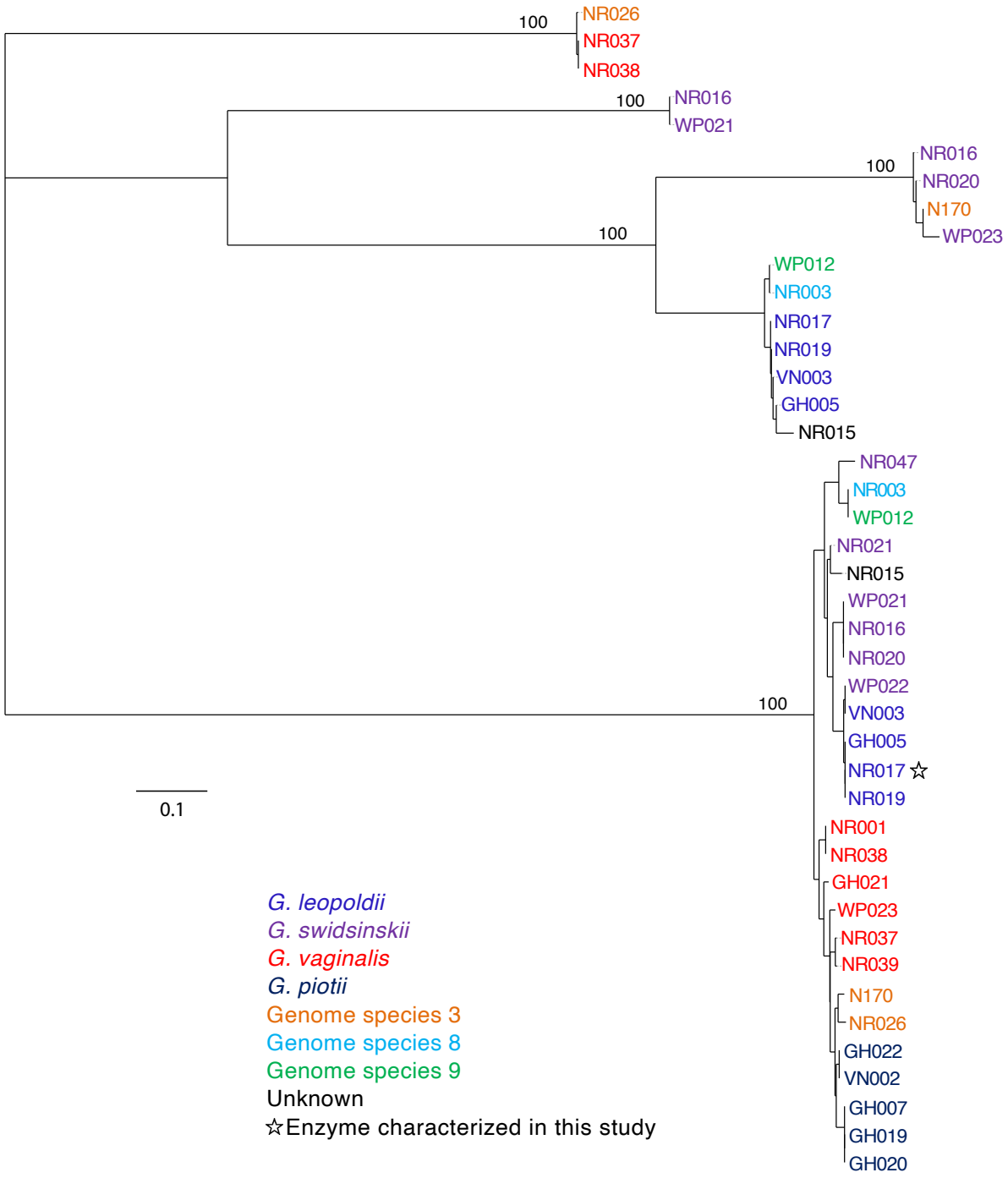


Figure 2. Phylogenetic analysis of predicted extracellular amylases protein sequences from 26 *Gardnerella* genomes. Trees are consensus trees of 100 bootstrap iterations. Bootstrap values are shown for major branch points. Species is indicated by label colour according to the legend.

