# A high-throughput microbial glycomics platform for prebiotic development

Jennifer L. Modesto[a,b,c], Seth G. Kabonick[a,b,c], Jennifer E. Lausch[a,b,c], Kamalesh Verma[a,b,c], Kailyn M. Winokur[a,b,c], Jessica E. Gaydos[a,b,c], Asia Poudel[a,b,c], Gregory Yochum[a,d], and Guy E. Townsend II[a,b,c]

[a]*Department of Molecular and Precision Medicine, Penn State College of Medicine, Hershey, PA, USA*

[b]*Penn State One Health Microbiome Center, Pennsylvania State University, State College, PA, USA*

[c]*Center for Molecular Carcinogenesis and Toxicology, Pennsylvania State University, State College, PA, USA*

[d]*Department of Surgery, Penn State College of Medicine, Hershey, PA, USA*

*Address correspondence to Guy E. Townsend II: gtownsend@pennstatehealth.psu.edu; Department of Molecular and Precision Medicine, Penn State College of Medicine, 700 HMC Crescent Road, Hershey, Pennsylvania, USA

**Abstract**

The mammalian intestine contains diverse carbohydrate pools that govern the gut microbiome composition. Structurally distinct polysaccharides, also called glycans, are differentially consumed by gut microbial subsets and direct their abundance by controlling gene expression and metabolite production. Therefore, identifying gut microbial accessible carbohydrates (MACs) is necessary to develop new prebiotics that beneficially manipulate the gut microbiome. However, no methods exist to efficiently examine MACs in biologically-derived mixtures. Here, we present a high-throughput platform to detect MACs from various plant, animal, and microbial sources using a genome-wide library of engineered *Bacteroides thetaiotaomicron* (*Bt*) strains that harness their endogenous glycan detection machinery. We demonstrate that this platform exhibits specific and sensitive responses to glycan mixtures and use bacterially-encoded proteins to characterize a previously unknown MAC from yeast. Expanding this technology across gut *Bacteroides* species will generate a broadly applicable approach to characterize heterogeneous glycan mixtures and identify prebiotic substrates.

38 **Main**

39 Commensal microbes harvest carbon from structurally diverse MACs, to achieve
40 extraordinary cell densities in the mammalian intestine[1-5]. The glycosidic linkages that tether
41 constituent monosaccharides into each MAC are frequently resistant to degradation in the
42 gastrointestinal tract and therefore require substrate-specific enzymes to degrade[6-11]. Gut
43 microbes can utilize available MACs by expressing carbohydrate-active enzymes that target
44 specific glycosidic linkages[1, 8, 11, 12]. Characterizing MACs present in biologically-derived mixtures
45 informs how and why the gut microbiota changes due to host dietary choices, genetics, or disease
46 because chemically-distinct glycans support the expansion of defined microbial subsets.
47 Additionally, identifying MACs that can influence microbial abundance and metabolism can
48 promote engineered strain engraftment and gene expression to leverage the gut microbiome as
49 a therapeutic target[13]. Glycan mixtures have traditionally been examined using tandem
50 chromatography and mass spectrometry, arrayed glycan binding aptamers, or lectins that probe
51 for a range of known carbohydrate structures[14-16]. Although these approaches can characterize
52 glycan content in various sample types, they are unable to directly detect MACs.

53 Microbes access unique MACs by producing proteins that sequester, transport, degrade,
54 and detect target molecules in their environment. For example, members of the dominant human
55 intestinal phylum, *Bacteroidetes*, possess glycan sensors encoded within polysaccharide
56 utilization genetic loci (PULs) that direct rapid, dramatic, and predictable increases to
57 corresponding PUL gene transcripts following recognition of their target glycan structures[4, 6, 7, 17].
58 Ultimately, these sensors direct increased expression of surface glycan binding proteins (SGBPs),
59 outer membrane transporters (SusCD), and glycolytic enzymes for consumption of available
60 MACs. Different *Bacteroides* species encode vast and partially unique PUL repertoires, which
61 endow each with the ability to access various target glycans that establish distinct intra-intestinal
62 niches. We previously demonstrated that introduction of PUL promoters into a plasmid-borne
63 *Bacteroides*-optimized *lux* cassette, p*Bolux,* converts these bacteria into biosensors of known and

