



HAL
open science

Mechanistic insights into consumption of the food additive xanthan gum by the human gut microbiota

Matthew Ostrowski, Sabina Leanti La Rosa, Benoit Kunath, Andrew Robertson, Gabriel Pereira, Live Hagen, Neha Varghese, Ling Qiu, Tianming Yao, Gabrielle Flint, et al.

► To cite this version:

Matthew Ostrowski, Sabina Leanti La Rosa, Benoit Kunath, Andrew Robertson, Gabriel Pereira, et al.. Mechanistic insights into consumption of the food additive xanthan gum by the human gut microbiota. 2024. hal-03666905

HAL Id: hal-03666905

<https://amu.hal.science/hal-03666905v1>

Preprint submitted on 12 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 **The Food Additive Xanthan Gum Drives Adaptation of the Human Gut Microbiota**

2

3 *Matthew P. Ostrowski¹, *Sabina Leanti La Rosa^{2,3}, Benoit J. Kunath², Andrew Robertson⁴,
4 Gabriel Pereira¹, Live H. Hagen², Neha J. Varghese⁵, Ling Qiu¹, Tianming Yao⁶, Gabrielle
5 Flint¹, James Li⁷, Sean McDonald⁷, Duna Buttner¹, Nicholas A. Pudlo¹, Matthew K. Schnizlein¹,
6 Vincent B. Young¹, Harry Brumer⁷, Thomas Schmidt¹, Nicolas Terrapon^{8,9}, Vincent Lombard^{8,9},
7 Bernard Henrissat^{8,9,10}, Bruce Hamaker⁶, Emiley A Eloie-Fadros⁵, Ashootosh Tripathi⁴, ^Phillip
8 B. Pope^{2,3}, ^Eric Martens¹

9

10 ¹*Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109,*
11 *USA*

12 ²*Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences,*
13 *Aas N-1433 Norge, Norway.*

14 ³*Faculty of Biosciences, Norwegian University of Life Sciences, Aas N-1433 Norge, Norway.*

15 ⁴*Life Sciences Institute: Natural Products Discovery Core, University of Michigan, Ann Arbor,*
16 *MI 48109, USA*

17 ⁵*DOE Joint Genome Institute, Berkeley, CA, USA*

18 ⁶*Department of Food Science and Whistler Center for Carbohydrate Research, Purdue*
19 *University, West Lafayette, IN 47907, USA*

20 ⁷*Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC,*
21 *V6T 1Z4*

22 ⁸*Centre National de la Recherche Scientifique, Aix-Marseille Univ., UMR7257 AFMB,*
23 *Marseille, France*

24 ⁹*Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, USC1408*
25 *AFMB, Marseille, France*

26 ¹⁰*Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia*

27

28 * These authors contributed equally to this work

29 ^ Correspondence to: phil.pope@nmbu.no, emartens@umich.edu

30

31

32 **Summary**

33 The diets of industrialized countries reflect the increasing use of processed foods, often with the
34 introduction of novel food additives. Xanthan gum is a complex polysaccharide with unique
35 rheological properties that have established its use as a widespread stabilizer and thickening
36 agent¹. However, little is known about its direct interaction with the gut microbiota, which plays
37 a central role in digestion of other, chemically-distinct dietary fiber polysaccharides. Here, we
38 show that the ability to digest xanthan gum is surprisingly common in industrialized human gut
39 microbiomes and appears to be contingent on the activity of a single bacterium that is a member
40 of an uncultured bacterial genus in the family *Ruminococcaceae*. We used a combination of
41 enrichment culture, multi-omics, and recombinant enzyme studies to identify and characterize a
42 complete pathway in this uncultured bacterium for the degradation of xanthan gum. Our data
43 reveal that this keystone degrader cleaves the xanthan gum backbone with a novel glycoside
44 hydrolase family 5 (GH5) enzyme before processing the released oligosaccharides using
45 additional enzymes. Surprisingly, some individuals harbor a *Bacteroides* species that is capable
46 of consuming oligosaccharide products generated by the keystone *Ruminococcaceae* or a
47 purified form of the GH5 enzyme. This *Bacteroides* symbiont is equipped with its own distinct
48 enzymatic pathway to cross-feed on xanthan gum breakdown products, which still harbor the
49 native linkage complexity in xanthan gum, but it cannot directly degrade the high molecular
50 weight polymer. Thus, the introduction of a common food additive into the human diet in the
51 past 50 years has promoted the establishment of a food chain involving at least two members of
52 different phyla of gut bacteria.

53

54 **Introduction**

55 Evidence is accumulating that food additives impact the symbiosis between humans and
56 their associated gut microbiomes, in some cases promoting intestinal inflammation and
57 metabolic syndrome² or promoting certain pathogens³. Often used as thickeners and emulsifiers,
58 polysaccharides are a prominent subset of these food additives. Since dietary polysaccharides
59 other than starch typically transit the upper intestinal tract undigested, polysaccharide-based
60 additives can potentially exert their influence by altering the composition and function of the
61 microbiome, which can in turn impact host health^{4,5}. Given generally regarded as safe (GRAS)
62 approval by the United States Food and Drug Administration in 1968, xanthan gum (XG) is an

63 exopolysaccharide produced by the bacterium *Xanthomonas campestris* that has been
64 increasingly used in the food supply for the last 50 years. This polymer has the same β -1,4-
65 linked backbone as cellulose, but contains trisaccharide branches on alternating glucose residues
66 consisting of α -1,3-mannose, β -1,2-glucuronic acid, and terminal β -1,4-mannose (**Figure 1a**).
67 The terminal β -D-mannose and the inner α -D-mannose are variably pyruvylated at the 4,6-
68 position or acetylated at the 6-position, respectively, with amounts determined by specific *X.*
69 *campestris* strains and culture conditions⁶. XG is typically added at concentrations of 0.05-0.5%
70 to foods including bakery products, condiments, and ice cream⁷. XG is also used as a
71 replacement for gluten in a gluten-free diet, which is a vital component for limiting intestinal
72 inflammation in patients with celiac disease, a lifelong condition estimated to affect 0.7-1.4% of
73 the population and increasing in prevalence⁸. In gluten-free baked goods, XG can be consumed
74 in up to gram quantities per serving. Although small doses of XG have not been connected to
75 immediate health impacts, its fate in the digestive tract is unknown⁹. The low-level, but constant
76 consumption of XG by a large portion of the population in the industrialized world and its higher
77 intake by specific subpopulations highlight the need to understand the effects of this
78 polysaccharide food additive on the ecology of the human gut microbiota.

79

80 **A member of an uncultured bacterial genus is a keystone XG degrader in the human gut** 81 **microbiome**

82 To identify potential XG degrading bacteria in the human gut microbiome, we surveyed a
83 group of healthy 18-20 year-old adults using a bacterial enrichment culture strategy in which a
84 partially defined minimal medium was combined with XG as the main carbon source¹⁰. This
85 medium was inoculated directly with feces that had been collected in anaerobic preservation
86 buffer within 24 h prior to testing and cultures containing individual samples were passaged 3
87 times with 1-2 days of growth in between. Growth of the final passage was monitored
88 quantitatively for bacterial growth or for loss of the gel-like viscosity that is characteristic of XG
89 in solution. We originally found one XG-degrading culture (see *Materials and Methods*) and in
90 an expanded experiment in which 60 individuals were sampled we identified 30 positive
91 cultures, indicating that the ability of intestinal bacteria to degrade XG is unexpectedly common
92 among the population surveyed.

93 Experiments with a culture derived from one positive subject revealed that bacterial
94 growth depended on the amount of XG provided in the medium, demonstrating specificity for
95 this nutrient (**Figure 1b**). Attempts to enrich the causal XG-consuming organism(s) with
96 additional passaging (total of 10-20 times) consistently yielded stable mixed microbial cultures
97 that contained multiple operational taxonomic units (OTUs; between 12-22 OTUs per culture
98 with relative abundance $\geq 0.5\%$) (**Figure 1c, 1d**). While these cultures had commonalities at the
99 genus level, there was surprisingly only one OTU that was $\geq 0.5\%$ and common across all 21
100 enrichment cultures examined. This common OTU was identified as a member of
101 Ruminococcaceae uncultured genus 13 (R. UCG13) in the Silva database¹¹ (**Figure 1d,**
102 **Supplemental Table 1**). Plating and passaging the same culture used in **Figure 1b** on a
103 commonly used anaerobic solid medium (brain-heart infusion with 10% horse blood) resulted in
104 loss of two previously abundant Gram-positive OTUs (loss defined as $< 0.01\%$ relative
105 abundance), which included the R. UCG13 OTU and corresponded with loss of the XG-
106 degrading phenotype when plate-passaged bacteria were re-inoculated into medium with XG
107 (**Figure 1c**).

108 Despite R. UCG13 and a *Bacteroides* OTU being present at $> 20\%$ relative abundance in
109 the original 12 OTU community, we repeatedly failed to isolate pure cultures that could degrade
110 XG using different solid media that are effective for Gram-positive and -negative bacteria (the
111 abundant *Bacteroides* OTU was captured multiple times, while R. UCG13 was never isolated).
112 Dilution of the active 12-OTU community to extinction in medium supplemented with either XG
113 or an equal amount of its component monosaccharides, resulted in loss of growth on XG at
114 higher dilutions than simple sugars (**Extended Data 2**). This observation suggests that the ability
115 to degrade XG in the medium conditions we employed requires multiple OTUs to be present
116 (*i.e.*, diluted into the same well together), which could be explained by either multiple species
117 being directly involved or the necessity of other species to promote sufficient growth conditions
118 for the XG degrader¹². Collectively, these results suggest that a member of an uncultured
119 Ruminococcaceae genus is necessary for XG degradation but may be unable to grow in isolation
120 in our media formulation.

121

122 **Community sequencing identifies two putative XG utilization loci in R. UCG13 and**
123 *Bacteroides intestinalis*

124 To identify XG-degrading genes within our bacterial consortium, we performed
125 combined metagenomics and metatranscriptomics analysis on the original XG-degrading culture,
126 using samples harvested throughout growth in liquid medium with XG (**Extended Data 3**). From
127 these samples, we reconstructed 18 metagenome assembled genomes (MAGs), 7 that were high
128 quality (completion >90% and contamination <5%) (**Supplemental Table 2**). To connect 16S
129 rRNA genes to MAGs, we performed additional long-read sequencing that yielded 2 MAGs that
130 were complete circular chromosomes and one of these was identified as R. UCG13 (with four
131 complete 16S rRNA operons, three of which were identical to the R. UCG13 OTU). This circular
132 R. UCG13 genome was distantly related (47.26% Average Amino Acid Identity, AAI) to the
133 recently cultured bacterium *Monoglobus pectinolyticus*¹³. Annotation of carbohydrate-active
134 enzymes (CAZymes) in the R. UCG13 MAG revealed a single locus encoding several highly
135 expressed enzymes that are candidates for XG degradation (**Figure 2, Extended Data 3**). These
136 included a polysaccharide lyase family 8 (PL8) with distant homology to known xanthan lyases
137 from soil bacteria *Paenibacillus alginolyticus* XL-1¹⁴ (36% identity/73% coverage) and *Bacillus*
138 sp. GL1¹⁵ (32% identity/81% coverage; **Figure 2**). Xanthan lyases typically remove the terminal
139 pyruvylated mannose prior to depolymerization of the β -1,4-glucose backbone, leaving a 4,5
140 unsaturated residue at the glucuronic acid position, although some tolerate non-pyruvylated
141 mannose^{16,17}. This same locus also contained two GH5 enzymes, a family that includes
142 endoglucanases and xyloglucanases, with the potential to cleave the xanthan gum backbone¹⁸, a
143 GH88 to remove the unsaturated glucuronic acid residue produced by the PL8¹⁹, and two GH38s
144 which could potentially cleave the α -D-mannose²⁰. Two carbohydrate esterases (CEs) could
145 potentially remove the acetylation from the mannose²¹. Secretion signal prediction detected
146 possible signal peptidase I (SPI) motifs for the two GH5s and one of the CEs (CE-A), while the
147 other enzymes lacked detectable membrane localization and secretion signals²². In addition to
148 putative enzymes to cleave the glycosidic bonds contained within xanthan gum, this locus also
149 contained proteins predicted to be involved in sensing, binding, and transporting sugars and
150 oligosaccharides.

151 Co-localization and expression of genes that saccharify a common polysaccharide into
152 discrete polysaccharide utilization loci (PULs) is common in the Gram-negative *Bacteroidetes*²³.
153 Although not present in most XG-degrading cultures, we obtained a second circular MAG
154 affiliated to *B. intestinalis*, which was conspicuously the most abundant OTU (up to ~50%) in

155 the mixed species culture that it was derived from (**Figure 1**). This MAG contained a putative
156 PUL that was highly expressed during growth on XG (**Figure 2, Extended Data 3**) and encodes
157 hallmark SusC-/SusD-like proteins, a sensor/regulator, and predicted GH88, GH92 and GH3
158 enzymes, which could potentially cleave the unsaturated β -glucuronyl, α -mannosyl, and β -
159 glucosyl linkages in XG, respectively. Like the candidate gene cluster in R. UCG13, this PUL
160 also contained a GH5 enzyme, which was assigned to subfamily GH5_5. Finally, a putative
161 polysaccharide lyase (PL) was predicted, remotely related to alginate lyases^{24,25}, as a candidate
162 for removing the terminal mannose. In addition to the lyase domain, this multi-modular protein
163 contains a carbohydrate esterase domain (CE) that could remove the acetyl groups positioned on
164 the mannose. Extensive work has been conducted to characterize the substrate-specificity of
165 PULs, which is demonstrated by hundreds of genomes with characterized and predicted PULs in
166 the PUL database (PUL-DB)²⁶. However, this database only harbored a single genome with a
167 partially related homolog of the *B. intestinalis* PUL (*B. salyersiae* WAL 10018 PUL genes
168 HMPREF1532_01924-HMPREF1532_01938), highlighting the diversity of polysaccharide
169 utilization machinery that remains for discovery and characterization.

170 Interestingly, neutral monosaccharide analysis from our XG-degrading culture showed a
171 relatively stable 1:1 ratio of glucose:mannose in the residual polysaccharide in the culture over
172 time, implying that lyase-digested xanthan gum was not accumulating as growth progressed
173 (**Extended Data 3**). This could be due to fast depolymerization and cellular importation of the
174 XG polymer following lyase removal of the terminal mannose. Alternatively, the data are also
175 consistent with a degradation model in which the XG backbone is cleaved prior to subsequent
176 hydrolysis of the repeating pentasaccharide, a pathway that has not been characterized in XG-
177 degrading bacteria and might result in pentasaccharides with full linkage complexity being
178 transported into the bacterial cell before depolymerization^{27,28}.

179

180 **R. UCG13 produces a unique endo-acting xanthanase activity**

181 To investigate the cellular location of the enzymes responsible for xanthan degradation,
182 the original 12-OTU culture was grown in XG medium and separated into filtered cell-free
183 supernatant, cells that were washed to remove supernatant and resuspended or lysed, or lysed
184 cells with supernatant. Incubation of these fractions with XG and subsequent analysis by thin
185 layer chromatography (TLC) revealed that the cell-free supernatant was capable of

186 depolymerizing XG into large oligosaccharides, while the intracellular fraction was required to
187 further saccharify these products into smaller components (**Extended Data 4**). Liquid
188 chromatography-mass spectrometry (LC-MS) analysis of cell-free supernatant incubated with
189 XG revealed the presence of pentameric and decameric oligosaccharides matching the structure
190 of xanthan gum (**Figure 3a**), supporting the model described above in which secreted *endo*-
191 cleaving enzymes first hydrolyze native XG before xanthan lyase and other enzymes cleave the
192 attached sidechains.

193 To identify enzymes responsible for XG hydrolysis and determine their cellular location,
194 we grew three independent cultures in liquid medium containing XG and subjected cell-free
195 supernatants to ammonium sulfate precipitation. Each of the resuspended protein preparations
196 was able to hydrolyze XG as demonstrated by a complete loss of viscosity after overnight
197 incubation. We proceeded to fractionate each sample with various purification methods (defined
198 in methods), collecting and pooling xanthan-degrading fractions for subsequent purification steps
199 and taking three different purification paths (**Extended Data 5**). Interestingly, the most pure
200 sample obtained ran primarily as a large smear when loaded onto an SDS-PAGE gel, but
201 separated into distinct bands after boiling, suggesting the formation of a multimeric protein
202 complex, which is reminiscent of cellulosomes or other complexes²⁹ (**Extended Data 5**).
203 Proteomic analysis of the samples from the three different activity-guided fractionation
204 experiments yielded 33 proteins present across all three experiments, including 22 from R.
205 UCG13, 11 of which were annotated as CAZymes (**Extended Data 5, Supplemental Table 3**).
206 While most of the proteins were either detected in low amounts or lacked functional predictions
207 consistent with polysaccharide degradation, one of the most abundant proteins across all three
208 samples was one of two GH5 enzymes (*RuGH5a*) encoded in the previously identified R.
209 UCG13 candidate xanthan locus.

210 *RuGH5a* consists of an N-terminal signal peptide sequence, its main catalytic domain
211 (which does not classify into any of the GH5 subfamilies), and 3 tandem carbohydrate-binding
212 modules (CBMs), which are often associated with CAZymes and can facilitate polysaccharide
213 degradation (**Figure 3b**)³⁰. The protein also contains a significant portion of undefined sequence
214 and *Listeria-Bacteroides* repeat domains (pfam³¹ PF09479), a β -grasp domain originally
215 characterized from the invasion protein InlB used by *Listeria monocytogenes* for host cell
216 entry^{32,33}. These small repeat domains are thought to be involved in protein-protein interactions

217 and are almost exclusively found in extracellular bacterial multidomain proteins. To test the
218 activity of *RuGH5a* on XG we expressed recombinant forms of the entire protein (*RuGH5a* full),
219 the GH5 domain only (*RuGH5a* GH5-only), and the GH5 domain with either one (*RuGH5a*
220 GH5+CBM-A), two (*RuGH5a* GH5+CBM-A/B), or all three of the CBMs (*RuGH5a*
221 GH5+CBMs, hereafter referred to as *RuGH5a* for simplicity). All but the full-length construct
222 yielded reasonably pure proteins, but only constructs with the GH5 and all three CBMs showed
223 activity on xanthan gum, suggesting a critical role in catalysis for these CBMs. (**Extended Data**
224 **6**). The alternate GH5 (*RuGH5b*) was also expressed in a variety of forms but did not display any
225 activity on XG (**Extended Data 6**).

226 Analysis of the reaction products showed that *RuGH5a* releases pentasaccharide
227 repeating units of XG, with various acetylation and pyruvylation (including di-acetylation as
228 previously described³⁴), and larger decasaccharide structures (**Figure 3a**). While isolation of
229 homogenous pentameric oligosaccharides proved difficult, coincubation of XG with *RuGH5a*
230 and a *Bacillus* sp. PL8 facilitated the isolation of pure tetrasaccharide followed by in-depth 1D
231 and 2D NMR structural characterization (**Extended Data 7**), which was useful in determining
232 *RuGH5a* regiospecificity in the XG backbone. Surprisingly, the NMR analysis suggested that
233 *RuGH5a* cleaves XG at the reducing end of the non-branching backbone glucosyl residue
234 (**Figure 3c, Extended Data 7**). This contrasts with the product of other known xanthanases
235 (such as the GH9 from *Paenibacillus nanensis*³⁵ or the β -D-glucanase in *Bacillus* sp. strain
236 GL1¹⁵), which hydrolyze xanthan at the reducing end of the branching glucose, demonstrating a
237 hitherto unknown enzymatic mechanism for the degradation of XG. While *RuGH5a* displayed
238 little activity on other polysaccharides (**Extended Data 6**), it was able to hydrolyze both native
239 and lyase-treated XG with comparable specificity, once more in contrast to most previously
240 known xanthanases, which show ≥ 600 -fold preference for the lyase-treated substrate³⁵ (**Figure**
241 **3d**). One exception is the xanthanase from *Microbacterium* sp. XT11, which also cleaves native
242 and lyase-treated xanthan gum with similar kinetic specificity²⁸; however, this enzyme only
243 produces intermediate XG fragments from the complete polysaccharide, whereas *RuGH5a* can
244 cleave XG down to its repeating pentasaccharide unit. Together these data highlight the novelty
245 of *RuGH5a*, which may be part of a multimeric protein complex *in vivo* and possesses a unique
246 enzymatic mechanism and specificity.

247

248 **R. UCG13 encodes all the enzymes required for XG saccharification**

249 In contrast to characterized PL8^{17,36,14} xanthan lyases, the R. UCG13 PL8 showed no
250 activity on the complete XG polymer but removed the terminal mannose from xanthan
251 pentasaccharides produced by *RuGH5a* (**Extended Data 8**). This further supports the model in
252 which the GH5 first depolymerizes XG, followed by further saccharification of the XG repeating
253 unit, likely inside the cell. Both R. UCG13 carbohydrate esterases were able to remove acetyl
254 groups from acetylated xanthan pentasaccharides (**Extended Data 8**). The tetrasaccharide
255 produced by the PL8 was processed by the GH88 and both GH38s, which were able to
256 saccharify the resulting trisaccharide (**Extended Data 8**). The GH94 catalyzed the
257 phosphorolysis of cellobiose in phosphate buffer, completing the full saccharification of XG
258 (**Extended Data 8**). Apparent redundancy of several enzymes (CEs and GH38s) could be
259 partially explained by different cell location (e.g. CE-A has an SPI signal while CE-B does not),
260 unique specificities for oligosaccharide variants in size or modification (i.e. acetylation or
261 pyruvylation), additional polysaccharides that the locus targets, or evolutionary hypotheses
262 where this locus is in the process of streamlining or expanding. Additional support for the
263 involvement of this locus in XG degradation is provided by RNA-seq based whole genome
264 transcriptome analysis, which showed the induction of genes in this cluster when the community
265 was grown on XG compared to another polysaccharide (polygalacturonic acid, PGA) that also
266 supports R. UCG13 abundance (**Extended Data 9**).