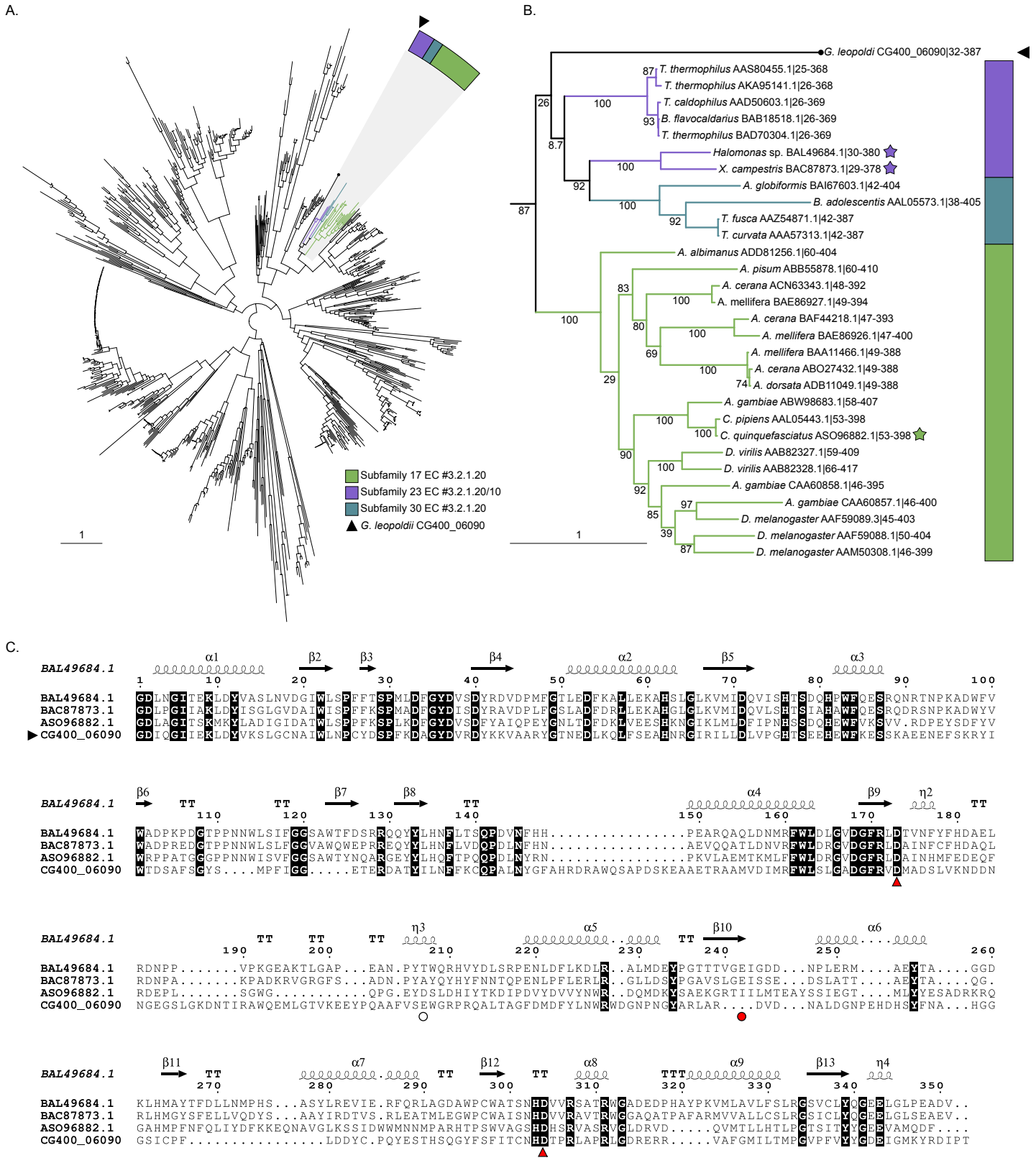


Figure 3. (A) Phylogenetic tree of characterized GH13 members in the CAZy database and predicted *G. leopoldii* NR017 GH13 sequence CG400_06090. (B) Expanded tree of subfamilies