64 unknown target glycans present in biologically-derived mixtures[18]. We reasoned that PUL
65 reporters could reflect the exquisite sensitivity and specificity of their cognate sensor proteins to
66 detect MACs in glycan mixtures and be implemented to isolate and characterize those molecules
67 in combination with corresponding PUL SGBPs and glycolytic enzymes.

68 Here, we generated a genome-wide *Bt* reporter library and arrayed these strains into a
69 scalable and distributable high-throughput platform that sensitively indicates target glycans from
70 various host-, microbial-, and plant-derived mixtures. We demonstrate that this approach
71 distinguishes structurally distinct MACs according to monosaccharide composition, glycosidic
72 linkages, and polysaccharide degree of polymerization. We use this technology to detect
73 previously unknown PUL target glycans and employ a recombinant SGBP to isolate and
74 characterize an unknown MAC in Baker's yeast (*Saccharomyces cerevisiae).* Collectively, these
75 results establish *Bacteroides* species as specific and sensitive toolkits to directly detect, isolate,
76 and characterize unknown PUL target glycans that could be developed into prebiotics. Expanding
77 these tools into additional *Bacteroides* species will support broad, untargeted glycomics
78 applications to examine heterogenous glycan mixtures and overcome current detection
79 limitations. Ultimately, our platform can be used to develop new prebiotics that manipulate gut
80 microbial abundance and metabolism to benefit human health.

81

82 **Results**

83 **Construction of an arrayed *Bt* PUL reporter library.**

84 We established that individual *Bt* PUL sensors direct predicted PUL transcript increases[19, 20]
85 because engineered strains expressing constitutively active sensor proteins from PULs 16, 19,
86 49, 53, and 57, exhibited dramatic increases in corresponding transcript abundances (Fig. 1a,
87 Supplementary Table 1)[21, 22]. To examine whether sensor-directed PUL transcription can be
88 faithfully reported by corresponding transcriptional reporters on a genome-wide scale, we
89 generated a complete library of *Bt* PUL reporter plasmids by introducing putative promoters

90 identified from RNAseq datasets[23] and mapped transcription start sites[24, 25] from annotated *Bt*
91 PULs[26, 27] into p*Bolux*[18] (Supplementary Table 2). Although *Bt* possesses 88 PULs, we generated
92 91 total reporter strains because PULs 14, 52, 74, and 75 possess two divergently transcribed
93 *susCD* pairs that were assigned unique reporters (denoted a and b) and PUL63 was omitted due
94 to the absence of a *susD*-like gene (Fig. 1a, Supplementary Table 1, 2). Following
95 cryopreservation of batch-produced 96-well plates, individual replicates were revived by the
96 addition of rich media and incubated overnight before transfer to a 384-well plate and supplied 4
97 distinct carbohydrate preparations, one of which was always galactose as a negative control (Fig.
98 1b). Fold changes in growth-adjusted bioluminescence generated over 18 hours (Fig. 1c) were
99 calculated from the area under each curve from strains supplied unique glycan preparations and
100 normalized to a strain harboring the promoter-less p*Bolux* plasmid and cultures supplied
101 galactose alone (Fig. 1d).
102

103 ***Bt* PUL reporters distinguish between glycosidic linkages and monosaccharide**
104 **composition.**