267

268 ***B. intestinalis* cross-feeds on XG oligosaccharides with its xanthan utilization PUL**

269 Although R. UCG13 was recalcitrant to culturing efforts, we isolated several bacteria
270 from the original consortium, including a representative strain of the *Bacteroides intestinalis* that
271 was the most abundant (**Figure 1c**) and also harbors a highly expressed candidate PUL for XG
272 degradation (**Figure 2**). While this strain was unable to grow on native XG as a substrate, we
273 hypothesized that it may be equipped to utilize smaller XG fragments, such as those released
274 during growth by R. UCG13 via its GH5 enzyme. Using the recombinant *RuGH5a*, we generated
275 sufficient quantities of mixed XG oligosaccharides (XGOs; primarily pentameric, but also some
276 decameric oligosaccharides) to test growth of *Bacteroides intestinalis*. While isolates of *P.*
277 *distasonis* and *B. clarus* from the same culture showed little or no growth (**Extended Data 9**),
278 the *B. intestinalis* strain achieved comparable density on the XGOs as cultures grown on a

279 stoichiometric mixture of the monosaccharides that compose XG, suggesting that it utilizes most
280 or all of the sugars contained in the XGOs (**Figure 4a**). Consistent with the candidate *B.*
281 *intestinalis* XGOs PUL being involved in this phenotype, all of the genes in this locus were
282 activated >100-fold (and some >1000-fold) during growth on XGOs compared to a glucose-
283 grown reference (**Figure 4b**). Whole genome RNA-seq analysis of the *B. intestinalis* strain
284 grown on XGOs revealed that the identified PUL was the most highly upregulated in the
285 genome, further validating its role in metabolism of XGOs (**Extended Data 9**). Interestingly,
286 *RuGH5a* XGOs treated with PL8 continued to support *B. intestinalis* growth, but tetramer
287 generated from the *P. nanensis* GH9 and PL8 failed to support any growth (**Extended Data 9**).
288 Growth was rescued in the presence of glucose but not in the presence of *RuGH5a* XGOs to
289 upregulate the PUL (**Extended Data 9**), suggesting that either the *B. intestinalis* transporters or
290 enzymes are incapable of processing this isomeric substrate.

291 To further test the role of the identified *B. intestinalis* PUL in XGOs degradation, we
292 tested recombinant forms of its constituent enzymes for their ability to degrade XGOs, and
293 confirmed the activity of several enzymes. The carbohydrate esterase domain C-terminal to the
294 PL-CE bimodular protein was able to remove acetyl groups from acetylated xanthan
295 pentasaccharides (**Extended Data 8**). While we were unable to detect xanthan lyase activity for
296 the PL-CE enzyme on full length XG or oligosaccharides it is likely that this enzyme or another
297 lyase acts to remove the terminal mannose residue since the GH88 was able to remove the
298 corresponding 4,5 unsaturated glucuronic acid residue from the corresponding tetrasaccharide
299 that would be generated by its action (**Extended Data 8**). The GH92 was active on the
300 trisaccharide produced by the GH88 as observed by loss of the trisaccharide and formation of
301 cellobiose (**Extended Data 8**). Finally, the GH3 was active on cellobiose, but did not show
302 activity on either tri- or tetra- saccharide, suggesting that this enzyme may be the final step in *B.*
303 *intestinalis* saccharification of XGOs (**Extended Data 8**). Signal peptidase II (SPII) secretion
304 signals were predicted for the GH5, GH3, GH88, and SusD proteins while the GH92, PL-CE,
305 HTCS, and SusC all had SPI motifs²². While signal peptides do not definitively determine
306 cellular location, these predictions and accumulated knowledge of Sus-type systems in
307 Bacteroidetes suggest a “selfish” model in which saccharification occurs primarily in the
308 periplasm^{23,37} (**Extended Data 1**).

309

310 **Metagenomics suggests additional cross-feeding modalities**

311 To determine if the original consortium was representative of all our XG-degrading
312 cultures, we performed metagenomic sequencing on 20 additional XG-degrading communities
313 and retrieved 16 high-quality and 3 low-quality R. UCG13 MAGs as well as an unbinned contig
314 affiliated with R. UCG13 (**Supplemental Table 2**). We found that the R. UCG13 XG utilization
315 locus is extremely well conserved across these cultures with only one variation in gene content,
316 the insertion of a GH125 coding gene, and >95% amino acid identity (**Extended Data 10**). The
317 additional GH125 gene was observed in most loci (14/17), suggesting that this gene provides a
318 complementary, but non-essential function, possibly as an accessory α -mannosidase³⁸. In
319 contrast, only a subset of the samples (4/17) contained the *B. intestinalis* XGOs PUL, which
320 showed essentially complete conservation in cultures that contained this PUL (**Extended Data**
321 **10**). Across all these cultures, R. UCG13 accounted for an average of only $23.1\% \pm 1.2$ (SEM) of
322 the total culture (**Figure 1d**), suggesting that additional microbes beyond *B. intestinalis* may
323 cross-feed on products released by R. UCG13, either from degradation products of XG or by
324 using other growth substrates generated by R. UCG13. For example, we found that the bacterial
325 communities in samples 1, 22, and 59 contain other microbes belonging to the Bacteroidaceae
326 family that harbor a PUL with a GH88, GH92, and GH3, suggesting that these bacteria can
327 potentially metabolize XG-derived tetramers (**Extended Data 10**).

328

329 **Xanthan utilization loci are widespread in modern microbiomes**

330 Next, we asked whether XG inclusion in the modern diet could have increased the
331 prevalence of the R. UCG13 and *B. intestinalis* xanthan loci compared to other populations, such
332 as hunter-gatherers, that are less likely to be exposed to this food additive. Using each locus as a
333 query, we searched several publicly available fecal metagenome datasets collected from
334 populations worldwide. All modern populations sampled displayed some presence of the R.
335 UCG13 XG locus, with the Chinese and Japanese cohorts being the highest (up to 51% in one
336 cohort) (**Figure 5**). The *B. intestinalis* locus was less prevalent, with two industrialized
337 population datasets (Japan and Denmark/Spain) lacking any incidence. Where the locus was
338 present, its prevalence ranged from 1-11%. The three hunter-gatherer or non-industrialized
339 populations sampled, the Yanomami, Hadza, and Burkina Faso had no detected presence of
340 either the R. UCG13 or *B. intestinalis* locus.

341 Although the size of the hunter-gatherer datasets is relatively small, excluding the
342 possibility of a false negative suggests several equally intriguing hypotheses. Most obviously,
343 inclusion of XG in the modern diet may have driven either the colonization or expansion of R.
344 UCG13 (and to a lesser extent *B. intestinalis*) into the gut communities of numerous human
345 populations. This is in concordance with the observations of Daly et al. who found that a set of
346 volunteers fed xanthan gum for an extended period produced stool with increased probability and
347 degree of xanthan degradation³⁹. Alternatively, the modern microbiome is drastically different
348 than that of hunter-gatherers and these differences simply correlate with the abundance of R.
349 UCG13, rather than any causal effect of XG in the diet. Another possible hypothesis is that the
350 microbiomes of hunter-gatherer populations can degrade XG, but through completely different
351 microbes and pathways, a hypothesis that could be tested by culturing microbes from hunter-
352 gatherer populations.

353 To further probe the presence of the identified XG utilization genes in other
354 environments, we conducted an expanded LAST search⁴⁰ of both loci in 72,491 sequenced
355 bacterial isolates and 102,860 genome bins extracted from 13,415 public metagenomes, as well
356 as 21,762 public metagenomes that are part of the Integrated Microbial Genomes &
357 Microbiomes⁴¹ (IMG/M) database using fairly stringent thresholds of 70% alignment over the
358 query and 90% nucleotide identity. This search yielded 35 hits of the R. UCG13 locus in human
359 microbiome datasets, including senior adults, children, and an infant (12-months of age,
360 Ga0169237_00111) (**Supplemental Table 4**). We also found 12 hits for the *B. intestinalis* XGOs
361 locus, all in human microbiome samples except for a single environmental sample from a
362 fracking water sample from deep shales in Oklahoma, USA (81% coverage, 99% identity)
363 (**Extended Data 10, Supplemental Table 4**). XG and other polysaccharides such as guar gum
364 are used in oil industry processes, and genes for guar gum catabolism have previously been
365 found in oil well associated microbial communities⁴². Since most samples searched were non-
366 gut-derived, this demonstrates that XG-degrading R. UCG13 and XGOs-degrading *B. intestinalis*
367 are largely confined to gut samples and can be present across the human lifetime.

368

369 **The mouse microbiome harbors a xanthan utilization locus**

370 To investigate the prevalence of XG-degrading populations beyond the human gut
371 microbiome, we used samples from a previous mouse experiment in which animals fed 5% XG

372 showed increased levels of short chain fatty acids propionate and butyrate, suggesting the ability
373 of members of the mouse microbiome to catabolize and ferment XG⁴³. After culturing mouse
374 feces from this experiment on XG media and confirming its ability to depolymerize XG, we used
375 metagenomics to characterize the community structure in two samples (M1741 and M737),
376 revealing the presence of a microbial species related to *R. UCG13* (AAI values between the
377 human *R. UCG13* and the mouse *R. UCG13* were 75.7% and 75.2% for M1741 and M737,
378 respectively) as well as a XG locus with strikingly similar genetic architecture to our previously
379 characterized human XG locus (**Extended Data 10, Supplemental Table 5**). Although several
380 genes are well conserved across both the human and mouse isolates, we observed significant
381 divergence in the sequences of the respective *RuGH5a* proteins that, based on data with the
382 human locus, initiate XG depolymerization. Specifically, this divergence was more pronounced
383 in the non-catalytic and non-CBM portions of the protein suggesting that while the XG-
384 hydrolyzing functions have been maintained, other domains may be more susceptible to genetic
385 drift. As with the human *RuGH5a*, recombinant versions of the mouse *RuGH5a* were able to
386 hydrolyze XG (**Extended Data 10**) but did not show significant activity on a panel of other
387 polysaccharides. These data suggest that the *R. UCG13* XG locus is more broadly present in
388 mammalian gastrointestinal microbiomes and can at least be recovered through XG-feeding.

389

390 **Prospectus**

391 Our results demonstrate the existence of a multi-phylum food chain in response to XG
392 that appears to have driven the colonization and expansion of *R. UCG13* and *B. intestinalis* in
393 industrialized human microbiomes. The absence of these XG-degraders in pre-industrialized
394 microbiomes and their variable presence across post-industrialized populations suggests that XG-
395 driven modulation of human microbiomes may be an ongoing process. The wide range in levels
396 of XG consumption, variable presence of XG-degrading microbes across human populations,
397 and our finding that *R. UCG13* can colonize infants at an early age highlight the profound
398 impacts that XG may be having on the assembly, stability, and evolution of industrialized human
399 microbiomes.

400 The discovery of XG loci in an environmental sample and mouse microbiome, raises
401 ecological questions about the transfer and evolution of XG utilization between host and non-
402 host associated environments. Although the mouse microbiome with a XG locus could have been

403 exposed to XG through herbivory of *X. campestris* infected plants, mice are affiliated with
404 human activities as pests and XG is used as a food additive in various domesticated animal
405 foodstuffs (e.g. in calf milk replacers⁴⁴), further solidifying a link between these loci and human
406 activities. Since XG is a naturally biosynthesized exopolysaccharide, it is also intriguing to
407 speculate about the role of R. UCG13's XG locus with respect to exopolysaccharides that other
408 microbes may be producing locally in the gut.

409 While many questions remain about the ecological, functional, and health-relevant
410 impacts of XG on the human microbiome, our study provides strong evidence that food additives
411 should not be considered inert and can be drivers of microbiome ecology with potentially broad
412 impacts.

413

414 **Acknowledgments**

415 We gratefully acknowledge Stephanie Theide for growth curve analysis suggestions and Tina
416 Johannessen and Aleksander Lysberg for help with Nanopore metagenomics. We thank the
417 University of Michigan Proteomics Resource Facility, Microbiome Core, and Natural Products
418 Discovery Core for their support in completion of this project. We are grateful for support from
419 the US National Institutes of Health (DK118024, DK125445 to ECM and UL1TR002240 in
420 support of MPO) and the Research Council of Norway (FRIPRO program, PBP and SLLR:
421 250479, LHH: 302639). We thank the University of Michigan Center for Gastrointestinal
422 Research (UMCGR), (NIDDK 5P30DK034933) for financial support with proteomics. The work
423 conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science
424 User Facility, is supported under Contract No. DE-AC02-05CH11231.

425

426 **Methods**

427 ***Isolation, culture, and phylogenetic analysis of xanthan degrading cultures***

428 The original culture was isolated from a survey of 80 healthy adults using a bacterial culture
429 strategy designed to enrich for members of the Gram-negative Bacteroidetes, a phylum that
430 generally harbors numerous polysaccharide-degrading enzymes²³. The original culture was the
431 only XG-degrading culture isolated from this initial survey, likely due to its bias for
432 Bacteroidetes. For subsequent surveys and further culturing fecal samples were collected into
433 pre-reduced phosphate buffered saline, then transferred to an anaerobic chamber (10% H₂, 5%

434 CO₂, and 85% N₂; Coy Manufacturing, Grass Lake, MI) maintained at 37°C. Fecal suspensions
435 were used to inoculate cultures and passaged using partially Defined Medium (DM), which was
436 generally prepared as a 2x stock then mixed 1:1 with 10 mg/mL carbon source (e.g. xanthan
437 gum). Each L of prepared DM medium (pH=7.2) contained 13.6 g KH₂PO₄ (Fisher, P284),
438 0.875 g NaCl (Sigma, S7653), 1.125 g (NH₄)₂SO₄ (Fisher, A702), 2 mg each of adenine,
439 guanine, thymine, cytosine, and uracil (Sigma, A2786, G11950, T0895, C3506, U1128, prepared
440 together as 100x solution), 2 mg of each of the 20 essential amino acids (prepared together as
441 100x solution), 1 mg vitamin K₃ (menadione, Sigma M5625), 0.4 mg FeSO₄ (Sigma, 215422),
442 9.5 mg MgCl₂ (Sigma, M8266), 8 mg CaCl₂ (Sigma, C1016), 5 µg Vitamin B₁₂ (Sigma,
443 V2876), 1 g L-cysteine, 1.2 mg hematin with 31 mg histidine (prepared together as 1,000x
444 solution), 1 mL of Balch's vitamins, 1 mL of trace mineral solution, and 2.5 g beef extract
445 (Sigma, B4888).

446 Each L of Balch's vitamins was prepared with 5 mg *p*-Aminobenzoic acid, 2 mg folic acid
447 (Sigma, F7876), 2 mg biotin (Sigma, B4501), 5 mg nicotinic acid (Sigma, N4126), 5 mg calcium
448 pantothenate (Sigma, P2250), 5 mg riboflavin (Sigma, R7649), 5 mg thiamine HCl (Sigma,
449 T4625), 10 mg pyridoxine HCl, 0.1 mg cyanocobalamin, 5 mg thioctic acid. Prepared Balch's
450 vitamins adjusted to pH 7.0, filter sterilized with 0.22 µm PES filters, and stored in the dark at 4
451 C.

452 Each L of trace mineral solution was prepared with 0.5 g EDTA (Sigma, ED4SS), 3 g
453 MgSO₄*7H₂O, 0.5 g MnSO₄*H₂O, 1 g NaCl (Sigma, S7653), 0.1 g FeSO₄*7H₂O (Sigma,
454 215422), 0.1 g CaCl₂, 0.1 g ZnSO₄*7H₂O, 0.01 g CuSO₄*5H₂O, 0.01 g H₃BO₃ (Sigma,
455 B6768), 0.01 g Na₂MoO₄*2H₂O, 0.02 g NiCl₂*6H₂O. Prepared trace mineral solution was
456 adjusted to pH 7.0, filter sterilized with 0.22 µm PES filters, and stored at room temperature.

457 Samples that showed growth on xanthan gum, as evidenced by loss of viscosity and increased
458 culture density, were subcultured 10 times by diluting an active culture 1:100 into fresh DM-XG
459 medium. For the original culture, multiple samples were stored for gDNA extraction and analysis
460 while for the larger sample set, samples were stored after 10 passages; samples were harvested
461 by centrifugation, decanted, and stored at -20 C until further processing).

462 Frozen cell pellets were resuspended in 500 µL Buffer A (200 mM NaCl, 200 mM Tris-HCl, 20
463 mM EDTA) and combined with 210 µL SDS (20% w/v, filter-sterilized), 500 µL
464 phenol:chloroform (alkaline pH), and ~250 µL acid-washed glass beads (212-300 µm; Sigma).

465 Samples were bead beaten on high for 2-3 minutes with a Mini-BeadBeater-16 (Biospec
466 Products, USA), then centrifuged at 18,000 g for 5 mins. The aqueous phase was recovered and
467 mixed by inversion with 500 μ L of phenol:chloroform, centrifuged at 18,000 g for 3 mins, and
468 the aqueous phase was recovered again. The sample was mixed with 500 μ L chloroform,
469 centrifuged, and then the aqueous phase was recovered and mixed with 0.1 volumes of 3 M
470 sodium acetate (pH 5.2) and 1 volume isopropanol. The sample was stored at -80 C for \geq 30
471 mins, then centrifuged at \geq 20,000 g for 20 mins at 4 C. The pellet was washed with 1 mL room
472 temperature 70% ethanol, centrifuged for 3 mins, decanted, and allowed to air dry before
473 resuspension in 100 μ L sterile water. Resulting samples were additionally purified using the
474 DNeasy Blood & Tissue Kit (QIAGEN, USA).

475 Illumina sequencing, including PCR and library preparation, were performed by the University
476 of Michigan Microbiome Core as described by Kozich et al⁴⁵. Barcoded dual-index primers
477 specific to the 16S rRNA V4 region were used to amplify the DNA. PCR reactions consisted of 5
478 μ L of 4 μ M equimolar primer set, 0.15 μ L of AccuPrime Taq DNA High Fidelity Polymerase, 2
479 μ L of 10x AccuPrime PCR Buffer II (Thermo Fisher Scientific, catalog no. 12346094), 11.85 μ L
480 of PCR-grade water, and 1 μ L of DNA template. The PCR conditions used consisted of 2 min at
481 95°C, followed by 30 cycles of 95°C for 20 s, 55°C for 15 s, and 72°C for 5 min, followed by
482 72°C for 10 min. Each reaction was normalized using the SequalPrep Normalization Plate Kit
483 (Thermo Fisher Scientific, catalog no. A1051001), then pooled and quantified using the Kapa
484 Biosystems Library qPCR MasterMix (ROX Low) Quantification kit for Illumina platforms
485 (catalog no. KK4873). After confirming the size of the amplicon library using an Agilent
486 Bioanalyzer and a high-sensitive DNA analysis kit (catalog no. 5067-4626), the amplicon library
487 was sequenced on an Illumina MiSeq platform using the 500 cycle MiSeq V2 Reagent kit
488 (catalog no. MS-102-2003) according to the manufacturer's instructions with with modifications
489 of the primer set with custom read 1/read 2 and index primers added to the reagent cartridge. The
490 "Preparing Libraries for Sequencing on the MiSeq" (part 15039740, Rev. D) protocol was used
491 to prepare libraries with a final load concentration of 5.5 pM, spiked with 15% PhiX to create
492 diversity within the run.

493 Sequencing FASTQ files were analyzed using mothur (v.1.40.5)⁴⁶ using the Silva reference
494 database¹¹. OTUs with the same genus were combined and displayed using R⁴⁷ with the
495 packages reshape2⁴⁸, RColorBrewer⁴⁹, and ggplot2⁵⁰.

496 ***Dilution to extinction experiment***

497 An overnight culture was serially diluted in 2x DM. Serial dilutions were split into two 50 mL
498 tubes and mixed 1:1 with either 10 mg/mL xanthan gum or 10 mg/mL monosaccharide mixture
499 (4 mg/mL glucose, 4 mg/mL mannose, 2 mg/mL sodium glucuronate), both of which also had 1
500 mg/mL L-cysteine. Each dilution and carbon source was aliquoted to fill a full 96-well culture
501 plate (Costar 3370) with 200 μ L per well. Plates were sealed with Breathe-Easy gas permeable
502 sealing membrane for microtiter plates (Diversified Biotech, cat #BEM-1). Microbial growth
503 was measured at least 60 hours by monitoring OD₆₀₀ using a Synergy HT plate reader (Biotek
504 Instruments) and BIOSTACK2WR plate handler (Biotek Instruments)⁵¹.

505 Maximum OD for each substrate was measured for each culture. Full growth on substrates was
506 conservatively defined as a maximum OD₆₀₀ of >0.7. For each unique 96 well plate of substrate
507 and dilution factor, the fraction of wells exhibiting full growth was calculated. Fractional growth
508 was plotted against dilution factor for each substrate. Data were fit to the Hill equation by
509 minimizing squared differences between the model and experimental values using Solver (GRG
510 nonlinear) in Excel. For each experiment, a 50% growth dilution factor (GDF 50) was calculated
511 for each substrate at which half of the wells would be predicted to exhibit full growth.