17, 23 and 30. Stars indicate structurally characterized sequences and all bootstrap values were included. (C) CLUSTAL alignment of the catalytic domain of CG400_06090 with catalytic domains of GH13 subfamily 23 proteins from *Halomonas* sp. (BAL49684.1) and *Xanthomonas campestris* (BAC87873.1), and subfamily 17 protein from *Culex quinquefasciatus* (ASO96882.1). Red arrows indicate the conserved members of the catalytic triad, and a red circle indicates the acid/base (Glu242) in BAL49684.1 and BAC87873.1. White circle indicates the predicted acid/base in CG400_06090. Invariant sequences are highlighted in black.

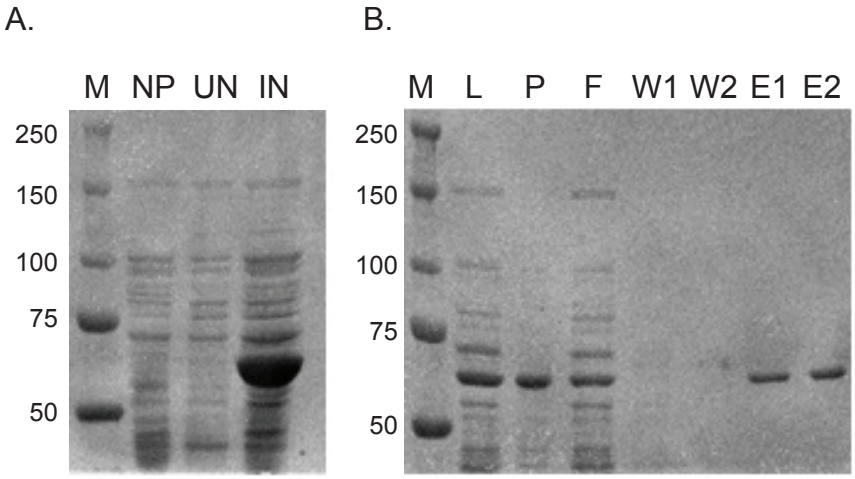


Figure 4. Production and purification of α-glucosidase protein. (A) A protein of the predicted mass of 62 kDa was observed in induced cultures. M: size marker, NP: *E. coli* with no vector, UN: uninduced culture, IN: induced with IPTG. Numbers on the left indicate the marker size in kDa. (B) Fractions from Ni-NTA affinity purification of His-tagged recombinant protein. M: size marker, L: lysate, P: pellet, F: flow-through, W1: first wash, W2: second wash, E1: first elution, E2: second elution.

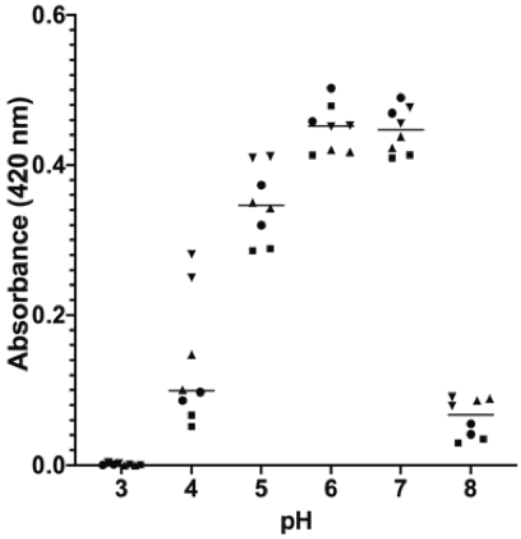


Figure 5. pH profile of α-glucosidase enzyme. The enzyme was incubated with 10 mM paranitrophenyl α-D-glucopyranoside substrate at different pH and release of paranitrophenol was measured as described in the methods. Results from four independent experiments each with two technical replicates are shown.