105 To examine the *Bt* PUL reporter array functionality, we examined changes in bioluminescence
106 from all strains following the introduction of various purified mono- and poly-saccharide
107 preparations. As expected, the monosaccharides, fructose and ribose, increased
108 bioluminescence from their respective PUL reporters, P-*PUL22* and P-*PUL37*, which agree with
109 previous findings (Extended Data Fig. 1a)[18]. However, ribose also increased *PUL22* transcripts[28]
110 and corresponding bioluminescence from P-*PUL22* (Extended Data Fig. 1a, b). Ribose is likely
111 contaminated with fructose because a *Bt* strain lacking *sensor*$^{PUL22}$ (*BT1754*) harboring P-*PUL22*
112 exhibited no bioluminescence increases when supplied ribose, whereas this strain harboring P-
113 *PUL37* exhibited similar responses to *wild-type Bt* (Extended Data Fig. 1c). In contrast, a strain
114 lacking *sensor*$^{PUL37}$ (*BT2802*), which is unable to increase *PUL37* transcripts or utilize ribose as a
115 sole carbon source[28], did not display increased activity when harboring P-*PUL37* (Extended Data

116 Fig. 1c). Similarly, P-*PUL22* also exhibited increased bioluminescence when supplied mannose
117 (Extended Data Fig. 1a), that required *sensor*$^{PUL22}$ (Extended Data Fig. 1b), suggesting that the
118 mannose preparation examined here was also contaminated with fructose. These results indicate
119 that the Sensor$^{PUL22}$, which binds fructose with high affinity[7], is responsible for increasing PUL22
120 expression when supplied ribose or mannose. Ultimately, these data highlight the specificity of
121 PUL reporters for their target ligands and reflect their extraordinary sensitivity in mixtures.

122 We established that *Bt* PUL reporters can distinguish between MACs based on glycosidic
123 linkages, monosaccharide composition, and degree of polymerization (d.o.p.). Distinct strains
124 exhibited increased bioluminescence according to the administration of various purified plant,
125 animal, and microbially-derived MACs consumed via known PULs[4, 6, 7, 11, 28, 29]. Strikingly, P-
126 *PUL66*, P-*PUL48*, and P-*PUL56* were specifically activated by amylose, dextran, and pustulan,
127 respectively, which are distinct glucose homopolymers differing by their α1,4-, α1,6-, and β1,6-
128 glycosidic linkages (Fig. 1e). Each PUL is regulated by their respective SusR-like sensor proteins,
129 encoded by BT3703 (Sensor$^{PUL66}$)[30], BT3091 (Sensor$^{PUL48}$)[31], and BT3309 (Sensor$^{PUL56}$)[32], to
130 increase target PUL transcription and utilize these MACs as growth substrates. Sensor$^{PUL66}$ and
131 Sensor$^{PUL56}$ differentiate between glycosidic linkage because their corresponding PUL reporters
132 only exhibit increased bioluminescence when supplied α1,4- or β1,6- linked glucose
133 disaccharides (Fig. 2a). However, P-*PUL48* can also distinguish between d.o.p. because
134 bioluminescence from this strain only increased when supplied isomaltooligosaccharides greater
135 than 3 α1,6-linked glucose units and requires at least 6 residues to reach activity similar to dextran
136 (Fig. 2b). These data demonstrate that SusR-like PUL sensors detect distinct glycosidic linkages
137 and d.o.p..

138 Further examination of purified oligosaccharides yielded several unexpected outcomes. First,
139 monomeric glucose and various glucose-disaccharides, including maltose, isomaltose, trehalose,
140 and β-gentibiose increased bioluminescence from P-*PUL58* (Fig. 2a). Similarly, supplying
141 glucose-fructose heterodimers, turanose and palatinose, stimulated greater than 10-fold