512 ***Metagenomic analysis***

513 Seven samples (15-mL) were collected at four time points (Extended Data 3; referred to as T1,
514 T2, T3 and T4) during growth of two biological replicates of the original XG-degrading culture.
515 Cells were harvested by centrifugation at 14,000 \times g for 5 min and stored at -20 °C until further
516 use. A phenol:chloroform:isoamyl alcohol and chloroform extraction method was used to obtain
517 high molecular weight DNA as previously described⁵². The gDNA was quantified using a
518 Qubit™ fluorimeter and the Quant-iT™ dsDNA BR Assay Kit (Invitrogen, USA), and the
519 quality was assessed with a NanoDrop One instrument (Thermo Fisher Scientific, USA).

520 Samples were subjected to metagenomic shotgun sequencing using the Illumina HiSeq 3000
521 platform at the Norwegian Sequencing Center (NSC, Oslo, Norway). Samples were prepared
522 with the TrueSeq DNA PCR-free preparation and sequenced with paired ends (2 \times 150 bp) on
523 one lane. Quality trimming of the raw reads was performed using Cutadapt⁵³ v1.3, to remove all
524 bases on the 3'-end with a Phred score lower than 20 and exclude all reads shorter than 100
525 nucleotides, followed by a quality filtering using the FASTX-Toolkit v.0.0.14

526 (http://hannonlab.cshl.edu/fastx_toolkit/). Retained reads had a minimum Phred score of 30 over

527 90% of the read length. Reads were co-assembled using metaSPAdes⁵⁴ v3.10.1 with default
528 parameters and k-mer sizes of 21, 33, 55, 77 and 99. The resulting contigs were binned with
529 MetaBAT⁵⁵ v0.26.3 in “very sensitive mode”. The quality (completeness, contamination, and
530 strain heterogeneity) of the metagenome assembled genomes (MAGs) was assessed by
531 CheckM⁵⁶ v1.0.7 with default parameters. Contigs were submitted to the Integrated Microbial
532 Genomes and Microbiomes system for open reading frames (ORFs) prediction and annotation⁵⁷.
533 Additionally, the resulting ORF were annotated for CAZymes using the CAZy annotation
534 pipeline⁵⁸. This MAG collection was used as a reference database for mapping of the
535 metatranscriptome data, as described below. Taxonomic classifications of MAGs were
536 determined using both MiGA⁵⁹ and GTDB-Tk⁶⁰.

537 Human fecal samples (20) from a second enrichment experiment (unbiased towards the
538 cultivation of Bacteroides) as well as two enrichments with mouse fecal samples were processed
539 for gDNA extraction and library preparation exactly as described above. Metagenomic shotgun
540 sequencing was conducted on two lanes of both Illumina HiSeq 4000 and Illumina HiSeq X Ten
541 platforms (Illumina, Inc.) at the NSC (Oslo, Norway), and reads were quality trimmed,
542 assembled and binned as described above. Open reading frames were annotated using
543 PROKKA⁶¹ v1.14.0 and resulting ORFs were further annotated for CAZymes using the CAZy
544 annotation pipeline and expert human curation⁵⁸. Completeness, contamination, and taxonomic
545 classifications for each MAG were determined as described above. AAI comparison between the
546 human R. UCG13 and the R. UCG13 found in the two mouse samples was determined using
547 CompareM (<https://github.com/dparks1134/CompareM>).

548 Extracted DNA from a second enrichment experiment on XG using the original culture was
549 prepared for long-reads sequencing using Oxford Nanopore Technologies (ONT) Ligation
550 Sequencing Kit (SQK-LSK109) according to the manufacture protocol. The DNA library was
551 sequenced with the ONT MinION Sequencer using a R9.4 flow cell. The sequencer was
552 controlled by the MinKNOW software v3.6.5 running for 6 hours on a laptop (Lenovo ThinkPad
553 P73 Xeon with data stored to 2Tb SSD), followed by base calling using Guppy v3.2.10 in ‘fast’
554 mode. This generated in total 3.59 Gb of data. The Nanopore reads were further processed using
555 Filtlong v0.2.0 (<https://github.com/rrwick/Filtlong>), discarding the poorest 5% of the read bases,
556 and reads shorter than 1000 bp.

557 The quality processed Nanopore long-reads were assembled using CANU⁶² v1.9 with the
558 parameters *corOutCoverage=10000 corMinCoverage=0 corMhapSensitivity=high*
559 *genomeSize=5m redMemory=32 oeaMemory=32 batMemory=200*. An initial polishing of the
560 generated contigs were carried out using error-corrected reads from the assembly with
561 minimap2⁶³ v2.17 -x *map-ont* and Racon⁶⁴ v1.4.14 with the argument *--include-unpolished*. The
562 racon-polished contigs were further polished using Medaka v1.1.3
563 (<https://github.com/nanoporetech/medaka>), with the commands *medaka_consensus --model*
564 *r941_min_fast_g303_model.hdf5*. Finally, Minimap2 -ax *sr* was used to map quality processed
565 Illumina reads to the medaka-polished contigs, followed by a final round of error correction
566 using Racon with the argument *--include-unpolished*. Circular contigs were identified by linking
567 the contig identifiers in the polished assembly back to *suggestCircular=yes* in the initial contig
568 header provided by CANU. These contigs were quality checked using CheckM⁵⁶ v1.1.3 and
569 BUSCO⁶⁵ v4.1.4. Circular contigs likely to represent chromosomes (> 1 Mbp) were further gene-
570 called and functionally annotated using PROKKA⁶¹ v1.13 and taxonomically classified using
571 GTDB-tk⁶⁰ v1.4.0 with the *classify_wf* command. Barnap v0.9
572 (<https://github.com/tseemann/barnap>) was used to predict ribosomal RNA genes. Average
573 nucleotide Identity (ANI) was measured between the short-reads and long-reads MAGs using
574 FastANI⁶⁶ v1.1 with default parameters. Short-reads MAGs were used as query while long-reads
575 MAGs were set as reference genomes. Short-reads MAG1 showed an Average Nucleotide
576 Identity (ANI) of 99.98% with the long-reads ONT_Circ01, while short-reads MAG2 showed an
577 ANI of 99.99% with the long-reads ONT_Circ02 (**Supplemental Table 2**). Phylogenetic
578 analysis revealed that ONT_Circ02 encoded four complete 16S rRNA operons, three of which
579 were identical to the aforementioned R. UCG13 OTU.

580 **Temporal metatranscriptomic analysis of the original XG-degrading community.**

581 Cell pellets from 6 mL samples collected at T1-T4 during growth of two biological replicates of
582 the original XG-degrading culture were supplemented with RNAprotect Bacteria Reagent
583 (Qiagen, USA) following the manufacturer's instructions and kept at -80 °C until RNA
584 extraction. mRNA extraction and purification were conducted as described in Kunath et al.⁶⁷.
585 Samples were processed with the TruSeq stranded RNA sample preparation, which included the
586 production of a cDNA library, and sequenced on one lane of the Illumina HiSeq 3000 system

587 (NSC, Oslo, Norway) to generate 2×150 paired-end reads. Prior to assembly, RNA reads were
588 quality filtered with Trimmomatic⁶⁸ v0.36, whereby the minimum read length was required to be
589 100 bases and an average Phred threshold of 20 over a 10 nt window, and rRNA and tRNA were
590 removed using SortMeRNA⁶⁹ v.2.1b. Reads were pseudo-aligned against the metagenomic
591 dataset using kallisto pseudo –pseudobam⁷⁰. Of the 58089 ORFs (that encode proteins with > 60
592 aa) identified from the metagenome of the original XG-degrading community, 7549 (13%) were
593 not found to be expressed, whereas 50540 (87%) were expressed, resulting in a reliable
594 quantification of the expression due to unique hits (reads mapping unambiguously against one
595 unique ORF).

596 *Neutral Monosaccharide analysis*

597 The hot-phenol extraction method originally described by Massie & Zimm⁷¹ and modified by
598 Nie⁷² was used for collecting and purifying the polysaccharides remaining at different
599 timepoints. Samples were heated to 65 °C for 5 mins, combined with an equal volume of phenol,
600 incubated at 65 °C for 10 mins, then cooled to 4 °C and centrifuged at 4 °C for 15 min at 12,000
601 g. The upper aqueous layer was collected and re-extracted using the same procedure, dialyzed
602 extensively against deionized water (2000 Da cutoff), and freeze-dried. Neutral monosaccharide
603 composition was obtained using the method described by Tuncil et al.⁷³ Briefly, sugar alditol
604 acetates were quantified by gas chromatography using a capillary column SP-2330 (SUPELCO,
605 Bellefonte, PA) with the following conditions: injector volume, 2 µl; injector temperature, 240
606 °C; detector temperature, 300 °C; carrier gas (helium), velocity 1.9 meter/second; split ratio, 1:2;
607 temperature program was 160 °C for 6 min, then 4 °C/min to 220 °C for 4 min, then 3 °C/min to
608 240 °C for 5 min, and then 11 °C/min to 255 °C for 5 min.

609 *Thin Layer Chromatography for Localization of Enzyme Activity*

610 Overnight cultures were harvested at 13,000 g for 10 minutes. Supernatant fractions were
611 prepared by vacuum filtration through 0.22 µm PES filters. Cell pellet fractions were prepared
612 by decanting supernatant, washing with phosphate buffered saline (PBS), spinning at 13,000 g
613 for 3 mins, decanting, and resuspending in PBS. Intracellular fractions were prepared by taking
614 cell pellet fractions and bead beating for 90 s with acid-washed glass beads (G1277, Sigma) in a
615 Biospec Mini Beadbeater. Lysed culture fractions were prepared by directly bead beating
616 unprocessed culture.

617 Each culture fraction was mixed 1:1 with 5 mg/mL xanthan gum and incubated at 37 C for 24
618 hours. Negative controls were prepared by heating culture fractions to 95 C for 15 mins, then
619 centrifuging at 13,000 g for 10 mins before the addition of xanthan gum. All reactions were
620 halted by heating to ≥ 85 C for 15 mins, then spun at 20,000 g for 15 mins at 4 C. Supernatants
621 were stored at -20 C until analysis by thin layer chromatography.

622 3 μ L sample were spotted twice onto a 10x20 cm thin layer chromatography plate (Millipore
623 TLC Silica gel 60, 20x20cm aluminum sheets), with intermediate drying using a Conair 1875
624 hairdryer. Standards included malto-oligosaccharides of varying lengths (Even: 2, 4, 6, Odd: 1,
625 3, 5, 7), glucuronic acid, and mannose. Standards were prepared at 10 mM and 3 μ L of each was
626 spotted onto the TLC plate. Plates were run in \sim 100 mL of 2:1:1 butanol, acetic acid, water,
627 dried, then run an additional time. After drying, plates were incubated in developing solution
628 (100 mL ethyl acetate, 2 g diphenylamine, 2 mL aniline, 10 mL of \sim 80% phosphoric acid, 1 mL
629 of \sim 38% hydrochloric acid) for \sim 30 seconds, then dried, and developed by holding over a flame
630 until colors were observed.

631 *Proteomic analysis*

632 Approximately 1 L of xanthan gum culture was grown until it had completely liquified (\sim 2-3
633 days). Supernatant was collected by centrifuging at 18,000 g and vacuum filtering through a 0.2
634 μ m PES filter. 4M ammonium sulfate was added to 200-400 mL of filtrate to a final
635 concentration of 2.4M and incubated for 30-60 mins at RT or, for one sample, overnight at 4 C.
636 Precipitated proteins were harvested by centrifugation at 18,000 g for 30-60 mins, then
637 resuspended in 50 mM sodium phosphate (pH 7.5). Three different fractionation protocols were
638 followed, but after every fractionation step, active fractions were identified by mixing \sim 500 μ L
639 with 10 mg/mL xanthan and incubating at 37 C overnight; active-fractions were identified by
640 loss of viscosity or production of xanthan oligosaccharides as visualized by TLC (method
641 previously described).

642 1. Resuspended protein was filtered and applied to a HiTrapQ column, running a gradient from
643 0-100% B (Buffer A: 50 mM sodium phosphate, pH 7.5; Buffer B: 50 mM sodium phosphate, 1
644 M NaCl, pH 7.5). Active fractions were pooled and concentrated with a 10 kDa MWCO
645 centricon and injected onto an S-200 16/60 column equilibrated in 50 mM sodium phosphate,
646 200 mM NaCl, pH 7.5. The earliest fractions to elute with significant A280 absorbance were also
647 the most active fractions; these were pooled and submitted for proteomics.

648 2. Resuspended protein was filtered and applied to an S-500 column equilibrated in 50 mM
649 sodium phosphate, 200 mM NaCl, pH 7.5. Active fractions eluted in the middle of the separation
650 were pooled and submitted for proteomics.

651 3. Resuspended protein was filtered and applied to an S-500 column equilibrated in 50 mM
652 sodium phosphate, 200 mM NaCl, pH 7.5. Pooled fractions were applied to a 20 mL strong anion
653 exchange column running a gradient from 0-100% B (Buffer A: 50 mM sodium phosphate, pH
654 7.5; Buffer B: 50 mM sodium phosphate, 1 M NaCl, pH 7.5). Active fractions were pooled and
655 applied to a 1 mL weak anion exchange column (ANX) running a gradient from 0-100% B
656 (Buffer A: 50 mM sodium phosphate, 10% glycerol, pH 7.5; Buffer B: 50 mM sodium
657 phosphate, 1 M NaCl, 10% glycerol, pH 7.5). Active fractions were pooled and submitted for
658 proteomics.

659 Cysteines were reduced by adding 50 ml of 10 mM DTT and incubating at 45 °C for 30 min.
660 Samples were cooled to room temperature and alkylation of cysteines was achieved by
661 incubating with 65 mM 2-Chloroacetamide, under darkness, for 30 min at room temperature. An
662 overnight digestion with 1 ug sequencing grade, modified trypsin was carried out at 37 C with
663 constant shaking in a Thermomixer. Digestion was stopped by acidification and peptides were
664 desalted using SepPak C18 cartridges using manufacturer's protocol (Waters). Samples were
665 completely dried using vacufuge. Resulting peptides were dissolved in 8 ml of 0.1% formic
666 acid/2% acetonitrile solution and 2 µls of the peptide solution were resolved on a nano-capillary
667 reverse phase column (Acclaim PepMap C18, 2 micron, 50 cm, ThermoScientific) using a 0.1%
668 formic acid/2% acetonitrile (Buffer A) and 0.1% formic acid/95% acetonitrile (Buffer B)
669 gradient at 300 nl/min over a period of 180 min (2-25% buffer B in 110 min, 25-40% in 20 min,
670 40-90% in 5 min followed by holding at 90% buffer B for 10 min and requilibration with Buffer
671 A for 30 min). Eluent was directly introduced into *Q exactive HF* mass spectrometer (Thermo
672 Scientific, San Jose CA) using an EasySpray source. MS1 scans were acquired at 60K resolution
673 (AGC target=3x10⁶; max IT=50 ms). Data-dependent collision induced dissociation MS/MS
674 spectra were acquired using Top speed method (3 seconds) following each MS1 scan (NCE
675 ~28%; 15K resolution; AGC target 1x10⁵; max IT 45 ms).

676 Proteins were identified by searching the MS/MS data against a database of all proteins
677 identified in the original culture metagenomes using Proteome Discoverer (v2.1, Thermo
678 Scientific). Search parameters included MS1 mass tolerance of 10 ppm and fragment tolerance

679 of 0.2 Da; two missed cleavages were allowed; carbamidimethylation of cysteine was considered
680 fixed modification and oxidation of methionine, deamidation of asparagine and glutamine were
681 considered as potential modifications. False discovery rate (FDR) was determined using
682 Percolator and proteins/peptides with an FDR of $\leq 1\%$ were retained for further analysis.

683 ***Plasmid Design and Protein Purification***

684 Plasmid constructs to produce recombinant proteins were made with a combination of
685 synthesized DNA fragments (GenScript Biotech, Netherlands) and PCR amplicons using
686 extracted culture gDNA as a template. In general, sequences were designed to remove N-
687 terminal signaling peptides and to add a histidine tag for immobilized metal affinity
688 chromatography (IMAC) (in many cases using the Lucigen MA101-Expresso-T7-Cloning-&-
689 Expression-System). Plasmid assembly and protein sequences are described in **Supplemental**
690 **Table 6**.

691 Constructs were transformed into HI-Control BL21(DE3) cells and single colonies were
692 inoculated in 5 mL overnight LB cultures at 37°C. 5 mL cultures were used to inoculate 1 L of
693 Terrific Broth (TB) with selective antibiotic, grown to OD ~0.8-1.1 at 37°C, and induced with
694 250 μM IPTG. *B. intestinalis* enzymes were expressed at RT, while R. UCG13 enzymes were
695 expressed at 18°C overnight. Cells were harvested by centrifugation and pellets were stored at -
696 80°C until further processing. Proteins were purified using standard IMAC purification
697 procedures employing sonication to lyse cells. R. UCG13 proteins were purified using 50 mM
698 sodium phosphate and 300 mM sodium chloride at pH 7.5; *B. intestinalis* proteins were purified
699 using 50 mM Tris and 300 mM sodium chloride at pH 8.0. All proteins were eluted from cobalt
700 resin using buffer with the addition of 100 mM imidazole, then buffer exchanged to remove
701 imidazole using Zeba 2 mL 7kDa MWCO desalting columns. Protein concentrations were
702 determined by measuring A280 and converting to molarity using calculated extinction
703 coefficients.

704

705 ***Characterization and isolation of Xanthan gum degradation products***

706 In general, pentameric xanthan oligosaccharides were produced by incubating ≥ 0.1
707 mg/mL *RuGH5a* with 5 mg/mL xanthan gum in PBS in approximately 1L total volume. For
708 xanthan tetrasaccharides, ~0.5 U/mL of Xanthan lyase (E-XANLB, Megazyme) was included.
709 After incubating 2-3 days at 37 °C to allow complete liquefaction, reactions were heat-

710 inactivated, centrifuged at $\geq 10,000$ g for 30 mins, and the supernatant was vacuum filtered
711 through 0.22 μm PES sterile filters. Supernatants were loaded onto a column containing ~ 10 g of
712 graphitized carbon (SupelcleanTM ENVI-CarbTM, 57210-U Supelco), washed extensively with
713 water to remove salt and unbound material, then eluted in a stepwise fashion with increasing
714 concentrations of acetonitrile. Fractions were dried, weighed, and analyzed by LC-MS and
715 fractions that contained the most significant yield of desired products were combined.
716 Highly pure products were obtained by reconstituting samples in 50% water:acetonitrile and
717 applying to a Luna[®] 5 μm HILIC 200 Å LC column (250 x 10 mm) (00G-4450-N0,
718 Phenomenex). A gradient was run from 90-20% acetonitrile, with peaks determined through a
719 combination of evaporative light scattering, UV, and post-run analytical LC-MS (Agilent qToF
720 6545) of resulting fractions.

721 NMR spectra were collected using an Agilent 600 NMR spectrometer (¹H: 600
722 MHz, ¹³C: 150 MHz) equipped with a 5 mm DB AUTOX PFG broadband probe and a Varian
723 NMR System console. All data analysis was performed using MestReNova NMR software. All
724 chemical shifts were referenced to residual solvent peaks [¹H (D₂O): 4.79 ppm].

725

726 ***Enzyme Reaction Analysis***

727 All enzyme reactions were carried out in 15-25 mM sodium phosphate buffer, 100-150
728 mM sodium chloride, and sometimes included up to 0.01 mg/mL bovine serum albumin
729 (B9000S, NEB) to limit enzyme adsorption to pipettes and tubes. All *R. UCG13* or *B. intestinalis*
730 enzymes were tested at concentrations from 1-10 μM . Cellobiose reactions were tested using 1
731 mM cellobiose at pH 7.5, while all other reactions used 2.5 mg/mL pentasaccharide (produced
732 using *RuGH5a*) and were carried out at pH 6.0. Reactions were incubated overnight at 37°C,
733 halted by heating at $\geq 95^\circ\text{C}$ for 5-10 minutes, and centrifugation at $\geq 20,000$ g for 10 mins.
734 Supernatants were mixed with 4 parts acetonitrile to yield an 80% acetonitrile solution,
735 centrifuged for 10 mins at $\geq 20,000$ g, and transferred into sample vials. 15 μL of each sample
736 was injected onto a Luna[®] Omega 3 μm HILIC 200 Å LC column (100 x 4.6 mm) (00D-4449-
737 E0, Phenomenex). An Agilent 1290 Infinity II HPLC system was used to separate the sample
738 using solvent A (100% water, 0.1% formic acid) and solvent B (95% acetonitrile, 5% water, with
739 0.1% formic acid added) at a flow rate of 0.4 mL/min. Prior to injection and following each
740 sample the column was equilibrated with 80% B. After injection, samples were eluted with a 30

741 minute isocratic step at 80% B, a 10 minute gradient decreasing B from 80% to 10%, and a final
742 column wash for 2 min at 10% B. Spectra were collected in negative mode using an Agilent
743 6545 LC/Q-TOF.

744

745

746 ***Kinetics of RuGH5a (BCA Assay)***

747 Lyase-treated xanthan gum was generated by mixing 5 mg/mL xanthan gum with 0.5 U/mL of
748 *Bacillus* sp. Xanthan lyase (E-XANLB, Megazyme) in 30 mM potassium phosphate buffer (pH
749 6.5). After incubating overnight at 37 °C, an additional 0.5 U/mL of xanthan lyase was added.
750 Both lyase-treated and native xanthan gum were dialyzed extensively against deionized water,
751 heated in an 80 °C water bath to inactivate the lyase, and centrifuged at 10,000 g for 20 mins to
752 remove particulate. Supernatants were collected and stored at 4 °C until use.