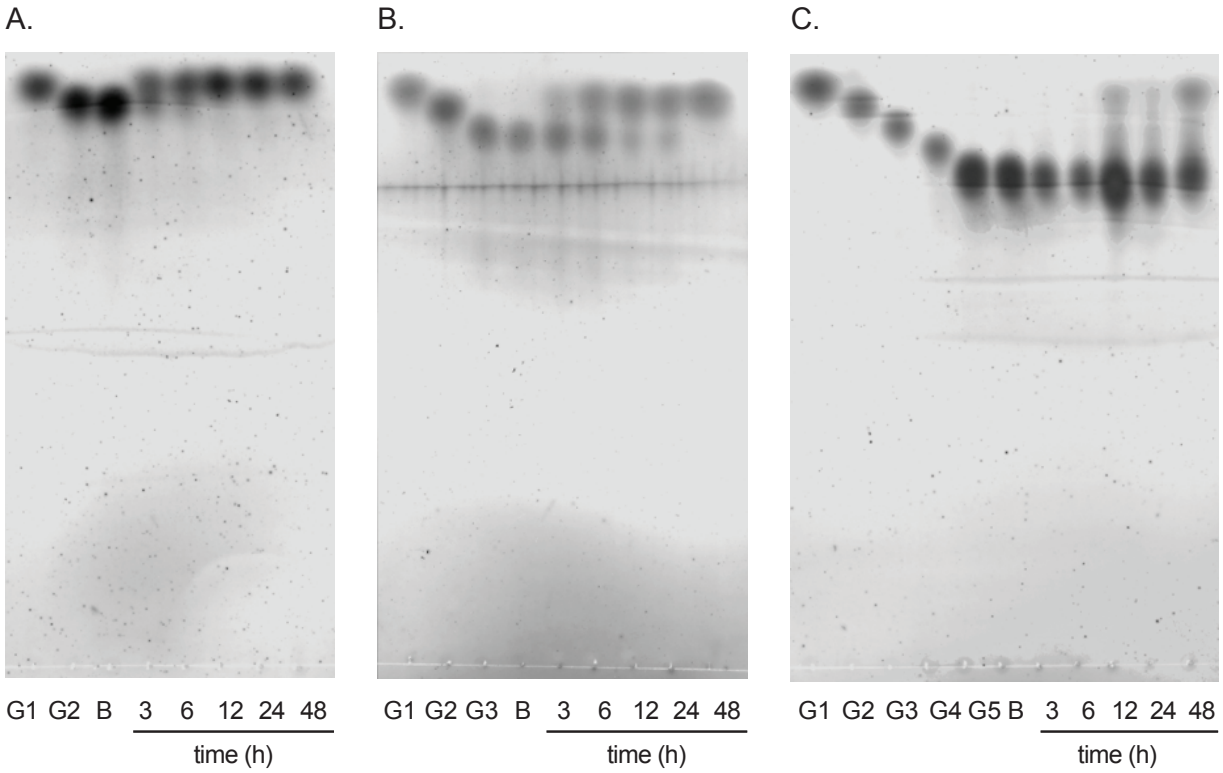


Figure 6. TLC of products of maltose (A), maltotriose (B) and maltopentose (C) hydrolysis by α-glucosidase enzyme. B: Substrate with no enzyme. Reaction mixtures were assessed at 3 h, 6 h, 12 h, 24 h and 48 h.