142 increases from this strain, whereas sucrose did not, suggesting that this PUL reporter responds
143 to glucose- and fructose- containing disaccharides subsets. We hypothesized that P-*PUL58* could
144 be governed by the SusR-like sensor, encoded by *BT2160*, that is an important *Bt* fitness
145 determinant during growth in trehalose and palatinose[33]. Accordingly, *Δsensor*$^{PUL58}$ harboring P-
146 *PUL58* did not exhibit increased bioluminescence in any condition, indicating that this sensor
147 facilitates PUL58 transcription in response to various disaccharides (Fig. 2c). Curiously, this
148 sensor also responds to glucosinolates derived from cruciferous vegetables and is required for
149 their conversion into isothiocyanates by increasing transcription of the linked genes *BT2159-56*[34].
150 A deeper examination of how and why this distally encoded sensor protein responds to these
151 signals is necessary to understand this regulatory relationship. Unexpectedly, P-*PUL46* detected
152 a *Bt*-inaccessible carbohydrate[35] because introduction of the β1,4-linked glucose disaccharide,
153 cellobiose, increased bioluminescence from this strain (Fig. 2a, d). This response was d.o.p.-
154 specific because there was no change when supplied cellulose-derived tetra- or hexa-saccharides
155 and this response did not require *sensor*$^{PUL58}$ (Fig. 2d). Because *Bt* cannot utilize cellulose or its
156 polysaccharide breakdown products including cellobiose[35], these results suggest that cellobiose
157 is a signal in the intestine.
158

159 ***Bt* PUL reporters detect discreet carbohydrate structures**

160 The *Bt* PUL array can also distinguish between various microbial-, animal-, and plant glycans
161 covalently tethered into polymeric complexes. For example, *Bt* utilization of yeast cell wall α-
162 mannans (Extended Data Fig. 2a) requires three PULs: 36, 68, and 69[36, 37], which increase activity
163 from the reporter strains, P-*PUL36*, P-*PUL68,* and P-*PUL69* by the addition of purified α-mannan
164 (Fig. 1e). The corresponding sensor proteins, BT2628 (Sensor$^{PUL36}$) and BT3786 (Sensor$^{PUL68}$)
165 facilitate mannan-responsive PUL36/68 transcription because bioluminescence from strains
166 harboring either P-*PUL36* or P-*PUL68* is not apparent in a strain lacking both sensors
167 (*Δsensors*$^{PUL36/68}$; Extended Data Fig. 2b), consistent with previous transcript measurements[36].

168 Alternatively, PUL69 transcription is facilitated by Sensor$^{PUL69}$, encoded by *BT3853*[36], which is
169 required for increased bioluminescence from P-*PUL69* when supplied α-mannan (Extended Data
170 Fig. 2b). Sensors$^{PUL36/68/69}$ are collectively required for mannan utilization as a sole carbon source
171 (Extended Data Fig. 2c) but detect discreet carbohydrate structures because the branched α1,3-
172 [α1,6] trisaccharide increases P-*PUL36/68* bioluminescence (Extended Data Fig. 2d, e).
173 Conversely, P-*PUL69* did not respond to any di- or tri-saccharide examined, indicating it responds
174 to a different oligo-mannan structure (Extended Data Fig. 2f). These data highlight that PUL
175 sensor proteins direct cognate PUL expression following recognition of distinct glycan moieties
176 assembled as larger polymers.

177 *Bt* possess 3 PULs that respond to glycosaminoglycans (GAGs), which are intestinally
178 abundant heteropolymers comprised of amine sugars and uronic acids, including previously
179 characterized PULs 57[6, 10] and 85[38], and the uncharacterized PUL20 (Fig. 3a). Strains harboring
180 P-*PUL57* exhibited increased bioluminescence when supplied chondroitin sulfate (CS) or
181 dermatan sulfate (DS; Fig. 1e), which are variably sulfated linear polymers of repeating,
182 disaccharide units comprised of glucuronic acid (GluA) and β1,3-linked N-acetylgalactosamine
183 (GalNAc). P-*PUL57* also responds to hyaluronic acid[18], an unsulfated GAG whose disaccharide
184 unit is comprised of gluA and β1,3-linked N-acetylglucosamine (GlcNAc) because the Sensor$^{PUL57}$
185 protein, BT3334, cannot distinguish between either CS- or HA-derived disaccharide subunits (Fig.
186 3b)[6]. Introduction of CS, DS, and HA also increased bioluminescence from P-*PUL20* (Fig. 1e),
187 consistent with previous transcriptional measurements[3], suggesting that all GAGs activate an
188 additional PUL. However, Sensor$^{PUL57}$ is not responsible for increasing bioluminescence from P-
189 *PUL20* because: 1.) CS- and HA-derived disaccharides increased activity from P-*PUL57* but not
190 P-*PUL20* (Fig. 3b), 2.) heterologous expression of a constitutively active Sensor$^{PUL57}$ (BT3334*)
191 increased *PUL57* but not *PUL20* transcripts (Fig. 1a), and 3.) BT3334* increased
192 bioluminescence from strains harboring P-*PUL57* but not P-*PUL20* (Fig. 3c). Furthermore, a
193 distinct GAG, heparin, also increased bioluminescence from only P-*PUL20* and P-*PUL85* (Fig.