753 Kinetic measurements were conducted using a slightly modified version of the low-volume
754 bicinchoninic acid (BCA) assay for glycoside hydrolases used by Arnal et al⁷⁴. Briefly, AEX and
755 SEC purified *RuGH5a* was diluted to a 10x stock of 5 µM enzyme, 50 mM sodium phosphate,
756 300 mM sodium chloride, and 0.1 mg/mL bovine serum albumin, pH7.5. Reactions were 20 µL
757 of enzyme stock mixed with 180 µL of various concentrations of xanthan gum. Negative controls
758 were conducted with heat-inactivated enzyme stock. Timepoints were taken by quenching
759 reactions with dilute, ice-cold, BCA working reagent. Reactions and controls were run with 4
760 independent replicates and compared to a glucose standard curve. Enzyme released reducing
761 sugar was calculated by subtracting controls from reaction measurements.

762 ***Growth curves of isolates on XG oligos***

763 Pure isolates from the xanthan culture were obtained by streaking an active culture onto a variety
764 of agar plates including LB and brain heart infusion with the optional addition of 10%
765 defibrinated horse blood (Colorado Serum Co.) and gentamicin (200 µg/mL). After passaging
766 isolates twice on agar plates, individual colonies were picked and grown overnight in tryptone-
767 yeast extract-glucose (TYG) broth medium, then stocked by mixing with 0.5 volumes each of
768 TYG and molecular biology grade glycerol and storing at -80 °C.

769 DM without beef extract (DM^{-BE}), with the addition of a defined carbon source, was used to test
770 isolates for growth on xanthan oligosaccharides. Some isolates (e.g. *Parabacteroides distasonis*)
771 required the inclusion of 5 mg/mL beef extract (Sigma, B4888) to achieve robust growth on

772 simple monosaccharides; in these cases, beef extract was included across all carbon conditions.
773 Unless otherwise specified, carbon sources were provided at a final concentration of 5 mg/mL.
774 Isolates were grown overnight in TYG media, subcultured 1:50 into DM^{BE}-glucose and grown
775 overnight, then subcultured 1:50 into DM^{BE} with either various carbon sources. Final cultures
776 were monitored for growth by measuring increase in absorbance (600 nm) using 96-well plates
777 as previously described.

778 ***qPCR and RNA-seq on B. intestinalis and original community***

779 For qPCR, *B. intestinalis* was grown as before but cells were harvested by centrifugation at mid-
780 exponential phase, mixed with RNA Protect (QIAGEN), and stored at -80 °C until further
781 processing. At collection, average OD₆₀₀ values were ~0.8 and ~0.6 for glucose- and
782 oligosaccharide-grown cultures, respectively. RNeasy mini kit buffers (QIAGEN) were used to
783 extract total RNA, purified with RNA-binding spin columns (Epoch), treated with DNase I
784 (NEB), and additionally purified using the RNeasy mini kit. SuperScript III reverse transcriptase
785 and random primers (Invitrogen) were used to perform reverse transcription. Target transcript
786 abundance in the resulting cDNA was quantified using a homemade qPCR mix as described
787 previously⁷⁵ and gene-specific primers (**Supplemental Table 7**). Each 20 uL reaction contained
788 1X Thermopol Reaction Buffer (NEB), 125uM dNTPs, 2.5mM MgSO₄, 1X SYBR Green I
789 (Lonza), 500nM gene specific or 65nM 16S rRNA primer and 0.5 units Hot Start Taq
790 Polymerase (NEB), and 10ng of template cDNA. Results were processed using the ddCT method
791 in which raw values were normalized to 16S rRNA values, then xanthan oligosaccharide values
792 were compared to those from glucose to calculate fold-change in expression.

793 For RNA-seq, total RNA was used from the *B. intestinalis* growths used for qPCR. For the
794 community grown on XG or PGA, 5 mL cultures of DM-XG or DM-PGA were inoculated with
795 a 1:100 dilution of a fully liquified DM-XG culture. PGA cultures were harvested at mid-log
796 phase at OD₆₀₀ ~0.85 whereas XG cultures were harvested at late-log phase at OD₆₀₀ ~1.2 to
797 allow liquification of XG, which was necessary to extract RNA from these cultures. As before,
798 cultures were harvested by centrifugation, mixed with RNA Protect (Qiagen) and stored at -80
799 °C until further processing. RNA was purified as before except that multiple replicates of DM-
800 XG RNA were pooled together and concentrated with Zymo RNA Clean and ConcentratorTM-25
801 to reach acceptable concentrations for RNA depletion input. rRNA was depleted twice from the
802 purified total RNA using the MICROBExpressTM Kit, each followed by a concentration step

803 using the Zymo RNA Clean and ConcentratorTM-25. About 90% rRNA depletion was achieved
804 for all samples. *B. intestinalis* RNA was sequenced using NovaSeq and community RNA was
805 sequenced using MiSeq. The resulting sequence data was analyzed for differentially expressed
806 genes following a previously published protocol⁷⁶. Briefly, reads were filtered for quality using
807 Trimmomatic v0.39⁶⁸. Reads were aligned to the each genome using BowTie2 v2.3.5.1⁷⁷. For the
808 *Bacteroides intestinalis* transcriptome reads were aligned to its genome, while for the community
809 data reads were aligned to either the *B. intestinalis* genome or the closed Ruminococcaceae
810 UCG-13 metagenome assembled genome (MAG). Reads mapping to gene features were counted
811 using htseq-count (release_0.11.1)⁷⁸. Differential expression analysis was performed using the
812 edgeR v3.34.0 package in R v.4.0.2 (with the aid of Rstudio v1.3.1093). The TMM method was
813 used for library normalization⁷⁹. Coverage data was visualized using Integrated Genome Viewer
814 (IGV)⁸⁰.

815 ***Extended metagenome analysis/comparison methodology***

816 Individual MAGs in each sample were searched by BlastP for the presence of proteins similar to
817 those encoded by the XG-degrading PUL of R. UCG13 and *B. intestinalis*. This was done using
818 the amino acid sequences of the proteins in the R. UCG13 and *B. intestinalis* PULs as the search
819 homologs; both BlastP probes were searched against all the individual MAGs in the different
820 samples with the default threshold e-value of 1e-5.

821 ***Looking for R. UCG13 and B. intestinalis. XG Loci in Metagenomes***

822 Available cohorts of human gut metagenomic sequence data (National Center for Biotechnology
823 Information projects: PRJNA422434⁸¹, PRJEB10878⁸², PRJEB12123⁸³, PRJEB12124⁸⁴,
824 PRJEB15371⁸⁵, PRJEB6997⁸⁶, PRJDB3601⁸⁷, PRJNA48479⁸⁸, PRJEB4336⁸⁹, PRJEB2054⁹⁰,
825 PRJNA392180⁹¹, and PRJNA527208⁹²) were searched for the presence of xanthan locus
826 nucleotide sequences from R. UCG13 (92.7 kb) and *B. intestinalis* (17.9kb) using the following
827 workflow: Each xanthan locus nucleotide sequence was used separately as a template and then
828 magic-blast v1.5.0⁹³ was used to recruit raw Illumina reads from the available metagenomic
829 datasets with an identity cutoff of 97%. Next, the alignment files were used to generate a
830 coverage map using bedtools v2.29.0⁹⁴ to calculate the percentage coverage of each sample
831 against each individual reference. We considered a metagenomic data sample to be positive for a
832 particular xanthan locus if it had at least 70% of the corresponding xanthan locus nucleotide
833 sequence covered.

834 The R. UCG13 locus and *B. intestinalis* XG locus were used as the query in a large-scale search
835 against the assembled scaffolds of isolates, metagenome assembled genomes (bins), and
836 metagenomes included into the Integrated Microbial Genomes & Microbiomes (IMG/M)
837 comparative analysis system⁴¹. Within the LAST software package, version 1066, the ‘lastal’
838 tool was used with default thresholds to search the 2 loci against 72,491 public high-quality
839 isolate genomes, and 102,860 bins from 13,415 public metagenomes, and 21,762 public
840 metagenomes in IMG/M. Metagenome bins were generated using the binning analysis method
841 described in A. Clum et al⁹⁵.

842 **Data availability.** All sequencing reads have been deposited at the European Nucleotide Archive
843 under BioProject PRJEB44146. All annotated MAGs are publicly available via Figshare (DOIs:
844 10.6084/m9.figshare.14494602, 10.6084/m9.figshare.14494536, 10.6084/m9.figshare.14494677,
845 10.6084/m9.figshare.14494683 and 10.6084/m9.figshare.14494689).

846

847 **REFERENCES**

- 848 1. García-Ochoa, F., Santos, V. E., Casas, J. A. & Gómez, E. Xanthan gum: Production,
849 recovery, and properties. *Biotechnol. Adv.* **18**, 549–579 (2000).
- 850 2. Chassaing, B. *et al.* Dietary emulsifiers impact the mouse gut microbiota promoting colitis
851 and metabolic syndrome. *Nature* **519**, 92–96 (2015).
- 852 3. Collins, J. *et al.* Dietary trehalose enhances virulence of epidemic *Clostridium difficile*.
853 *Nature* **553**, 291–294 (2018).
- 854 4. Laudisi, F. *et al.* The Food Additive Maltodextrin Promotes Endoplasmic Reticulum
855 Stress–Driven Mucus Depletion and Exacerbates Intestinal Inflammation. *Cmgh* **7**, 457–
856 473 (2019).
- 857 5. Etienne-Mesmin, L. *et al.* Experimental models to study intestinal microbes–mucus
858 interactions in health and disease. *FEMS Microbiol. Rev.* **43**, 457–489 (2019).
- 859 6. Casas, J. A., Santos, V. E. & García-Ochoa, F. Xanthan gum production under several
860 operational conditions: Molecular structure and rheological properties. *Enzyme Microb.*
861 *Technol.* **26**, 282–291 (2000).
- 862 7. Sworn, G. Xanthan gum. in *Handbook of Hydrocolloids* **262**, 833–853 (Elsevier, 2021).

- 863 8. King, J. A. *et al.* Incidence of Celiac Disease Is Increasing Over Time. *Am. J.*
864 *Gastroenterol.* 1 (2020). doi:10.14309/ajg.0000000000000523
- 865 9. Mortensen, A. *et al.* Re-evaluation of xanthan gum (E 415) as a food additive. *EFSA J.* **15**,
866 (2017).
- 867 10. Hehemann, J.-H., Kelly, A. G., Pudlo, N. A., Martens, E. C. & Boraston, A. B. Bacteria of
868 the human gut microbiome catabolize red seaweed glycans with carbohydrate-active
869 enzyme updates from extrinsic microbes. *Proc. Natl. Acad. Sci.* **109**, 19786–19791 (2012).
- 870 11. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data
871 processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013).
- 872 12. Goodman, A. L. *et al.* Extensive personal human gut microbiota culture collections
873 characterized and manipulated in gnotobiotic mice. *Proc. Natl. Acad. Sci. U. S. A.* **108**,
874 6252–6257 (2011).
- 875 13. Kim, C. C. *et al.* Genomic insights from *Monoglobus pectinilyticus*: a pectin-degrading
876 specialist bacterium in the human colon. *ISME J.* **13**, 1437–1456 (2019).
- 877 14. Ruijsenaars, H. J., De Bont, J. A. M. & Hartmans, S. A pyruvated mannose-specific
878 xanthan lyase involved in xanthan degradation by *Paenibacillus alginolyticus* XL-1. *Appl.*
879 *Environ. Microbiol.* **65**, 2446–2452 (1999).
- 880 15. Nankai, H., Hashimoto, W., Miki, H., Kawai, S. & Murata, K. Microbial system for
881 polysaccharide depolymerization: Enzymatic route for xanthan depolymerization by
882 *Bacillus* sp. strain GL1. *Appl. Environ. Microbiol.* **65**, 2520–2526 (1999).
- 883 16. Hashimoto, W., Nankai, H., Mikami, B. & Murata, K. Crystal structure of *Bacillus* sp.
884 GL1 xanthan lyase, which acts on the side chains of xanthan. *J. Biol. Chem.* **278**, 7663–
885 7673 (2003).
- 886 17. Jensen, P. F. *et al.* Structure and Dynamics of a Promiscuous Xanthan Lyase from
887 *Paenibacillus nanensis* and the Design of Variants with Increased Stability and Activity.
888 *Cell Chem. Biol.* **26**, 191-202.e6 (2019).
- 889 18. Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H. & Henrissat, B. Evolution,

- 890 substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5).
891 *BMC Evol. Biol.* **12**, (2012).
- 892 19. Jongkees, S. A. K. & Withers, S. G. Unusual enzymatic glycoside cleavage mechanisms.
893 *Acc. Chem. Res.* **47**, 226–235 (2014).
- 894 20. Rovira, C., Males, A., Davies, G. J. & Williams, S. J. Mannosidase mechanism: at the
895 intersection of conformation and catalysis. *Curr. Opin. Struct. Biol.* **62**, 79–92 (2020).
- 896 21. Kool, M. M. *et al.* Characterization of an acetyl esterase from *Myceliophthora*
897 *thermophila* C1 able to deacetylate xanthan. *Carbohydr. Polym.* **111**, 222–229 (2014).
- 898 22. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using
899 deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).
- 900 23. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W. & Brumer, H. Polysaccharide
901 utilization loci: Fueling microbial communities. *J. Bacteriol.* **199**, 1–15 (2017).
- 902 24. Pilgaard, B., Vuillemin, M., Holck, J., Wilkens, C. & Meyer, A. S. Specificities and
903 synergistic actions of novel PL8 and PL7 alginate lyases from the marine fungus
904 *Paradendryphiella salina*. *J. Fungi* **7**, 1–16 (2021).
- 905 25. Zhu, B. & Yin, H. Alginate lyase: Review of major sources and classification, properties,
906 structure-function analysis and applications. *Bioengineered* **6**, 125–131 (2015).
- 907 26. Terrapon, N. *et al.* PULDB: The expanded database of Polysaccharide Utilization Loci.
908 *Nucleic Acids Res.* **46**, D677–D683 (2018).
- 909 27. Sun, Z., Liu, H., Wang, X., Yang, F. & Li, X. Proteomic Analysis of the Xanthan-
910 Degrading Pathway of *Microbacterium* sp. XT11. *ACS Omega* **4**, 19096–19105 (2019).
- 911 28. Yang, F. *et al.* Novel Endotype Xanthanase from Xanthan-Degrading *Microbacterium*
912 *Microbacterium* sp. Strain XT11. **85**, 1–16 (2019).
- 913 29. Artzi, L., Bayer, E. A. & Morais, S. Cellulosomes: Bacterial nanomachines for
914 dismantling plant polysaccharides. *Nat. Rev. Microbiol.* **15**, 83–95 (2017).
- 915 30. Guillén, D., Sánchez, S. & Rodríguez-Sanoja, R. Carbohydrate-binding domains:
916 Multiplicity of biological roles. *Appl. Microbiol. Biotechnol.* **85**, 1241–1249 (2010).

- 917 31. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**,
918 D412–D419 (2021).
- 919 32. Ebbes, M. *et al.* Fold and function of the InlB B-repeat. *J. Biol. Chem.* **286**, 15496–15506
920 (2011).
- 921 33. Bley Müller, W. M. *et al.* MET-activating residues in the B-repeat of the *Listeria*
922 *monocytogenes* invasion protein InlB. *J. Biol. Chem.* **291**, 25567–25577 (2016).
- 923 34. Kool, M. M., Gruppen, H., Sworn, G. & Schols, H. A. Comparison of xanthans by the
924 relative abundance of its six constituent repeating units. *Carbohydr. Polym.* **98**, 914–921
925 (2013).
- 926 35. Moroz, O. V. *et al.* Structural Dynamics and Catalytic Properties of a Multi - Modular
927 Xanthanase. *ACS Catal.* **8**, 6021–6034 (2018).
- 928 36. Yang, F. *et al.* Production and purification of a novel xanthan lyase from a xanthan-
929 degrading microbacterium sp. Strain XT11. *Sci. World J.* **2014**, (2014).
- 930 37. Glenwright, A. J. *et al.* Structural basis for nutrient acquisition by dominant members of
931 the human gut microbiota. *Nature* **541**, 407–411 (2017).
- 932 38. Gregg, K. J. *et al.* Analysis of a new family of widely distributed metal-independent α -
933 mannosidases provides unique insight into the processing of N-linked glycans. *J. Biol.*
934 *Chem.* **286**, 15586–15596 (2011).
- 935 39. Daly, J., Tomlin, J. & Read, N. W. The effect of feeding xanthan gum on colonic function
936 in man: correlation with in vitro determinants of bacterial breakdown. *Br. J. Nutr.* **69**,
937 897–902 (1993).
- 938 40. Kiełbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame
939 genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
- 940 41. Chen, I. M. A. *et al.* The IMG/M data management and analysis system v.6.0: New tools
941 and advanced capabilities. *Nucleic Acids Res.* **49**, D751–D763 (2021).
- 942 42. Liang, R. *et al.* Metabolic capability of a predominant *Halanaerobium* sp. in hydraulically
943 fractured gas wells and its implication in pipeline corrosion. *Front. Microbiol.* **7**, 1–10

- 944 (2016).
- 945 43. Schnizlein, M. K., Vendrov, K. C., Edwards, S. J., Martens, E. C. & Young, V. B. Dietary
946 xanthan gum alters antibiotic efficacy against the murine gut microbiota and attenuates
947 *Clostridioides difficile* colonization. *bioRxiv* **5**, 1–10 (2019).
- 948 44. Katzbauer, B. Properties and applications of xanthan gum. *Polym. Degrad. Stab.* **59**, 81–
949 84 (1998).
- 950 45. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D.
951 Development of a dual-index sequencing strategy and curation pipeline for analyzing
952 amplicon sequence data on the miseq illumina sequencing platform. *Appl. Environ.*
953 *Microbiol.* **79**, 5112–5120 (2013).
- 954 46. Schloss, P. D. *et al.* Introducing mothur: Open-source, platform-independent, community-
955 supported software for describing and comparing microbial communities. *Appl. Environ.*
956 *Microbiol.* **75**, 7537–7541 (2009).
- 957 47. Team, R. C. R: A language and environment for statistical computing. (2020).
- 958 48. Wickham, H. Reshaping Data with the reshape Package. *J. Stat. Softw.* **21**, 1–20 (2007).
- 959 49. Neuwirth, E. RColorBrewer: ColorBrewer Palettes. (2014). Available at: [https://cran.r-](https://cran.r-project.org/package=RColorBrewer)
960 [project.org/package=RColorBrewer](https://cran.r-project.org/package=RColorBrewer).
- 961 50. Wickham, H. *Elegant Graphics for Data Analysis: ggplot2. Applied Spatial Data Analysis*
962 *with R* (2008).
- 963 51. Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by
964 two human gut symbionts. *PLoS Biol.* **9**, (2011).
- 965 52. Pope, P. B. *et al.* Isolation of Succinivibrionaceae implicated in low methane emissions
966 from Tammar wallabies. *Science (80-.)*. **333**, 646–648 (2011).
- 967 53. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
968 *EMBnet.journal* **17**, 10 (2011).
- 969 54. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: A new versatile
970 metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).

- 971 55. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately
972 reconstructing single genomes from complex microbial communities. *PeerJ* **2015**, 1–15
973 (2015).
- 974 56. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM:
975 Assessing the quality of microbial genomes recovered from isolates, single cells, and
976 metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
- 977 57. Chen, I. M. A. *et al.* IMG/M: Integrated genome and metagenome comparative data
978 analysis system. *Nucleic Acids Res.* **45**, D507–D516 (2017).
- 979 58. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The
980 carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, 490–495
981 (2014).
- 982 59. Rodriguez-R, L. M. *et al.* The Microbial Genomes Atlas (MiGA) webserver: Taxonomic
983 and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic*
984 *Acids Res.* **46**, W282–W288 (2018).
- 985 60. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: A toolkit to
986 classify genomes with the genome taxonomy database. *Bioinformatics* **36**, 1925–1927
987 (2020).
- 988 61. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–
989 2069 (2014).
- 990 62. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k -mer
991 weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
- 992 63. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–
993 3100 (2018).
- 994 64. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome
995 assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
- 996 65. Seppey, M., Manni, M. & Zdobnov, E. M. *BUSCO: Assessing Genome Assembly and*
997 *Annotation Completeness BT - Gene Prediction: Methods and Protocols.* (2019).

- 998 66. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High
999 throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.
1000 *Nat. Commun.* **9**, 1–8 (2018).
- 1001 67. Kunath, B. J. *et al.* From proteins to polysaccharides: lifestyle and genetic evolution of
1002 *Coprothermobacter proteolyticus*. *ISME J.* **13**, 603–617 (2019).
- 1003 68. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina
1004 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- 1005 69. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: Fast and accurate filtering of ribosomal
1006 RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
- 1007 70. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq
1008 quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
- 1009 71. Massie, H. R. & Zimm, B. H. THE USE OF HOT PHENOL IN PREPARING DNA.
1010 *Proc. Natl. Acad. Sci.* **54**, 1641–1643 (1965).
- 1011 72. Nie, X. Relationships between dietary fiber structural features and growth and utilization
1012 patterns of human gut bacteria. *ProQuest Diss. Theses* 136 (2016).
- 1013 73. Tuncil, Y. E., Thakkar, R. D., Marcia, A. D. R., Hamaker, B. R. & Lindemann, S. R.
1014 Divergent short-chain fatty acid production and succession of colonic microbiota arise in
1015 fermentation of variously-sized wheat bran fractions. *Sci. Rep.* **8**, 1–13 (2018).
- 1016 74. Arnal, G., Attia, M. A., Asohan, J. & Brumer, H. A Low-Volume, Parallel Copper-
1017 Bicinchoninic Acid (BCA) Assay for Glycoside Hydrolases. in *Protein-Carbohydrate*
1018 *Interactions. Methods and Protocols* (eds. Abbott, D. W. & Lammerts van Bueren, A.)
1019 **1588**, 209–214 (Springer New York, 2017).
- 1020 75. Speer, M. A. DEVELOPMENT OF A GENETICALLY MODIFIED SILAGE
1021 INOCULANT FOR THE BIOLOGICAL PRETREATMENT OF LIGNOCELLULOSIC
1022 BIOMASS. (Pennsylvania State University, 2013).
- 1023 76. Anders, S. *et al.* Count-based differential expression analysis of RNA sequencing data
1024 using R and Bioconductor. *Nat. Protoc.* **8**, 1765–1786 (2013).