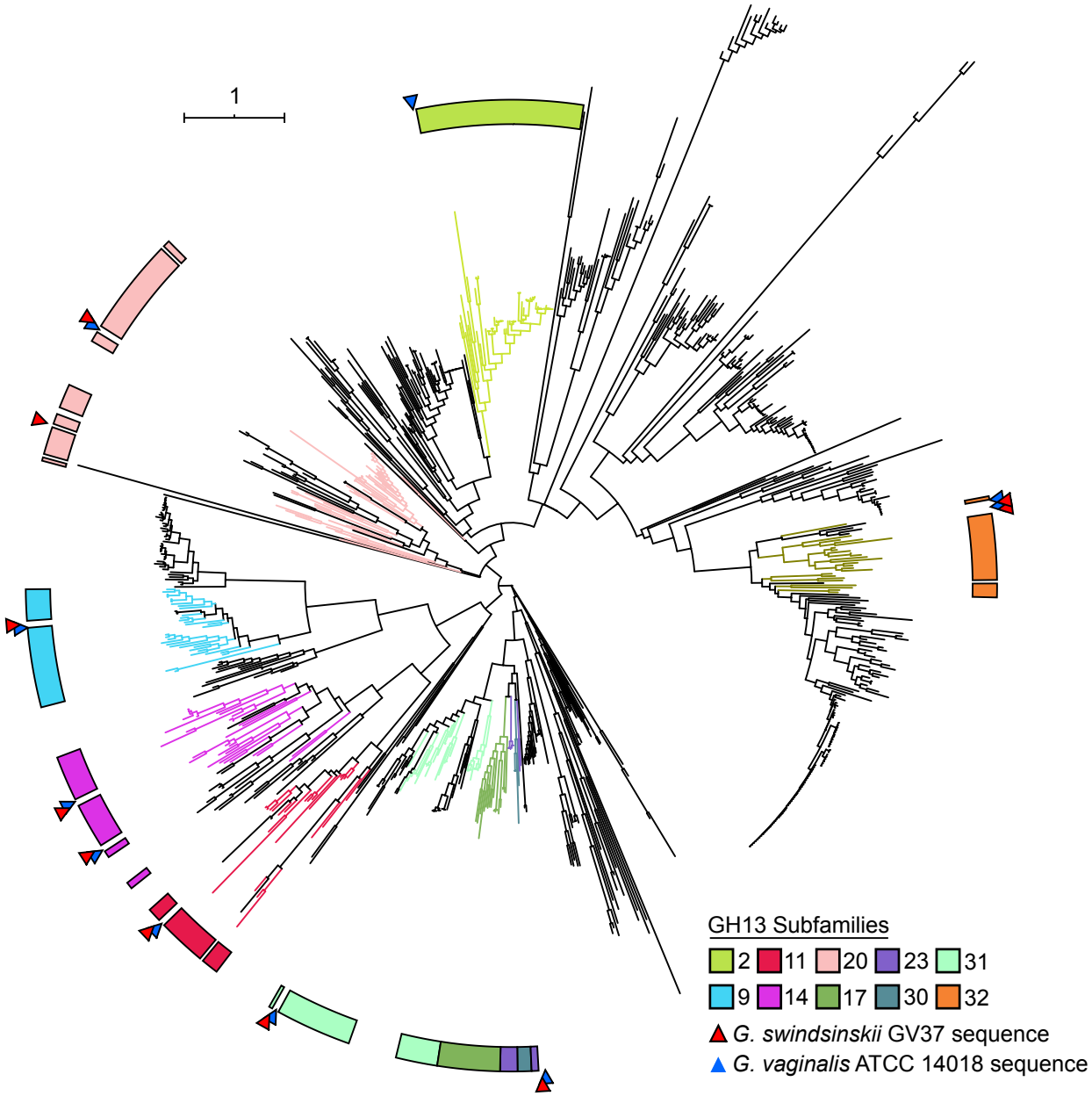