194 1e), whereas the composite disaccharide comprised of alternating N-acetylglucosamine (GlcNAc)
195 and uronic acid increased activity from only P-*PUL85* (Fig. 3b). We propose that PUL20 responds
196 to the common GAG tetrasaccharide linker, GlcA-[β1-3]-Gal-[β1-3]-Galβ-[1-4]-Xyl2P-β-Ser (Fig.
197 1a), because CS from shark cartilage stimulates increased bioluminescence from P-*PUL20* and
198 P-*PUL57* but no other contaminating glycans following co-incubation with SGBP$^{PUL57}$ (Fig. 3d)[18],
199 indicating that the PUL20-target glycan is covalently tethered to CS.

200 Plant cell wall glycans called pectins and hemicelluloses, are frequently assembled into
201 complex polymeric structures (Extended Data Fig. 3a) that are targeted by distinct *Bt* PULs[1, 4, 9,
202 11]. This is exemplified by P-*PUL7*, which exhibited increased bioluminescence when supplied
203 arabinan (Fig. 1e). However, arabinan also elicited increases from strains harboring P-*PUL5*, P-
204 *PUL13*, P-*PUL22*, P-*PUL47*, P-*PUL48*, P-*PUL65*, P-*PUL77*, and P-*PUL86*, indicating this
205 preparation also contains arabinogalactan (PUL5), rhamnogalacturonan II (PULs 13 & 21),
206 fructose (PUL22), dextran (PUL48), rhamnogalacturonan I (PUL77), and pectic galactan (PUL86)
207 (Extended Data Fig. 3b). Reporter activity faithfully reflects sensor-dependent activity because
208 the Sensor$^{PUL7}$ (BT0366)[39], which recognizes arabino-oligomers of at least 6 α1,5-linked
209 arabinose residues[4], was only required to stimulate P-*PUL7* when supplied arabinan (Extended
210 Data Fig. 3b). Consistent with this notion, purified oligo-galacturonate increased bioluminescence
211 from its corresponding reporter, P-*PUL75a&b* (Fig. 2a), whereas highly polymeric PGA also
212 increased activity from P-*PUL7*, P-*PUL22*, P-*PUL47*, P-*PUL48*, and P-*PUL66*, and P-*PUL86* (Fig.
213 1e), demonstrating that PUL reporters respond predictably to complex MACs comprised of
214 multiple ligands[1, 4, 7, 9, 11]. Finally, the ubiquitous food additives carrageenan, xantham gum, locust
215 bean gum, karaya gum, and gum arabic are complex heteropolymers of largely *Bt*-inaccessible
216 carbohydrates[9, 35] but contain small amounts of detectable structures because each elicits
217 increased bioluminescence from at least one PUL reporter (Fig. 1e). For example, P-*PUL5* is
218 dramatically activated by purified arabinogalactan and exhibits significant increases when
219 supplied karaya gum and gum arabic[9] (Fig. 1e). Collectively, these data demonstrate the exquisite