- 1025 77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
1026 **9**, 357–359 (2012).
- 1027 78. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-
1028 throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
- 1029 79. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for
1030 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–
1031 140 (2009).
- 1032 80. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV):
1033 High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–
1034 192 (2013).
- 1035 81. Wang, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes.
1036 *Nature* **490**, 55–60 (2012).
- 1037 82. Yu, J. *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-
1038 invasive biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).
- 1039 83. Liu, R. *et al.* Gut microbiome and serum metabolome alterations in obesity and after
1040 weight-loss intervention. *Nat. Med.* **23**, 859–868 (2017).
- 1041 84. Gu, Y. *et al.* Analyses of gut microbiota and plasma bile acids enable stratification of
1042 patients for antidiabetic treatment. *Nat. Commun.* **8**, (2017).
- 1043 85. He, Q. *et al.* Two distinct metacommunities characterize the gut microbiota in Crohn’s
1044 disease patients. *Gigascience* **6**, 1–11 (2017).
- 1045 86. Zhang, X. *et al.* The oral and gut microbiomes are perturbed in rheumatoid arthritis and
1046 partly normalized after treatment. *Nat. Med.* **21**, 895–905 (2015).
- 1047 87. Nishijima, S. *et al.* The gut microbiome of healthy Japanese and its microbial and
1048 functional uniqueness. *DNA Res.* **23**, 125–133 (2016).
- 1049 88. Lloyd-Price, J. *et al.* Strains, functions and dynamics in the expanded Human Microbiome
1050 Project. *Nature* **550**, 61–66 (2017).
- 1051 89. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic

- 1052 markers. *Nature* **500**, 541–546 (2013).
- 1053 90. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic
1054 sequencing. *Nature* **464**, 59–65 (2010).
- 1055 91. Smits, S. A. *et al.* Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers
1056 of Tanzania. *Science* (80-.). **357**, 802–805 (2017).
- 1057 92. Conteville, L. C., Oliveira-Ferreira, J. & Vicente, A. C. P. Gut microbiome biomarkers
1058 and functional diversity within an Amazonian semi-nomadic hunter-gatherer group. *Front.*
1059 *Microbiol.* **10**, 1–10 (2019).
- 1060 93. Boratyn, G. M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B. & Madden, T. L. Magic-
1061 BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics* **20**,
1062 1–19 (2019).
- 1063 94. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing
1064 genomic features. *Bioinformatics* **26**, 841–842 (2010).
- 1065 95. Clum, A. *et al.* The DOE JGI Metagenome Workflow. *bioRxiv* (2020).
1066 doi:<https://doi.org/10.1101/2020.09.30.320929>

1067

1068 **FIGURE LEGENDS**

1069 **Figure 1.** R. UCG13 was a common factor across xanthan gum degrading cultures.

1070 **a**, Xanthan gum is a repeating structure of glucose (blue circles), mannose (green circles), and
1071 glucuronic acid (blue and white diamond). The inner and outer mannose residues are variably
1072 modified by acetylation and pyruvylation, respectively.

1073 **b**, Growth curves of the original xanthan-degrading culture showed greater culture density as
1074 xanthan gum concentration was increased (n=12, SEM \leq 3%), and, **c**, displayed relatively stable
1075 composition over sequential passaging. Passaging the culture on BHI-blood plates resulted in a
1076 loss of R. UCG13 as well as xanthan degrading activity.

1077 **d**, An additional 20 samples were sequentially passaged in xanthan containing media (10x) and
1078 analyzed for composition by 16S rRNA sequencing (16 of the most abundant genus are
1079 displayed for clarity). All cultures shared an abundant OTU, classified as R. UCG13.

1080

1081 **Figure 2.** Metagenomics, metatranscriptomics, and activity-guided proteomics identified two
1082 putative xanthan gum degrading loci.

1083 Putative xanthan utilization loci color-coded and annotated by predicted protein family. The four
1084 boxes below each gene are colored to represent expression levels of each gene at timepoints
1085 taken throughout the culture's growth on xanthan gum. MAG taxonomy is indicated in
1086 parentheses.

1087

1088 **Figure 3.** *R. UCG13* encodes a novel GH5 that depolymerizes native xanthan gum.

1089 **a,** Extracted ion chromatograms showing various acetylated and pyruvylated penta- and deca-
1090 saccharides produced by incubating culture supernatant with XG.

1091 **b,** Annotated domains of the xanthan-degrading *RuGH5a*, showing its signal peptide (SP), three
1092 carbohydrate binding modules (CBMs), and multiple *Listeria*-*Bacteroides* repeat domains.

1093 **c,** Proton NMR contrasting tetrasaccharide products obtained from incubating lyase-treated
1094 xanthan gum with either *RuGH5a* or *P. nanensis* GH9.

1095 **d,** Kinetics of *RuGH5a* on native and lyase-treated xanthan gum (error bars represent mean and
1096 standard deviation, n=4)

1097

1098 **Figure 4.** *B. intestinalis* cross-feeds on xanthan oligosaccharides.

1099 **a,** Growth curves of *B. intestinalis* isolated from the original xanthan-degrading culture. (curves
1100 represent mean and SEM, n=2).

1101 **b,** Fold-change in expression of *B. intestinalis* genes when grown on xanthan oligosaccharides
1102 relative to glucose.

1103

1104

1105 **Figure 5.** Xanthan degrading loci are present in modern human microbiomes but not in hunter-
1106 gatherers'.

1107 Multiple microbiome metagenome datasets were searched for the presence or absence of the *R.*
1108 *UCG13* and *B. intestinalis* xanthan loci. Map colors correspond to where populations were
1109 sampled for each dataset displayed on the outside of the figure. Circle segments are sized
1110 proportionately to total number of individuals sampled for each dataset. Lines represent presence

1111 of either the R. UCG13 xanthan locus (green) or the *B. intestinalis* xanthan locus (red).
1112 Percentages display the total abundance of R. UCG13 or *B. intestinalis* locus in each dataset.

1113

1114 **Extended Data 1.** Cellular model of xanthan degradation

1115

1116 **Extended Data 2.** Xanthan degradation is a multi-species phenotype.

1117 An active xanthan culture was diluted in 2x defined media without a major carbon source, then
1118 divided and diluted 1:1 with either 2x xanthan gum or 2x monosaccharide mix (2:2:1
1119 mannose;glucose;glucuronic acid), then aliquoted into 200 uL cultures in 96-well plates. Each
1120 datapoint represents the fraction of cultures (out of 96) growing above OD₆₀₀ 0.7 at each
1121 dilution, grown in either xanthan gum or monosaccharide mix media. Data were fit to the Hill
1122 equation to calculate a 50% growth dilution factor (GDF 50) at which half of the cultures would
1123 grow above OD 0.7. Across 5 independent experiments, there was a GDF50 difference of 1.8
1124 (standard deviation = 0.4, SEM = 0.2). This demonstrates that at comparable dilutions, microbes
1125 were present that could grow on monosaccharides but were unable to grow using XG, suggesting
1126 that several microbes are required in this media to allow growth on XG.

1127

1128 **Extended Data 3.** Neutral monosaccharide and metatranscriptomic analysis

1129 a, Two replicates of the original xanthan culture were grown and sampled at multiple timepoints
1130 for **b**, neutral monosaccharide analysis of residual xanthan gum (n=3, standard deviation shown)
1131 and **c**, metatranscriptomic analysis of annotated CAZymes in each of the MAGs (completeness
1132 value > 75%) reconstructed from metagenomic data from the enrichment culture. MAG
1133 taxonomy (**Supplementary Table 2**) is indicated in parentheses. An “#” indicates a low
1134 AAI%.

1135

1136 **Extended Data 4.** Culture supernatant contains enzymes capable of depolymerizing xanthan
1137 gum, while intracellular contents are required for complete saccharification.

1138 Thin layer chromatography of xanthan gum incubated with different fractions of an active
1139 xanthan gum culture (supernatant, washed cell pellet, lysed cell pellet, or lysed culture).

1140 Negative controls were prepared by heating fractions at 95 C for 15 minutes prior to initiating
1141 with xanthan gum. EDTA was added to a final concentration of ~50 mM to determine the

1142 necessity of divalent cations for enzyme activity. Strong color development in circles at baseline
1143 is undigested polysaccharide while bands that migrated with solvent are digested
1144 oligosaccharides and monosaccharides.

1145

1146 **Extended Data 5.** Activity-guided fractionation and proteomics narrow list of potential
1147 xanthanases.

1148 **a**, Venn diagram depicting activity-guided fractionation of culture supernatants followed by
1149 proteomic identification of candidate proteins present across all active preparations.

1150 **b**, SDS-PAGE of one of the fractionated culture supernatants (ANX processed sample)
1151 submitted for proteomic analysis (without and with boiling prior to analysis). Protein complexes
1152 or fragments that are larger are retained at the top of the gel while smaller proteins migrate
1153 towards the bottom of the gel. The ladder on the left shows how 180, 115, and 82 kDa standards
1154 are retained by the gel matrix. Boiling results in separation of protein complexes that cause
1155 streaking in the first lane and resolution into single protein bands that are denatured and migrate
1156 with respect to size.

1157 **c**, Proteomics narrows potential xanthanases to 33 proteins, 22 of which were from *R. UCG13*.
1158 Final candidates were obtained by collating proteins that were present across all three activity-
1159 based fractionation experiments with proteomic identification. Colors assist with visualizing
1160 number of peptides associated with each protein at different thresholds (<5, red; 5-24, orange;
1161 25-74, blue; ≥ 75 , green)

1162

1163 **Extended Data 6. Activity of *R. UCG13* GH5 enzymes on various polysaccharides.**

1164 **a**, SDS-PAGE gel of purified GH5 constructs and their resultant activity as assessed by TLC on

1165 **b**, xanthan gum, **b-c**, carboxymethyl cellulose (CMC), **c**, hydroxyethyl cellulose (HEC), **d**,

1166 barley β -glucan, **d-e**, yeast β -glucan, **e**, tamarind xyloglucan, **f**, xylan, and **f-g**, wheat

1167 arabinoxylan. Enzymes are 1, *RuGH5b* (GH5 only); 2, *RuGH5b* (GH5 with CBM-A); 3,

1168 *RuGH5b* (GH5 with CBM-A/B); 4, *RuGH5b* (full protein); 5, *RuGH5a* (GH5 only); 6, *RuGH5a*

1169 (GH5 with CBM-A); 7, *RuGH5a* (GH5 with CBM-A/B); 8, *RuGH5a* (GH5 with CBM-A/B/C);

1170 9, *RuGH5a* (full protein); 10, replicate of 8. Strong color development in circles at baseline is

1171 undigested polysaccharide while bands or streaking that migrated with solvent are digested

1172 oligosaccharides and monosaccharides. Although minor streaking appears in some substrates due

1173 to residual oligosaccharides, comparing untreated substrate with enzyme incubated substrate
1174 allows determination of enzyme activity. *RuGH5a* constructs with all 3 CBMs (8-10) showed
1175 clear activity on XG.

1176

1177 **Extended Data 7.** NMR of tetrasaccharide produced by *RuGH5a* (and PL8 xanthan lyase).

1178 **a,** ^{13}C -NMR spectrum of **tetrasaccharide** in D_2O (^{13}C : 150 MHz).

1179 **b,** COSY (^1H - ^1H) spectrum of **tetrasaccharide** in D_2O (600 MHz).

1180 **c,** HSQC (^1H - ^{13}C) spectrum of **tetrasaccharide** in D_2O . Red contours represent CH and CH_3
1181 groups, blue contours represent CH_2 groups.

1182 **d,** HMBC (^1H - ^{13}C) spectrum of **tetrasaccharide** in D_2O .

1183 **e,** Selective 1H 1D-TOCSY spectrum of **tetrasaccharide** in D_2O (^1H : 600 MHz). Starred peak
1184 indicates the frequency irradiated (5.83 ppm) and arrow on the structure illustrates corresponding
1185 proton position being irradiated (Δ 4,5-GlcA H5).

1186 **f,** Selective 1H 1D-TOCSY spectrum of **tetrasaccharide** in D_2O (^1H : 600 MHz). Starred peak
1187 indicates the frequency irradiated (5.35 ppm) and arrow on the structure illustrates corresponding
1188 proton position being irradiated (Man H1).

1189 **g,** Selective 1H 1D-TOCSY spectrum of **tetrasaccharide** in D_2O (^1H : 600 MHz). Starred peak
1190 indicates the frequency irradiated (4.27 ppm) and arrow on the structure illustrates corresponding
1191 proton position being irradiated (Man H2).

1192 **h,** Selective 1H 1D-TOCSY spectrum of **tetrasaccharide** in D_2O (^1H : 600 MHz). Starred peak
1193 indicates the frequency irradiated (4.52 ppm) and arrow on the structure illustrates corresponding
1194 proton position being irradiated (Glc H1).

1195 **i,** Selective 1H 1D-TOCSY spectrum of **tetrasaccharide** in D_2O (^1H : 600 MHz). Starred peak
1196 indicates the frequency irradiated (4.64 ppm) and arrow on the structure illustrates corresponding
1197 proton position being irradiated (β -Glc H1).

1198 **j,** Selective 1H 1D-TOCSY spectrum of **tetrasaccharide** in D_2O (^1H : 600 MHz). Starred peak
1199 indicates the frequency irradiated (5.20 ppm) and arrow on the structure illustrates corresponding
1200 proton position being irradiated (α -Glc H1).

1201

1202 ^1H -NMR analysis illustrated an inconsistent spectrum to that of the known degradation product
1203 from other xanthanases (including GH9 from *Paenibacillus nanensis* or the β -D-glucanase in

1204 *Bacillus* sp. Strain GL1), that hydrolyze xanthan at the reducing end of the branching glucose.
1205 LCMS analysis was consistent with a **tetrasaccharide** containing a Δ 4,5-ene-GlcA moiety, but
1206 despite the appropriate mass, these differences in the $^1\text{H-NMR}$ suggested an alternative cut site.
1207 To confirm, full structural elucidation was conducted by extensive NMR-analysis. In a similar
1208 fashion to Wilson and coworkers, spin systems for each monosaccharide were established via
1209 selective 1D-TOCSY experiments, selectively irradiating individual anomeric protons between
1210 δ_{H} 4.52 and δ_{H} 5.35, and the one vinylic proton of the Δ 4,5-ene-GlcA residue at δ_{H} 5.83. This
1211 vinylic proton was easily identified by HSQC analysis via its distinct $^1J_{\text{H,C}}$ correlation (δ_{C} 107.9
1212 ppm), and its HMBC correlations to C-5 (δ_{C} 144.4 ppm) and C-6 (δ_{C} 168.8 ppm) of Δ 4,5-ene-
1213 GlcA. This proton was used as a starting point for structural elucidation, and in conjunction with
1214 the data obtained from 2D-HSQC and selective 1D-TOCSY experiments allowed for
1215 identification of the remainder of the Δ 4,5-ene-GlcA spin system, including the anomeric
1216 position (δ_{H} 5.06, δ_{C} 99.2). Further HMBC analysis identified a correlation between H-1 of Δ 4,5-
1217 ene-GlcA and the C-2 position of Man (δ_{C} 76.4 ppm). The inverse HMBC correlation was also
1218 observed from H-2 of the Man residue (δ_{H} 4.27) to the anomeric carbon of Δ 4,5-ene-GlcA (δ_{C}
1219 99.2), confirming the expected connectivity through a 1 \rightarrow 2 linkage. COSY analysis identified a
1220 correlation between H-2 and H-1 (δ_{H} 5.35) of the of the α -Man residue, and as was the case for
1221 Δ 4,5-ene-GlcA, the remainder of the positions were assigned from 2D-HSQC and selective 1D-
1222 TOCSY experiments. Interestingly, the anomeric position of α -Man appeared as a singlet, in
1223 contrast to the reported two anomeric proton signals associated with this residue in the
1224 tetrasaccharide isolated by Wilson. HMBC analysis from this anomeric position showed
1225 correlations to C-2, C-3, and C-5 (δ_{C} 76.4, δ_{C} 69.4, and δ_{C} 72.5 respectively) of Man. An
1226 additional correlation was observed to a carbon external to the Man subunit, with a chemical
1227 shift of 82.5 ppm. This shift was identified as belonging to the C-3 position of a nonreducing
1228 glucosyl residue. This was confirmed via HMBC correlations from both H-2 and H-4 of Glc(n.r).
1229 The H-2 peak was free of any overlap in the $^1\text{H-NMR}$ spectrum, allowing for unambiguous
1230 assignment through a COSY correlation between itself and H-1 (δ_{H} 3.37 and δ_{H} 4.52
1231 respectively), as well as the H-3 proton with a chemical shift of 3.63 ppm. This gave us
1232 confidence that the α -Man residue was connected via a 1 \rightarrow 3 linkage to this Glc(n.r) subunit.
1233 Importantly, the anomeric proton of the Glc(n.r) residue appeared as a single doublet, integrating
1234 with a value of one in the $^1\text{H-NMR}$ spectrum with a coupling constant of 8.0 Hz, consistent with

1235 a β -configured Glc(n.r) monomer. This confirmed connectivity between the α -Man and β -
1236 Glc(n.r) residues, suggesting a disparate structure to the previously reported degradation product.
1237 Finally, a key HMBC correlation was observed from the anomeric proton (δ_{H} 4.52) to an external
1238 carbon with a ^{13}C -chemical shift of ~ 79 ppm. Upon closer inspection, this carbon was actually
1239 two separate peaks, corresponding to the C-4 position of the alpha (minor) and beta (major)
1240 anomers (δ_{C} 79.0 and δ_{C} 78.8 respectively) of the reducing Glc. This confirmed the expected 1 \rightarrow
1241 4 linkage of the backbone Glc residues, albeit illustrating hydrolysis had occurred at the reducing
1242 end of the nonbranching Glc. To complete structural elucidation, the remainder of the positions
1243 were assigned from 2D-HSQC and selective 1D-TOCSY experiments for both the alpha and beta
1244 anomers separately.

1245
1246 **Extended Data 8.** Activity of *R. UCG13* and *B. intestinalis* enzymes.
1247 LC-MS analysis was used to track relative increases and decreases of intermediate
1248 oligosaccharides with the addition of enzymes, verifying their abilities to successively cleave XG
1249 pentasaccharides to their substituent monosaccharides. Integrated extracted ion counts (n=4,
1250 SEM) that correlate with compound abundance are shown for **a**, acetylated pentasaccharide (M-
1251 H ions: 883.26, 953.26, 925.27), **b**, deacetylated pentasaccharide (M-H ions: 841.25, 911.25), **c**,
1252 acetylated tetrasaccharide (2M-H ion: 1407.39), **d**, tetrasaccharide (M-H ion: 661.18), **e**,
1253 acetylated trisaccharide (M+Cl ion: 581.15), **f**, trisaccharide (M+Cl ion: 539.14), **g**, cellobiose
1254 (M+Cl ion: 377.09), and **h**, pyruvylated mannose (M-H ion: 249.06). Reactions were carried out
1255 using xanthan oligosaccharides produced by the *RuGH5a* to test activities of the *R. UCG13* (A-I)
1256 and *B. intestinalis* (J-O) enzymes. *R. UCG13* enzymes were tested in reactions that included (A)
1257 no enzyme, (B) *R. UCG13* CE-A, (C) *R. UCG13* CE-B, (D) *R. UCG13* PL8, (E) *R. UCG13* PL8
1258 and CE-A, (F) *R. UCG13* PL8 and CE-B, (G) *R. UCG13* PL8, both CEs, and GH88, (H) *R.*
1259 *UCG13* PL8, both CEs, GH88, and GH38-A, (I) *R. UCG13* PL8, both CEs, GH88, and GH38-B.
1260 *B. intestinalis* enzymes were tested in reactions that included (J) no enzyme, (K) Bi PL-only, (L)
1261 Bi PL-CE, (M) Bi PL-CE and Bacillus PL8, (N) Bi PL-CE and GH88 and Bacillus PL8, (O) Bi
1262 PL-CE, GH88, and GH92 and Bacillus PL8.
1263 **i**, Legend of enzymes included in each reaction.

1264

1265 Recombinant enzymes were purified and analyzed for expression and purity by SDS-PAGE.
1266 Proteins generally expressed well with a single dense band indicating overexpression of the
1267 enzyme at its predicted molecular weight as compared to a size ladder. Exceptions included the
1268 R. UCG13 GH88 and CE-A, both of which had bands at the predicted enzyme size but also
1269 showed bands of comparable density at other sizes resulting from either proteolysis or co-
1270 purification of undesired *E. coli* proteins.

1271 **j**, SDS-PAGE gel of purified enzymes with 4.5 µg loaded, including (1-2) ladder, (3) *B.*
1272 *intestinalis* GH3, (4) *B. intestinalis* GH5, (5) *B. intestinalis* PL-only, (6) *B. intestinalis* PL-CE,
1273 (7) *B. intestinalis* GH88, (8) *B. intestinalis* GH92, (9) R. UCG 13 GH38-A, (10) R. UCG13
1274 GH38-B, (11) R. UCG13 GH94, (12) R. UCG13 PL8, (13) R. UCG13 CE-A.