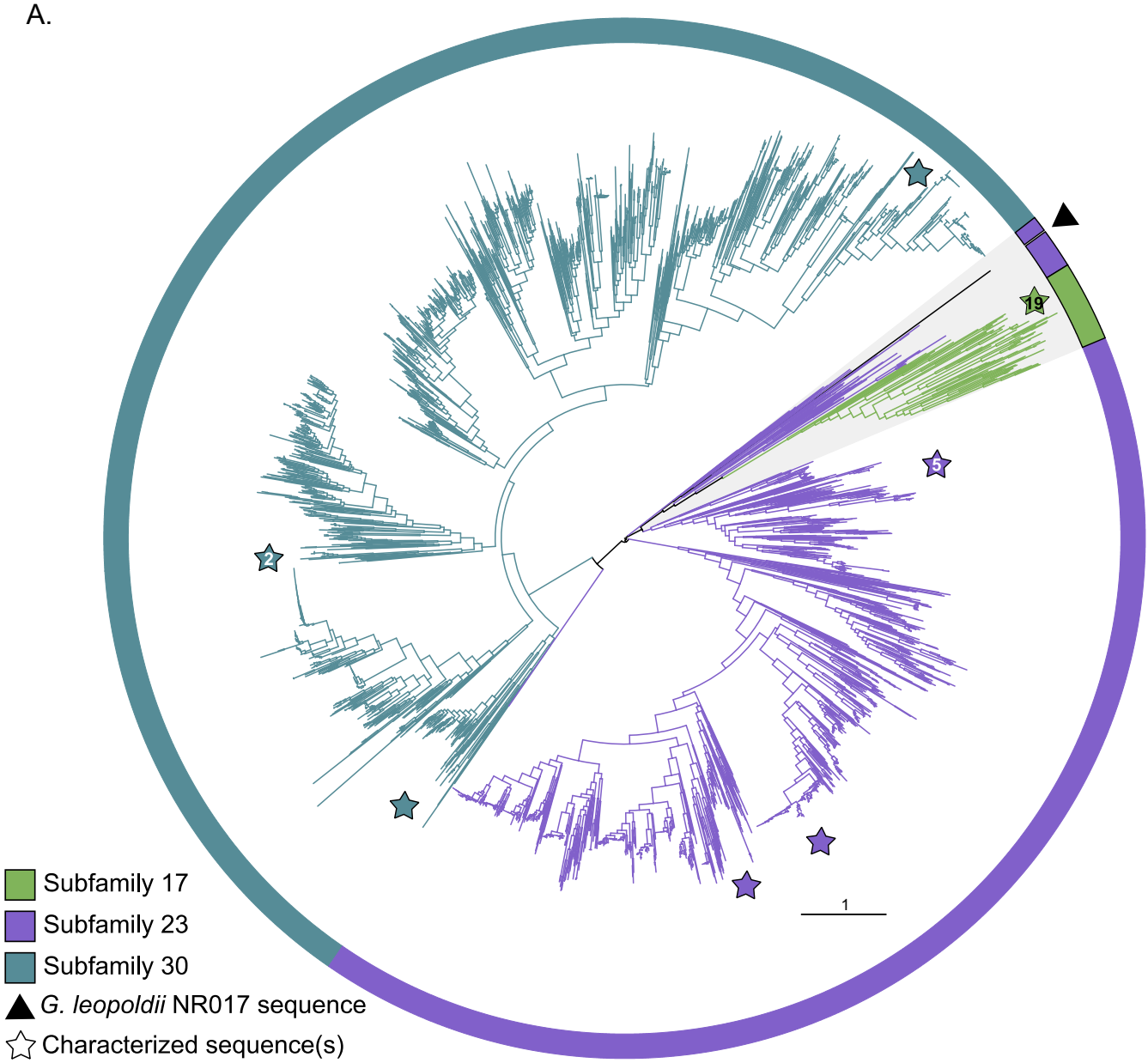