220 specificity of *Bt* PUL reporters for complex multi-valent MACs, facilitated by the cognate sensor
221 proteins.
222
223 **The *Bt* PUL reporter array can faithfully survey glycan mixtures.**
224 *Bt* PUL reporters can simultaneously examine MACs present in heterogeneous glycan
225 mixtures because P-*PUL5*, P-*PUL22*, P-*PUL57*, and P-*PUL48* exhibited increased
226 bioluminescence when supplied an equal mixture of purified arabinogalactan (AG, PUL5), levan
227 (PUL22), hyaluronan (HA, PUL57), and dextran (PUL48) (Fig. 1e). Additionally, this mixture
228 increased activity from P-*PUL20* and P-*PUL47*, which were both increased by HA (Fig. 1c-e),
229 displaying faithful reporting of lab-prepared mixtures. To characterize biologically derived
230 mixtures, we measured reporter responses to PMOGs, a mixture of porcine mucosal-derived O-
231 glycans that increases many *Bt* PUL transcripts[3, 18, 40]. When supplied PMOG, we detected
232 increased activity in GAG-, and O-glycan-specific PUL reporters, consistent with previous
233 transcript measurements[3, 18, 40] indicating that the reporter strains can faithfully examine
234 biologically-derived mixtures (Fig. 4a). To prepare other host-, diet-, and microbially-derived
235 mixtures, we adapted the oxidative-release of natural glycans (ORNG) method that can liberate
236 glycans from a wide variety of biological materials[41]. We determined that a preparation of porcine
237 mucosal glycans using ORNG (PMG), yielded similar reporter profiles as PMOG (Fig. 4a), but did
238 display a 2-fold decrease in bioluminescence from a strain harboring P-*PUL14b* and a 2-fold
239 increase in signal from P-*PUL48*, indicating glycan subsets are differentially extracted or detected
240 when this approach is applied (Extended Data Fig. 4a).
241 To determine how gut microbial responses differ between animal and human mucosal
242 material, we compared human-derived intestinal glycan mixtures collected from various patient
243 intestinal samples recovered following surgery (HIG) and porcine cecal mixtures from freshly
244 slaughtered boars (BCG). HIG elicited similar patterns of reporter activity as PMOG and fresh
245 porcine cecal glycans, suggesting that mammalian hosts largely supply gut microbes with similar

246 growth substrates (Fig. 4a). For example, HIG and BCG elicited similar bioluminescence from P-
247 *PUL14*b, P-*PUL20*, P-*PUL38*, P-*PUL66*, and P-*PUL72* (Extended Data Fig. 4b). However, the
248 following distinctions were readily apparent: on one hand, human material elicited reduced
249 bioluminescence from the CS/HA reporter P-*PUL57* and heparin reporter P-*PUL85*, but on the
250 other hand they elicited increased responses from P-*PUL76*, *P-PUL81, and P-PUL86* (Extended
251 Data Fig. 4b). Furthermore, human milk glycans (HMG) elicited similar responses to HIG,
252 however, signals were improved when HMG were prepared by defatting and proteinase K
253 treatment (Extended Data Fig. 4c), suggesting that ORNG can result in glycan peeling from some
254 source materials[42].

255 We used ORNG to prepare various animal, plant, and fungal material that are abundant
256 components of the human diet (Fig. 4a) and examined *Bt* reporter responses. As expected, plant-
257 , animal-, and microbially-derived material occupied distinct space following PCoA analysis,
258 indicating that MAC distribution reflects source phylogeny at the phylum level (Fig. 4b). For
259 example, dietary plant products emanating from both mono- and di-cot species distinctly
260 increased bioluminescence from strains harboring P-*PUL5* (arabinogalactan), P-*PUL7* (arabinan),
261 P-*PUL13* and P-*PUL21* (RGII), P-*PUL22* (fructan), P-*PUL66* (starch/glycogen), P-*PUL75* (PGA),
262 and P-*PUL86* (RGI), which is consistent with common plant fibers. Conversely, mammalian
263 derived mixtures stimulated P-*PUL14b* (N-glycans), P-*PUL20* (GAG associated), P-*PUL38*
264 (unknown mucin *O*-glycan), P-*PUL57* (CS/HA), P-*PUL66* (glycogen), and P-*PUL76* (unknown
265 mucin *O*-glycan). Finally, fungal products consistently increased activity from strains containing
266 P-*PUL29* (unknown), P-*PUL36/68* (α-mannan), P-*PUL56* (β-glucan), P-*PUL66* (starch/glycogen),
267 and P-*PUL69* (α-mannan). Accordingly, processed foods containing both plant and fungal
268 products, including breads and beer, contained reporter increases consistent with both grains and
269 yeast (Fig. 4a).