1275
1276 **k**, SDS-PAGE gel of purified enzymes with 4.5 µg loaded, including (1) ladder, (2) *B.*
1277 *intestinalis* PL-only, (3) *B. intestinalis* PL-CE, (4) *B. intestinalis* GH88, (5) *B. intestinalis* GH92,
1278 (6) R. UCG13 GH38-A, (7) R. UCG13 GH38-B, (8) R. UCG13 CE-A, (9) R. UCG13 GH88,
1279 (10) R. UCG13 CE-B, (11) R. UCG13 PL8.

1280
1281 **l**, TLC analysis showed that R. UCG13 GH94 and *B. intestinalis* GH3 are active on
1282 cellobiose. From left to right lane show (A) *RuGH5b* (full protein), (B) *RuGH5a* (full protein),
1283 (C) *B. intestinalis* GH3, (D) *B. intestinalis* GH5, (E) R. UCG13 GH94, (F) odd standards, (G)
1284 even standards, (H) cellobiose. Odd and even standards are maltooligosaccharides with 1, 3, 5,
1285 and 7 hexoses or 2, 4, and 6 hexoses, respectively. While the *B. intestinalis* GH3 only produced
1286 one product, the R. UCG13 GH94 produced two, one matching the approximate R_f of glucose
1287 while the other had a much lower R_f which presumably is phosphorylated glucose (matching the
1288 known phosphorylase activity of the GH94 family).

1289

1290 **Extended Data 9.**

1291 **a**, Traces of RNA-seq expression data from triplicates of the original culture grown on either XG
1292 or polygalacturonic acid (PGA), illustrating overexpression of the XG PUL. **b**, *Bacteroides*
1293 *clarus* and **c**, *Parabacteroides distasonis* isolated from the original culture did not grow on
1294 XGOs. **d**, *Bacteroides intestinalis* did not grow on tetramer generated with *P. nanensis* GH9 and
1295 PL8 (Psp Tetramer) even in the presence of 1 mg/mL *RuGH5a* generated XGOs to activate the

1296 PUL. Growth on glucose confirmed that the Psp Tetramer was not inherently toxic to cells. All
1297 substrates were used at 5 mg/mL unless otherwise noted. Growths are $n \geq 2$, error bars show
1298 SEM (in most cases, smaller than the marker). **e**, Traces of RNA-seq expression data from
1299 triplicates of *B. intestinalis* grown on either glucose (Glc) or XG oligosaccharides (XGOs),
1300 illustrating overexpression of the XGO PUL.

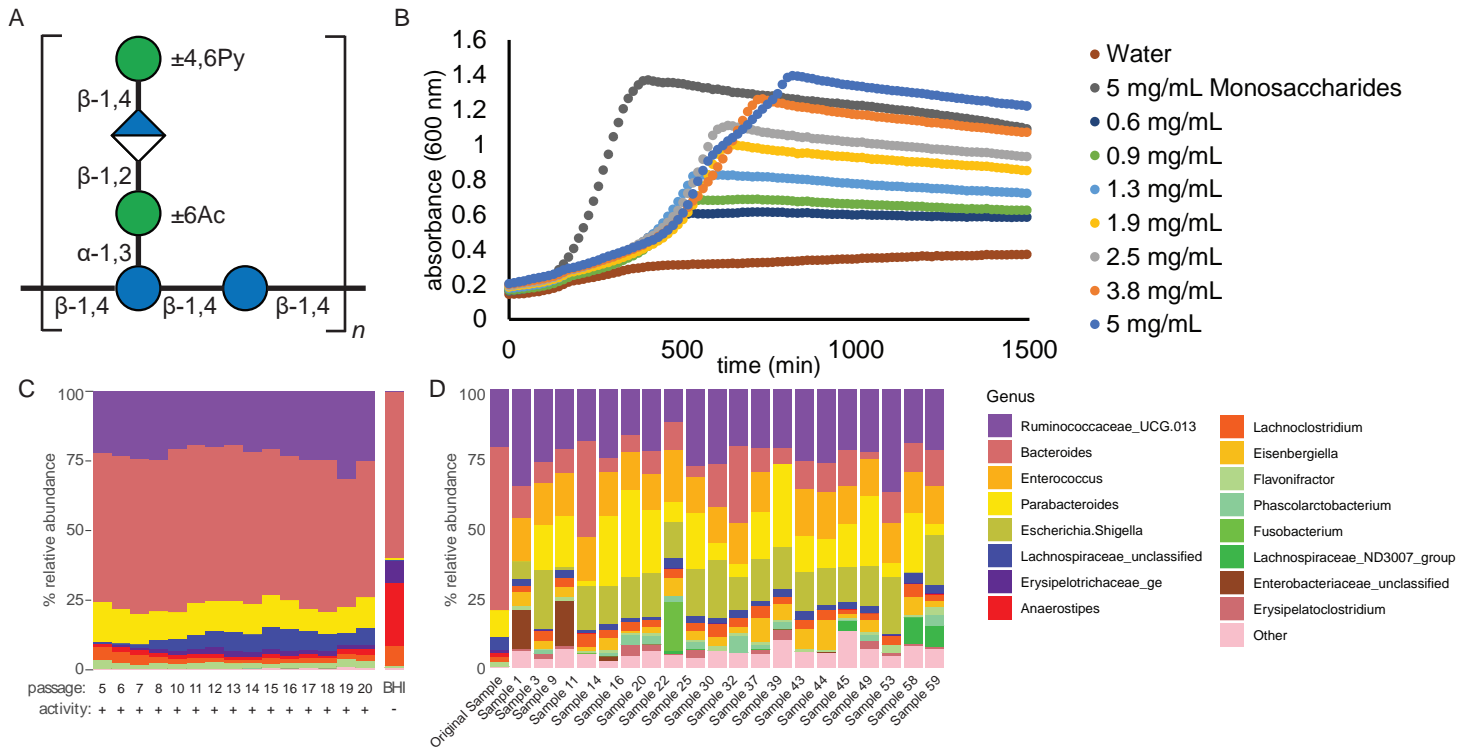
1301

1302 **Extended Data 10.**

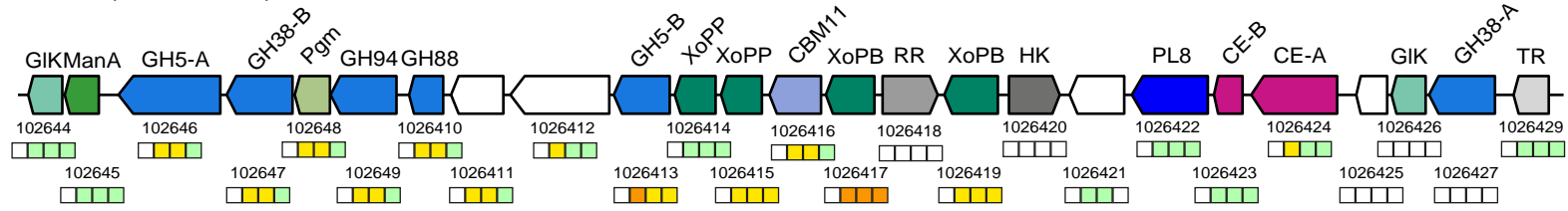
1303 **a**, Metagenomic sequencing of additional 16 cultures (S, human fecal sample) that actively grew
1304 on and degraded xanthan gum revealed two architectures of the R. UCG13. The more prevalent
1305 locus contained a GH125 insertion. The 10 additional samples with this locus architecture
1306 include: S22, S25, S39, S43, S44, S45, S49, S53, S58, and S59. **b**, The *B. intestinalis* xanthan
1307 locus was present in 3 additional cultures. **c**, Additional members of the Bacteroidaceae family
1308 harbor a PUL with a GH88, GH92 and GH3 that could potentially enable utilization of XG-
1309 oligosaccharides.

1310 **d**, The GH125-containing version of the R. UCG13 xanthan locus was detected in two mouse
1311 fecal samples (M, mouse fecal sample). **e**, Comparison of the human and mouse *RuGH5a* aa
1312 sequence, showing the annotated signal peptide (SP), GH5 domain, three carbohydrate binding
1313 modules (CBMs), and multiple Listeria-Bacteroides repeat domains. **f**, Genetic organization and
1314 aa identity (%) between the *B. intestinalis* xanthan locus in the original human sample and a PUL
1315 detected in a fracking water microbial community (FWMC) using LAST-searches. **g**, SDS-
1316 PAGE gel of purified enzymes with 4.5 μ g loaded, including ladder and the different mouse
1317 *RuGH5a* constructs. A, B, and C are all versions of the GH5 domain alone, D is a construct
1318 designed to terminate at a site homologous to the last CBM in the human *RuGH5a*, and E is a
1319 full-length construct of the mouse *RuGH5a*. **h**, TLC of each mouse *RuGH5a* construct incubated
1320 with XG and also odd (1, 3, 5, and 7 residues) and even (2, 4, and 6 residues) malto-
1321 oligosaccharide standards. The GH5-only constructs did not degrade XG, but constructs D and E
1322 (with regions homologous to the human *RuGH5a* CBMs) were able to hydrolyze XG.

Figure 1



MAG2 (*R. UCG13*) XG PUL



MAG1 (*B. intestinalis*) XG PUL

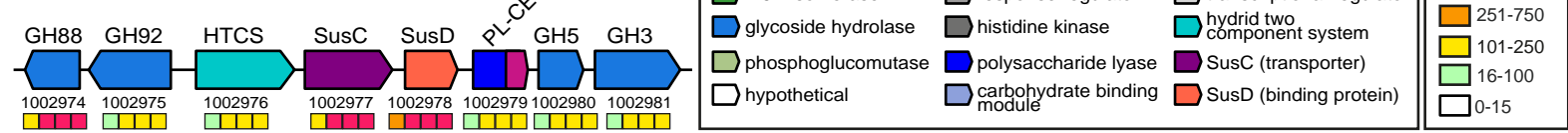


Figure 3

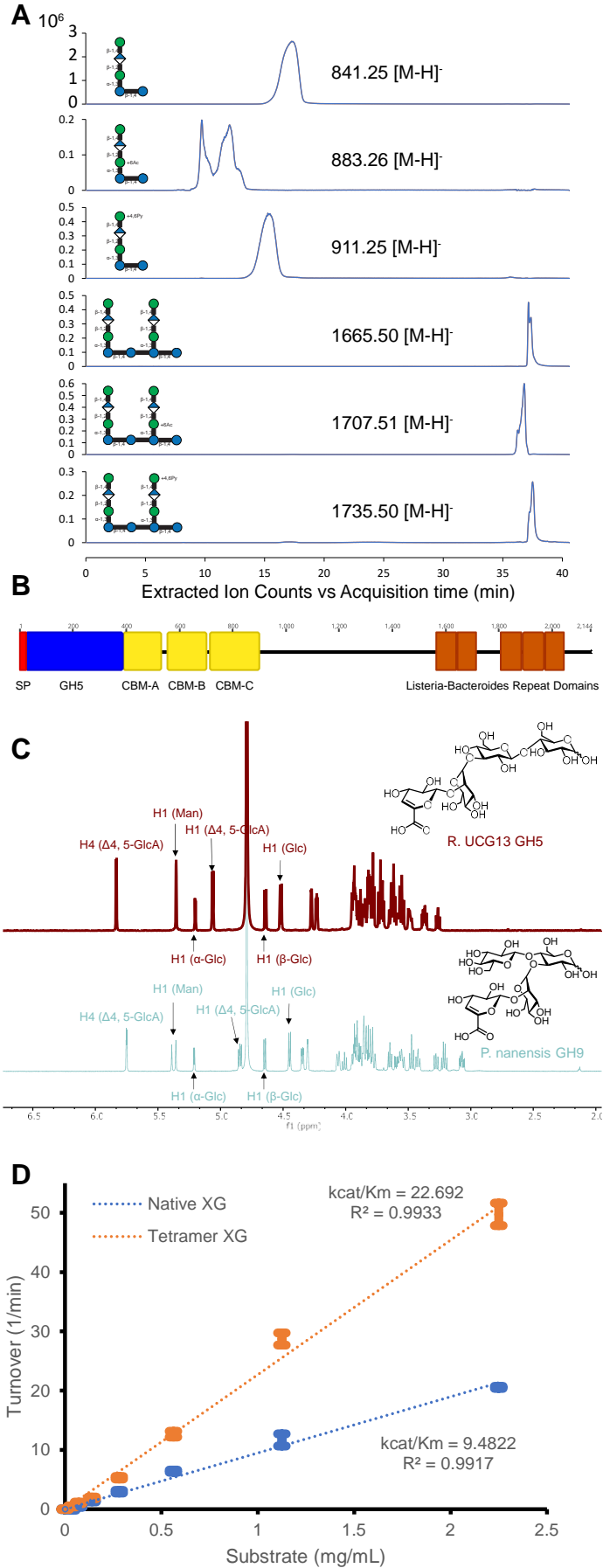
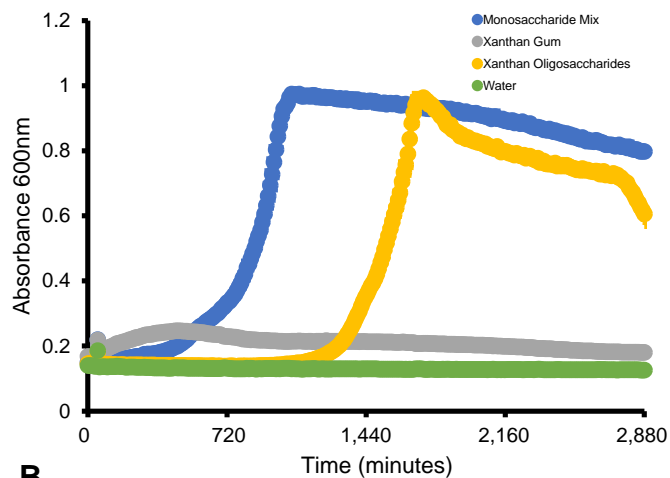


Figure 4

A



B

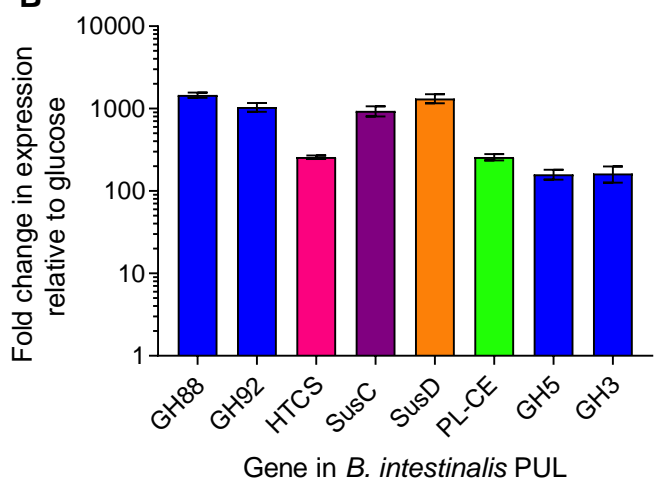
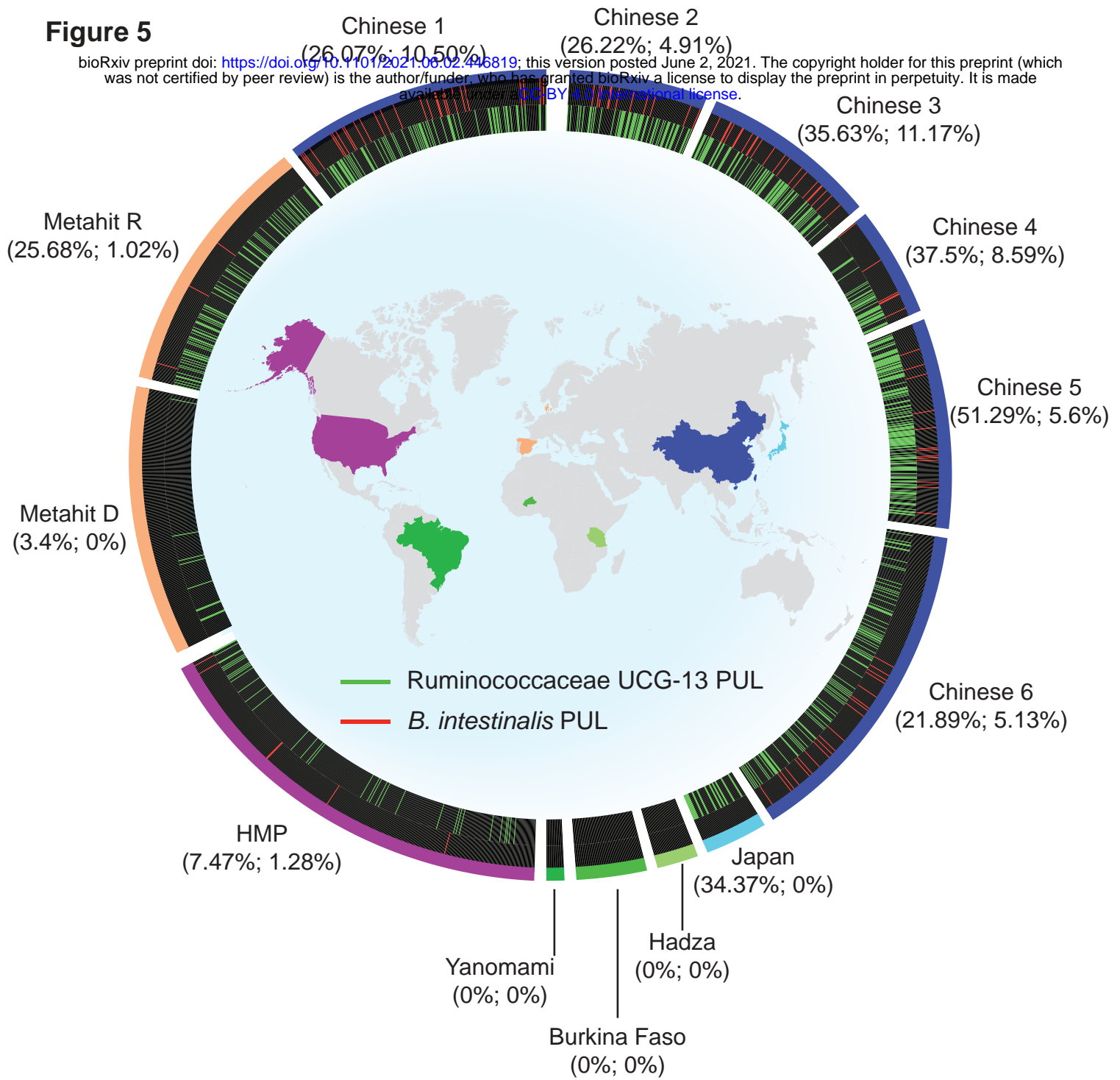
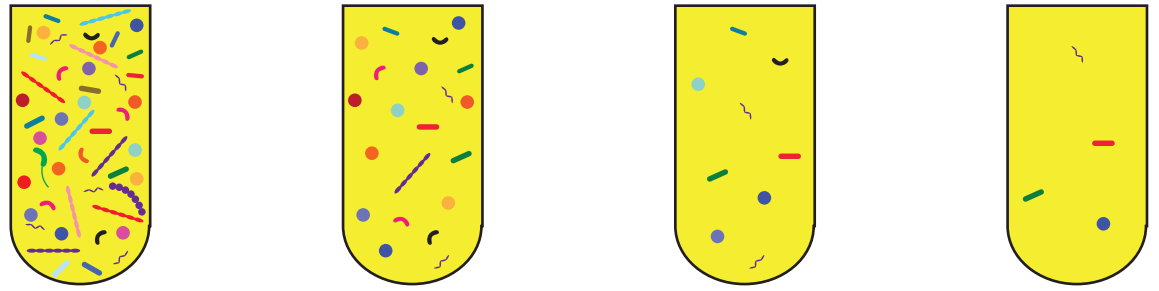


Figure 5

bioRxiv preprint doi: <https://doi.org/10.1101/2021.06.02.446819>; this version posted June 2, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Extended Data 2