Figure S1. (A) Phylogenetic tree of GH13 functionally characterized members with predicted GH13 domains from *G. swindsinskii* GV37 (red arrow) and *G. vaginalis* ATCC 14018 (blue arrow). Tree branches are coloured based on GH13 subfamily. Trees were generated using SACCHARIS and viewed in iTOL.

A.


19
5
2
1
Subfamily 17
Subfamily 23
Subfamily 30
G. leopoldii NR017 sequence
Characterized sequence(s)

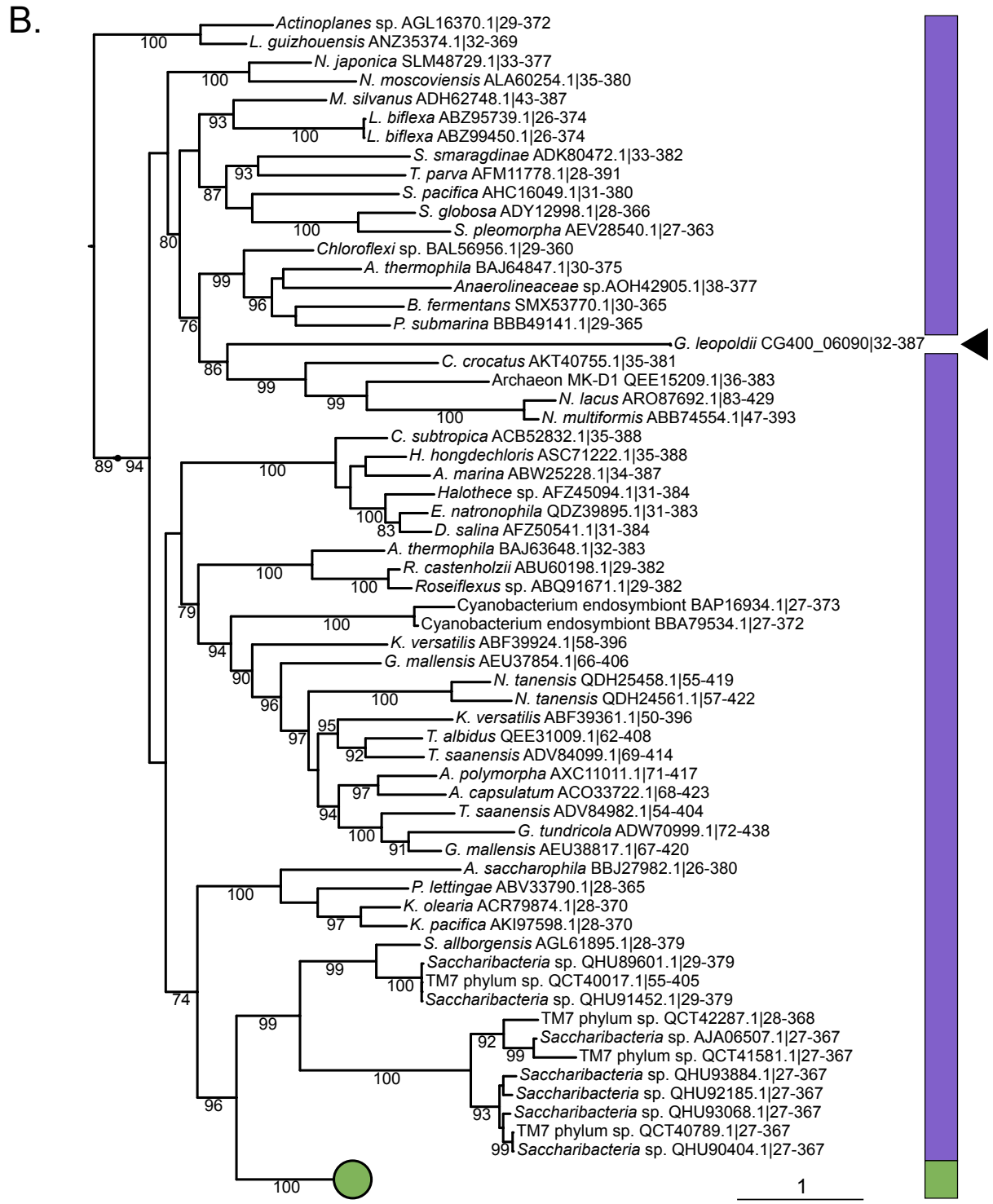


Figure S2. (A) Phylogenetic tree of all GH13 subfamily 17, 23 and 30 members in the CAZy database and *G. leopoldii* CG400_06090. Branch colour is based on GH13 subfamily and functionally characterized proteins are indicated with stars. The clade highlighted in grey including *G. leopoldii* CG400_06090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