Dilution Factor

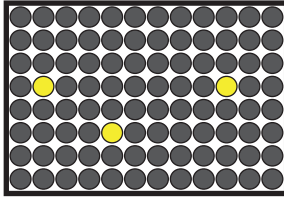
10^6

10^7

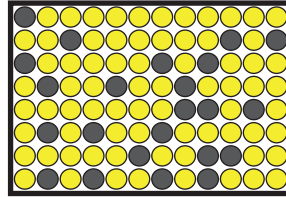
10^8

10^9

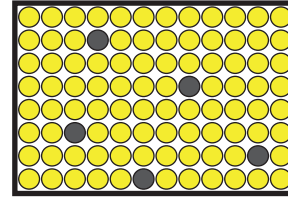
Xanthan Gum



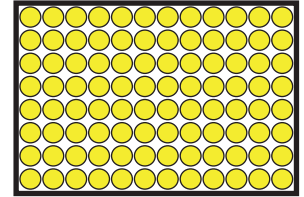
93/96



23/96

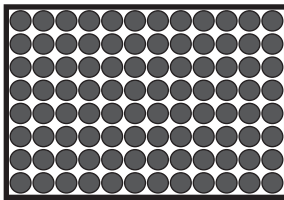


5/96

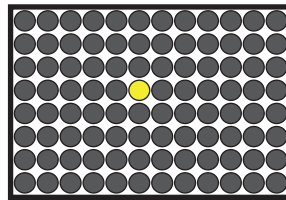


0/96

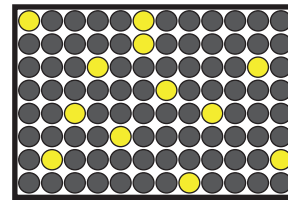
Monosaccharides



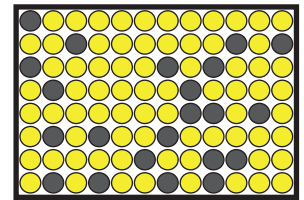
96/96



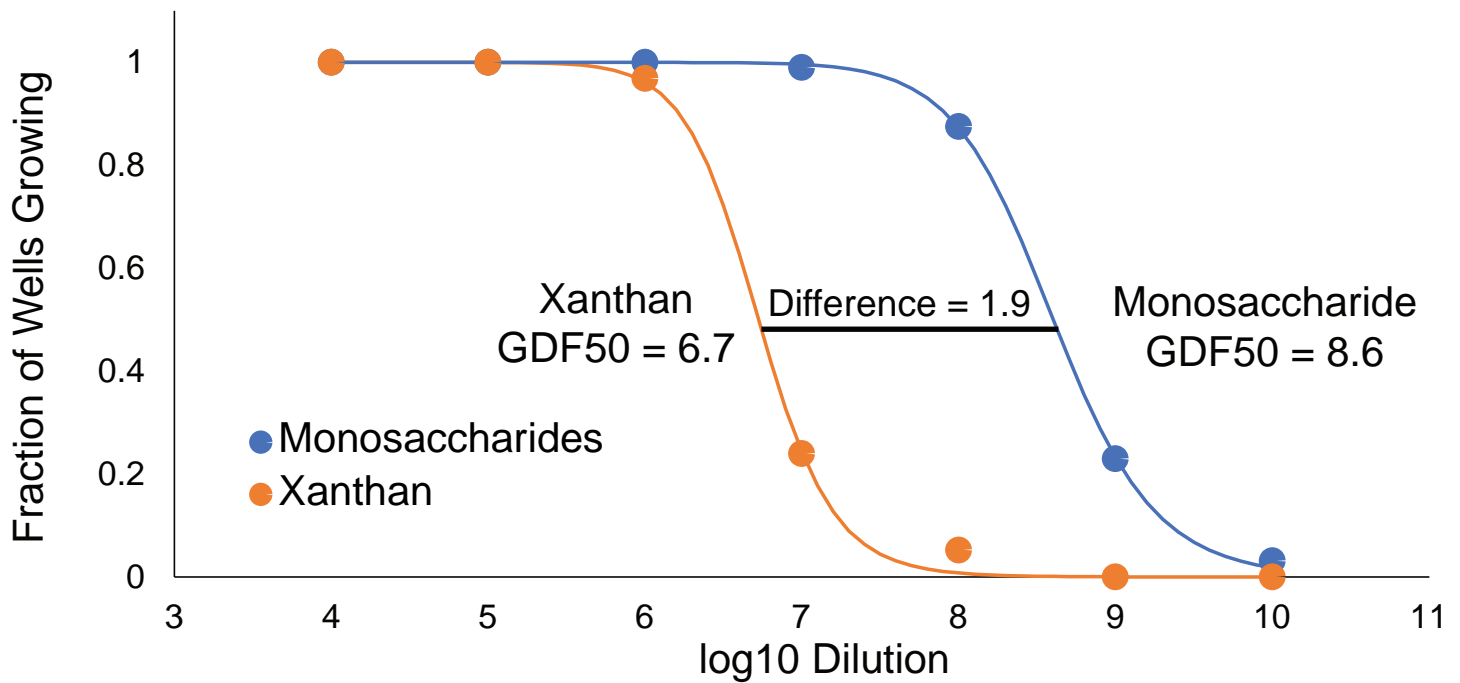
95/96



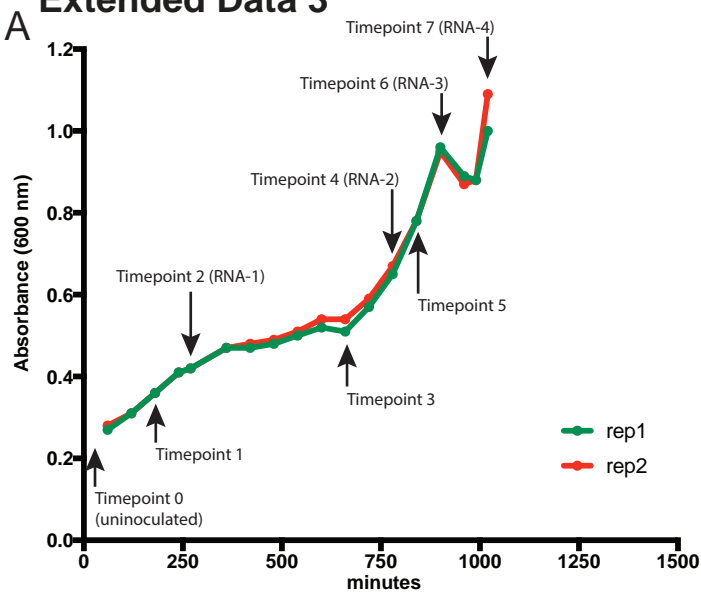
84/96



22/96



Extended Data 3



B

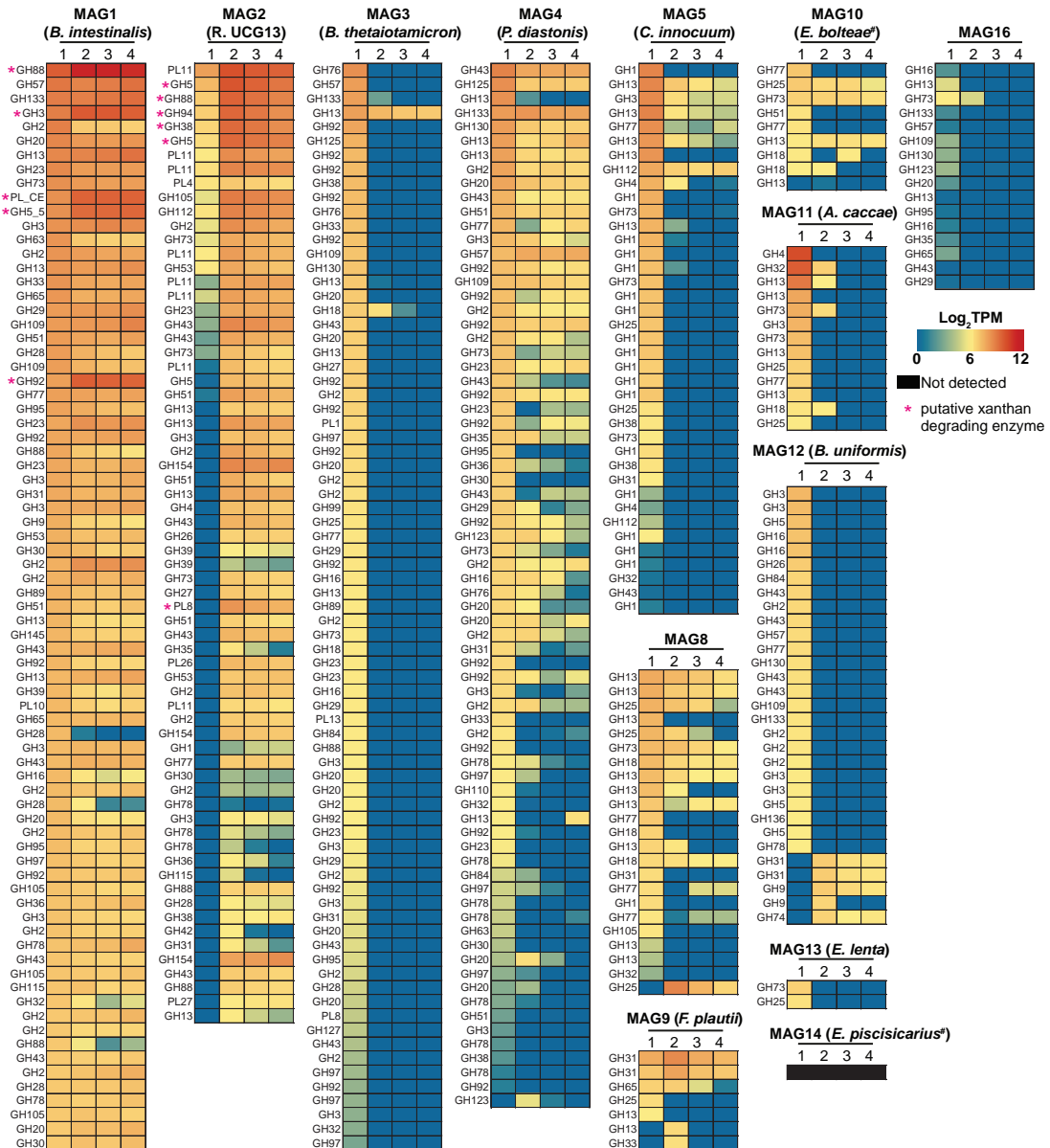
Replicate 1 composition

Samples	Glucose (%)	Mannose (%)
Timepoint 0	54.82 ± 0.12	45.18 ± 0.49
Timepoint 1	54.37 ± 1.10	45.63 ± 1.05
Timepoint 2	55.52 ± 0.78	44.48 ± 0.77
Timepoint 3	54.97 ± 0.57	45.03 ± 0.57
Timepoint 4	56.73 ± 0.42	43.27 ± 0.29
Timepoint 5	55.84 ± 0.14	44.26 ± 0.11

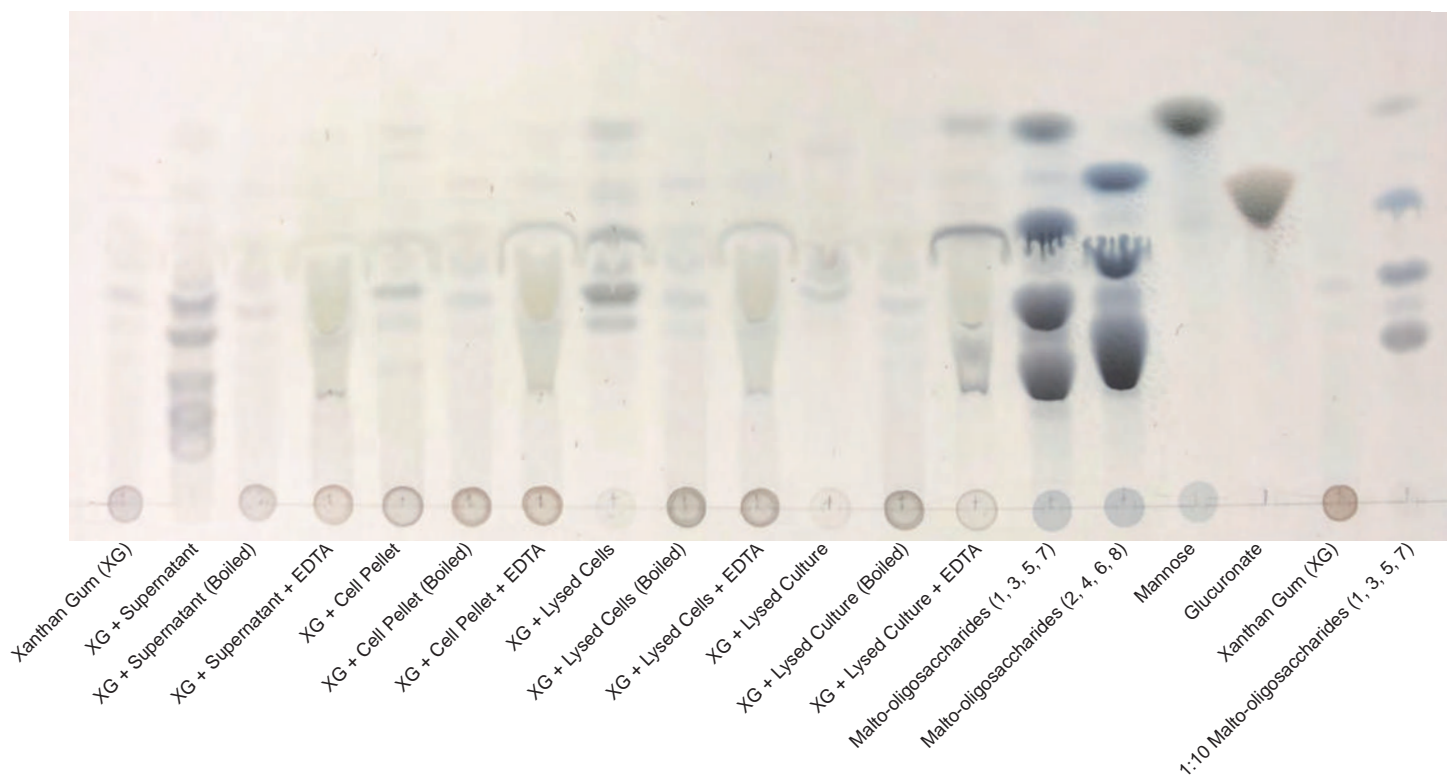
Replicate 2 composition

Samples	Glucose (%)	Mannose (%)
Timepoint 0	54.82 ± 0.12	45.18 ± 0.49
Timepoint 1	54.99 ± 0.65	45.01 ± 1.03
Timepoint 2	55.04 ± 0.55	44.96 ± 0.84
Timepoint 3	55.24 ± 0.32	44.76 ± 0.54
Timepoint 4	55.99 ± 0.17	44.01 ± 1.32
Timepoint 5	55.28 ± 0.88	44.72 ± 1.09

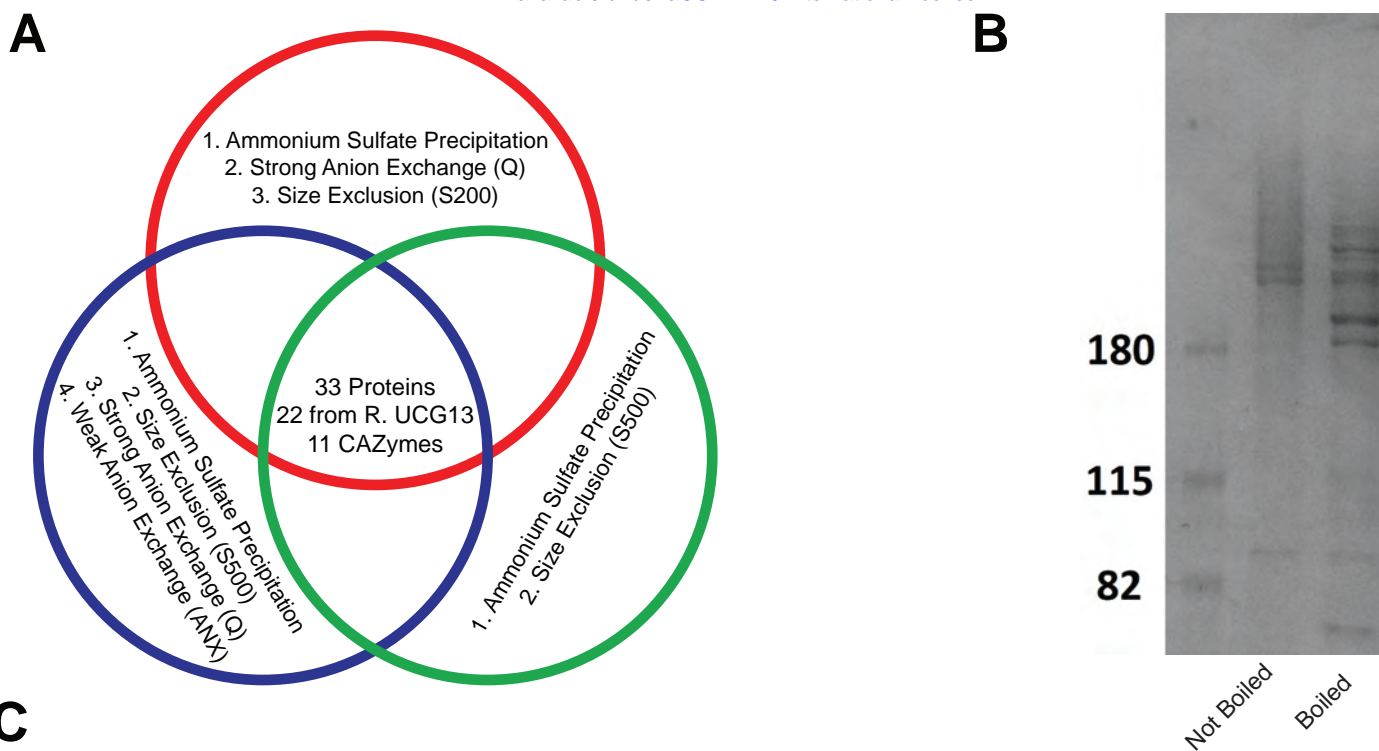
C



Extended Data 4



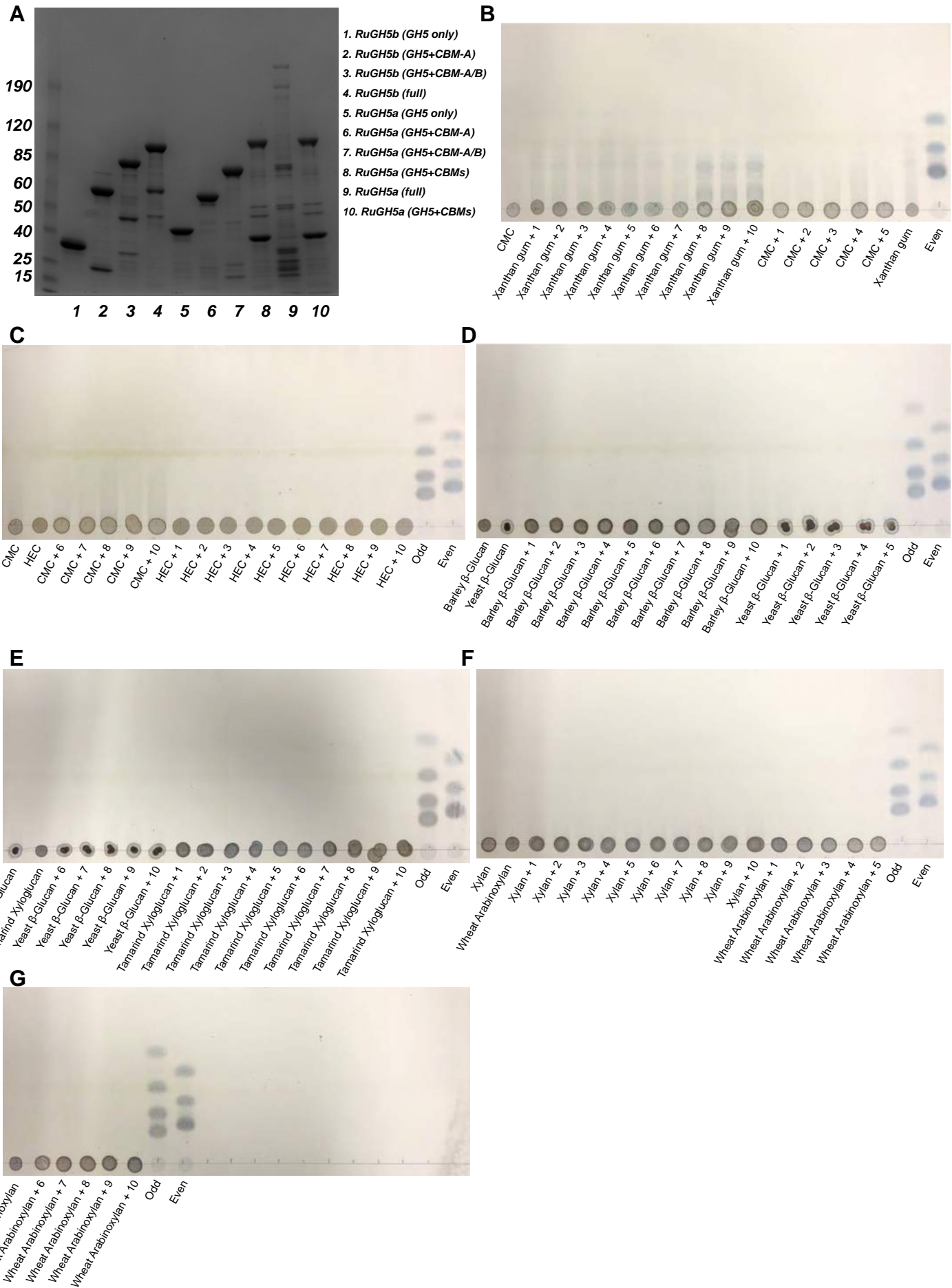
Extended Data Fig. 6



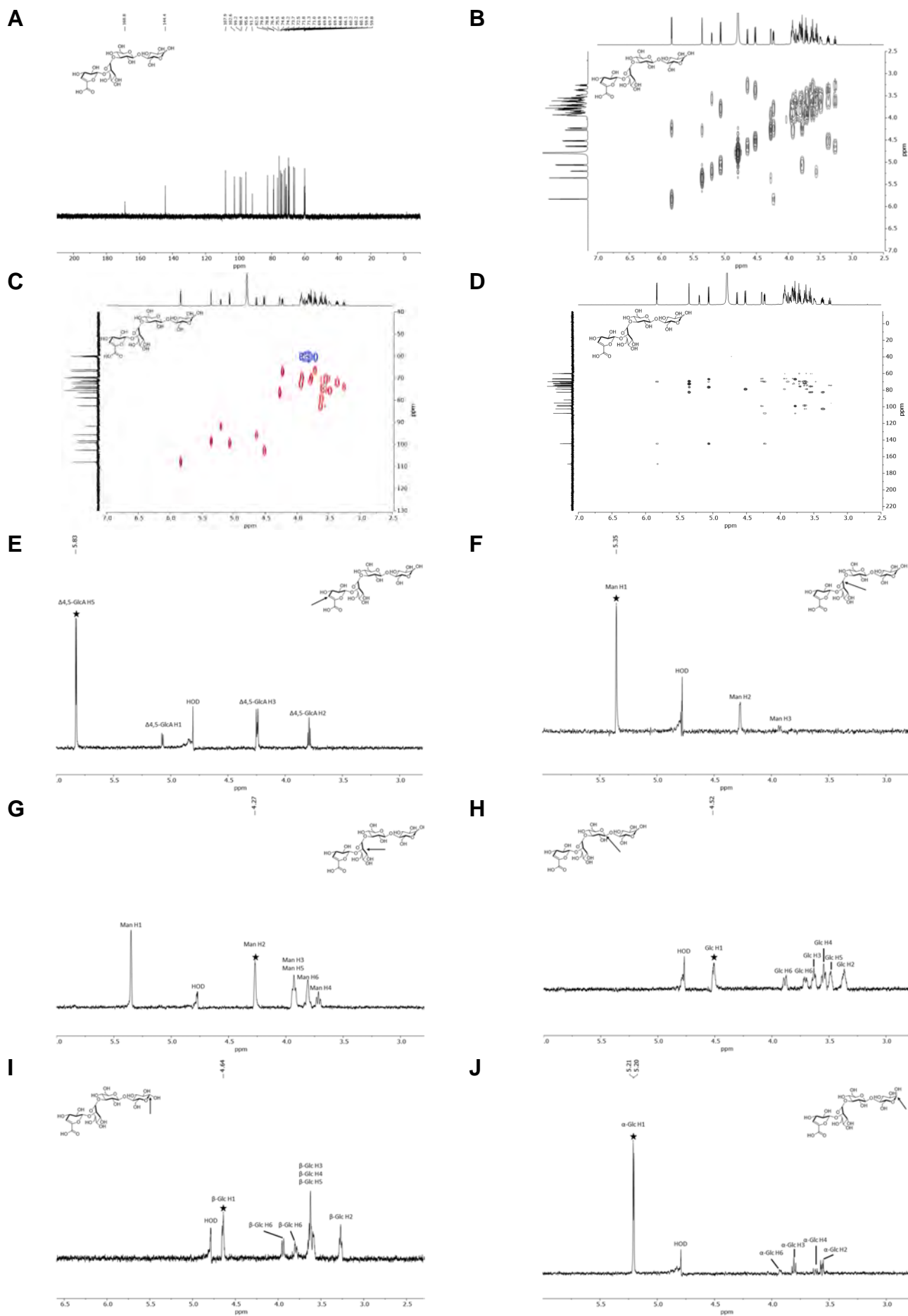
C

Protein ID	Protein Annotation	Species Prediction	CAZyme Annotation	#Peptides (E95)	#Peptides (E97)	#Peptides (E131)
Ga0308426_1011562	beta-galactosidase/beta-glucuronidase	Ruminococcaceae UCG-13	GH2	102	110	114
Ga0308426_10042164	alpha-tubulin suppressor-like RCC1 family protein	Ruminococcaceae UCG-13	#N/A	106	196	8
Ga0308426_102646	cellulase (glycosyl hydrolase family 5)/List-Bact-rpt repeat protein	Ruminococcaceae UCG-13	GH5	78	94	91
Ga0308426_100982	uncharacterized protein RhaS with RHS repeats	Ruminococcaceae UCG-13	#N/A	120	94	13
Ga0308426_1061813	uncharacterized protein YjdB	Ruminococcaceae UCG-13	PL11	71	64	61
Ga0308426_1039440	beta-xylosidase	Ruminococcaceae UCG-13	GH43_4-GH43_29	75	61	49
Ga0308426_110622	N-acetylneuraminic acid mutarotase	Clostridium formicaceticum A1, DSM92	#N/A	66	65	49
Ga0308426_1046623	O-glycosyl hydrolase	Ruminococcaceae UCG-13	GH95-GH43_22-GH30_5	70	53	12
Ga0308426_1044429	S-layer family protein	Ruminococcaceae UCG-13	#N/A	55	26	44
Ga0308426_1010896	SdrD B-like protein	Anaerostipes hadrus	#N/A	75	22	26
Ga0308426_115143	List-Bact-rpt repeat protein	Lachnospiraceae bacterium sp. 7_1_58FAA	#N/A	34	33	31
Ga0308426_110623	S-layer family protein	Ruminococcaceae UCG-13	#N/A	46	44	5
Ga0308426_108116	S-layer family protein	Ruminococcaceae UCG-13	#N/A	55	33	6
Ga0308426_1009815	formate C-acetyltransferase	Ruminococcaceae UCG-13	#N/A	48	16	16
Ga0308426_10026226	S-layer family protein	Ruminococcaceae UCG-13	#N/A	30	10	17
Ga0308426_1046628	alpha-L-arabinofuranosidase	Ruminococcaceae UCG-13	GH51-GH43_34	24	14	19
Ga0308426_1026413	cellulase (glycosyl hydrolase family 5)	Ruminococcaceae UCG-13	GH5	27	18	9
Ga0308426_109238	hypothetical protein	Bacteroides intestinalis	#N/A	26	10	18
Ga0308426_1011531	Ig-like protein group 3	Ruminococcaceae UCG-13	PL11	27	9	15
Ga0308426_1015472	chaperonin GroEL	Clostridium saccharolyticum	#N/A	26	12	7
Ga0308426_10042105	hypothetical protein	Ruminococcaceae UCG-13	CBM66	25	14	5
Ga0308426_1038838	surface antigen	Ruminococcaceae UCG-13	COH1-DOC-COH1	24	12	7
Ga0308426_1000896	putative aldouronate transport system substrate-binding protein	Ruminococcaceae UCG-13	#N/A	28	5	3
Ga0308426_10026230	carboxyl-terminal processing protease	Ruminococcaceae UCG-13	#N/A	20	1	14
Ga0308426_10042104	copper amine oxidase-like protein/concanavalin A-like lectin/glucanase superfamily protein	Ruminococcaceae UCG-13	#N/A	19	1	15
Ga0308426_109371	hypothetical protein	Clostridium sp. ATCC BAA-442	#N/A	22	7	5
Ga0308426_10026260	hypothetical protein	Ruminococcaceae UCG-13	#N/A	22	7	2
Ga0308426_1000893	rhamnogalacturonyl hydrolase YesR	Ruminococcaceae UCG-13	GH105	16	8	1
Ga0308426_112844	hypothetical protein	Bacteroides pectinophilus ATCC 43243	#N/A	18	1	6
Ga0308426_108113	copper amine oxidase-like protein/leucine rich repeat (LRR) protein	Ruminococcaceae UCG-13	#N/A	15	2	3
Ga0308426_1013474	hypothetical protein	Bacteroides intestinalis	#N/A	10	3	1
Ga0308426_10029134	hypothetical protein	Bacteroides intestinalis	#N/A	7	1	2
Ga0308426_115141	S-layer family protein	Flavonifractor plautii YL31	#N/A	6	1	2

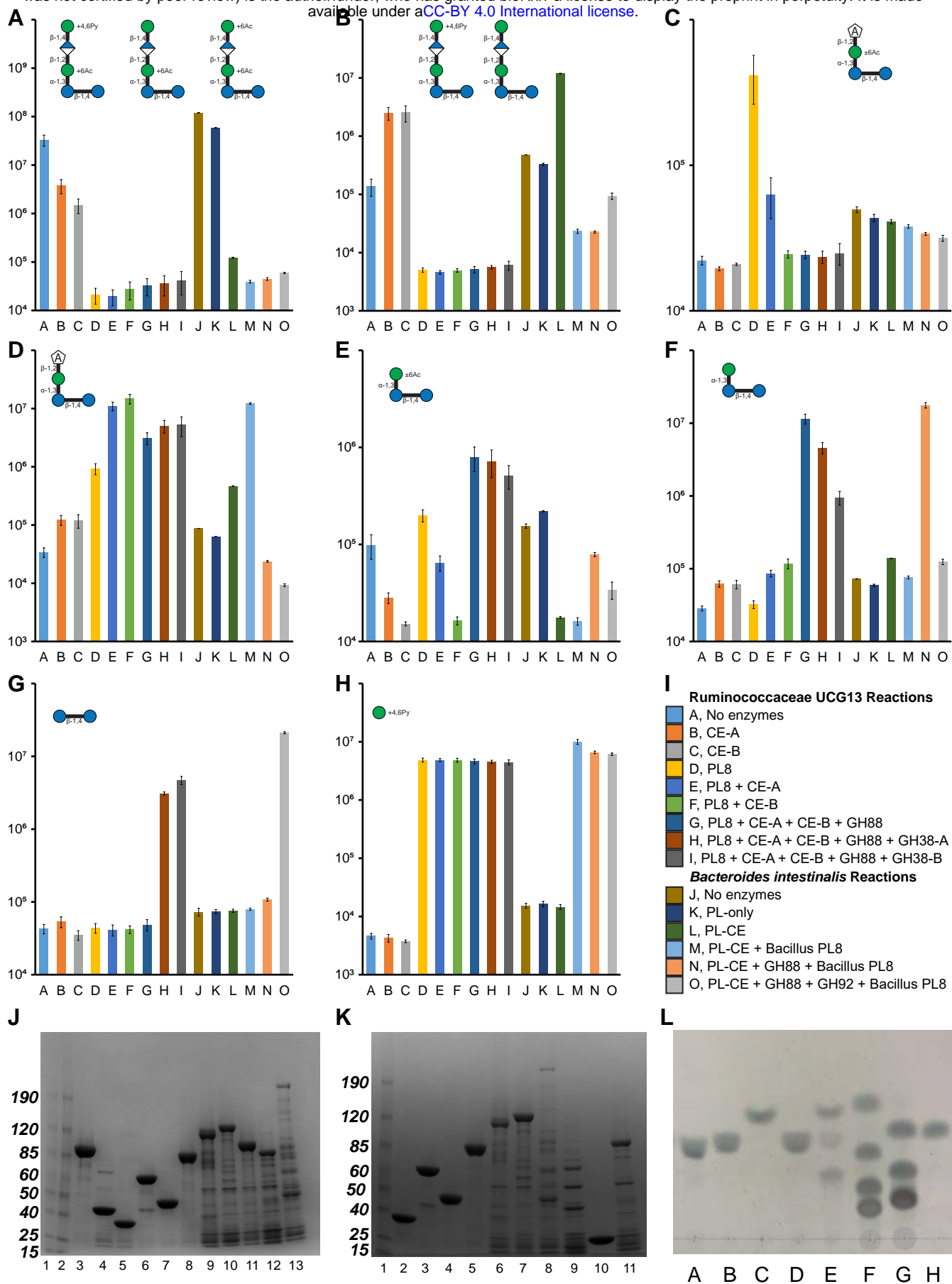
Extended Data 6



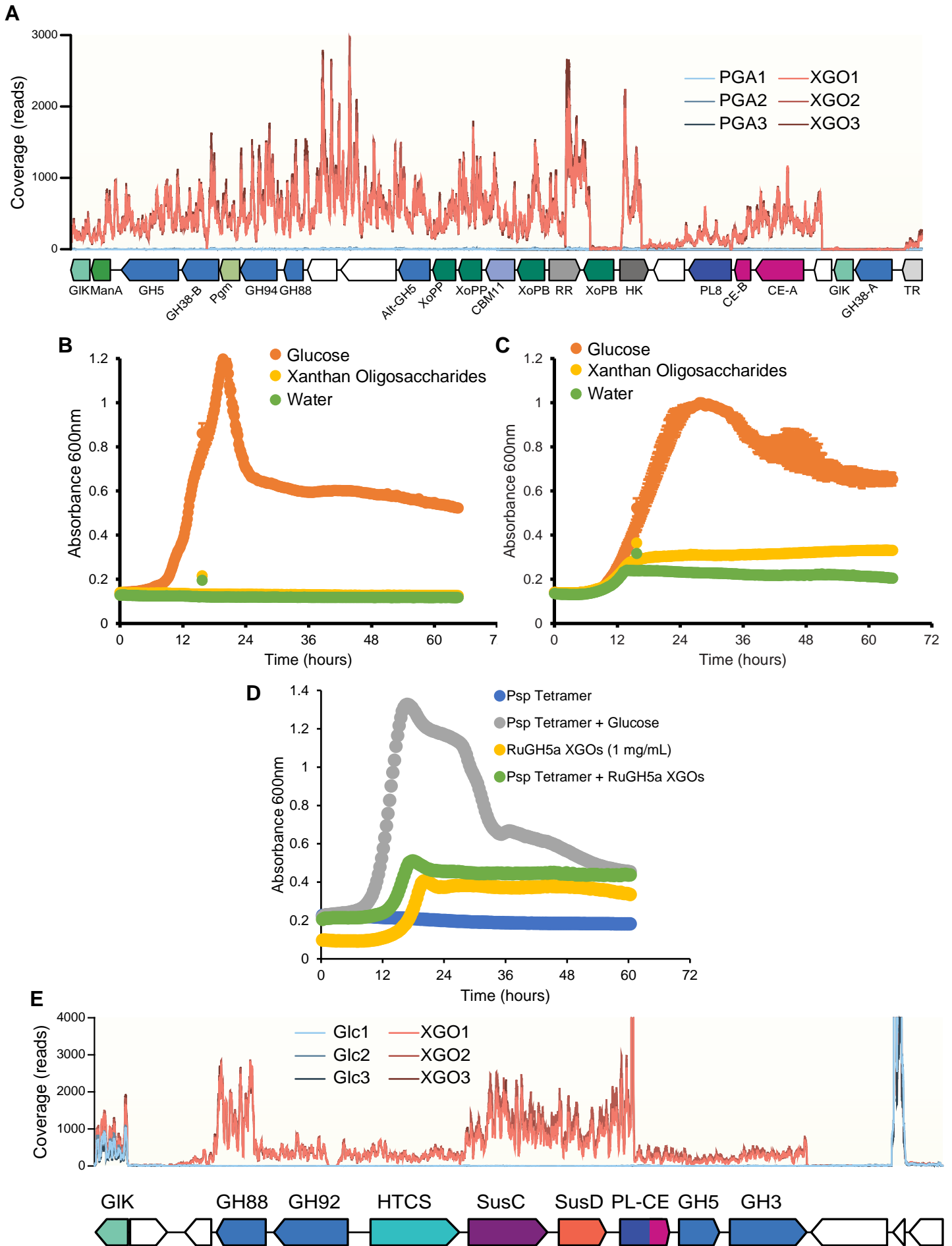
Extended Data 7



Extended Data 8

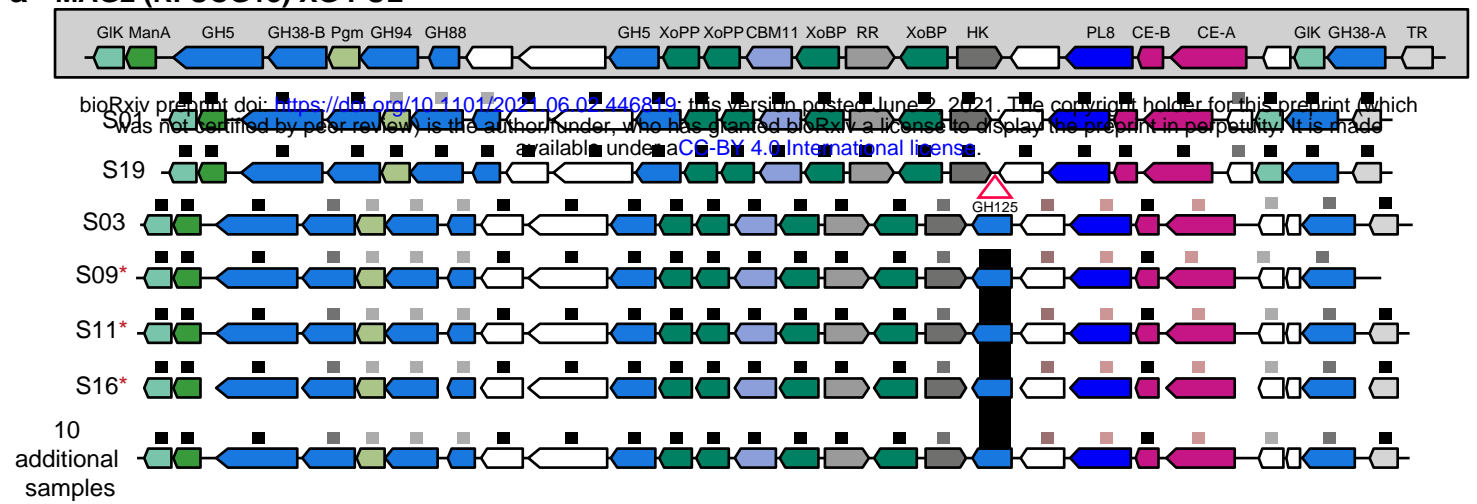


Extended Data 9

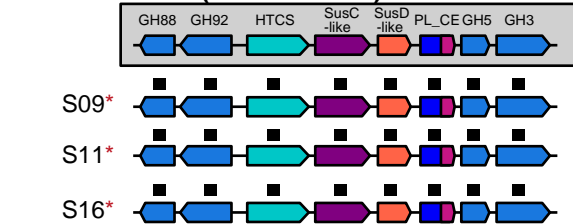


a **MAG2 (R. UCG13) XG PUL**

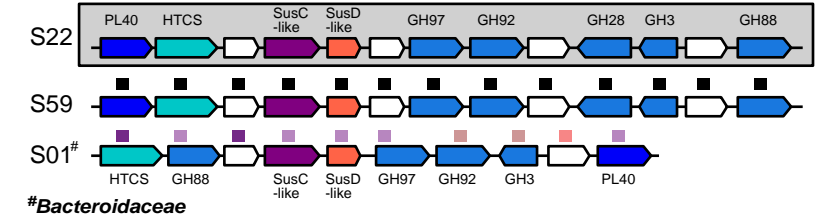
Extended Data 10



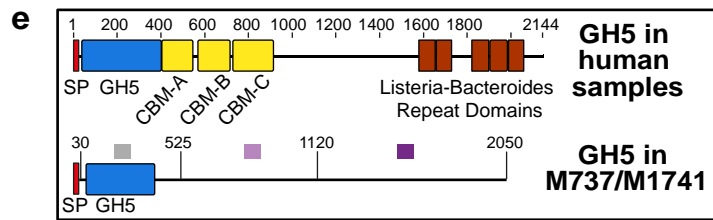
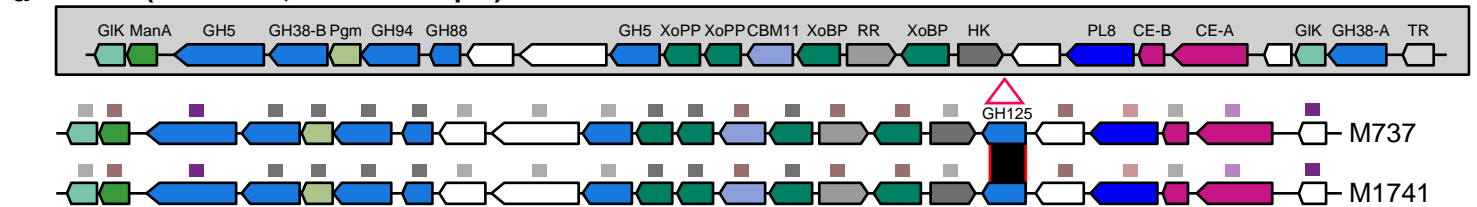
b **MAG1 (*B. intestinalis*) XG PUL**



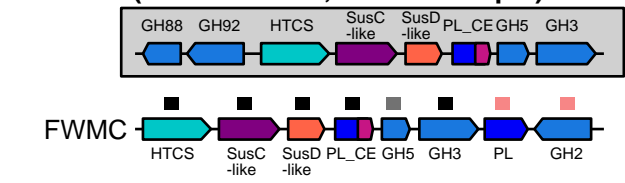
c ***Bacteroides uniformis***



d **MAG2 (R. UCG13, human sample) XG PUL**

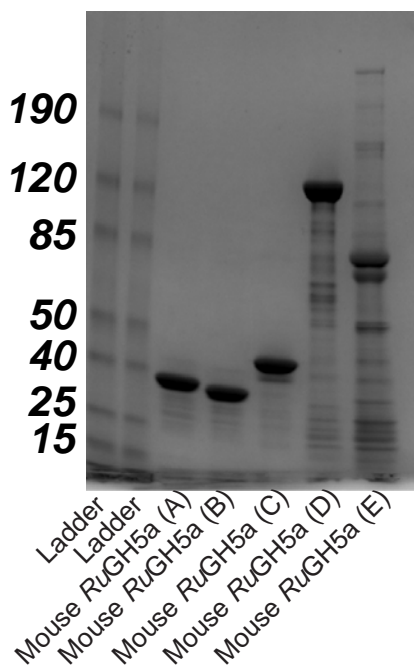


f **MAG1 (*B. intestinalis*, human sample) XG PUL**



Predicted gene functions						aa identity (%)	
glucokinase	phosphoglucomutase	response regulator	carbohydrate binding module	hydrid two component system	100-99	84-78	
M6P isomerase	hypothetical	histidine kinase	carbohydrate esterase	SusC-like	98-95	77-55	
glycoside hydrolase	ABC-transporter	polysaccharide lyase	transcriptional regulator	SusD-like	94-90	53-28	
					89-85	0	

g



h