270 Unexpectedly, several PUL reporters did not exhibit bioluminescence increases when
271 supplied any purified carbohydrate or mixture examined in this study, suggesting either these

272 target glycans are not present in the mixtures prepared for this work or the corresponding reporter
273 plasmid may not contain the appropriate regulatory sequence to facilitate increased
274 bioluminescence. However, we determined that the latter is an unlikely scenario because qPCR
275 analysis for *susC* transcripts from PULs whose corresponding reporters failed to increase
276 bioluminescence were not increased when supplied various mixtures (Extended Data Fig. 5a).
277 Furthermore, expression of constitutively active sensor proteins from PULs 16, 19, 49, 53, and
278 57, which increased target PUL transcription (Fig. 1a)[21, 22], stimulated bioluminescence increases
279 when these strains contained the corresponding PUL reporter (Fig. 3c, Extended Data Fig. 5b).
280 A notable exception is PUL2, which targets an unknown glycan, exhibits 180-fold increased
281 transcripts when supplied yeast extract and red-wine derived glycan mixtures (Extended Data
282 Fig. 5a), but whose corresponding reporter did not exhibit increased bioluminescence despite
283 engineering an additional plasmid that included an internal transcription start site[25, 40] (Extended
284 Data Fig. 5c). Thus, the majority of PULs can be converted into glycan biosensors with a small
285 minority that may require additional optimization to use p*Bolux*. Finally, establishing the
286 functionality of various PUL reporters with unknown target glycans (Extended Data Fig. 5b) that
287 are not activated across the common foods examined in this study, indicate that future
288 investigations should focus on other glycan sources, including co-resident microbes and
289 fermented products.
290

291 **Unique PULs across p*Bolux*-compatible *Bacteroides* species**

292 This study demonstrates how *Bt* PUL reporters can be readily constructed and deployed
293 to detect MAC content in various glycan preparations. However, other *Bacteroides* species
294 possess unique PUL repertoires that putatively confer distinct glycan accessibility, indicating that
295 they recognize a partially unique subset of carbohydrate structures (Fig. 4c, Supplementary Table
296 3). For example, *B. fragilis* PULs facilitate its growth in host-derived glycan preparations including
297 those distinct from *Bt* target glycans[22], whereas *B. ovatus*[43] and *B. uniformis*[44] possess PUL

298 repertoires that target *Bt*-inaccessible hemicelluloses. We previously demonstrated that p*Bolux*
299 can report promoter-dependent transcription in *B. ovatus*[18] and *B. fragilis*[45], and this is consistent
300 in *B. uniformis* and *B. cellulosilyticus* because introduction of their corresponding *rpoD* promoters
301 in this plasmid increase bioluminescence from all five species (Extended Data Fig. 6). Therefore,
302 construction of future PUL reporter libraries in these species will expand the breadth of glycan
303 detection capabilities to comprehensively characterize MAC content in host-, microbial-, and
304 dietary- components.
305
306 **Isolation and characterization of a *Saccharomyces cerevisiae* glycan consumed by *Bt.***
307 To explore the utility of PULs as glycomics tools for prebiotic development, we
308 characterized an unknown glycan detected in a biologically derived mixture. We selected PUL71
309 (Fig. 5a) because yeast extract glycans (YEG) elicited 32-fold increased bioluminescence from
310 strains harboring P-*PUL71*, which was also stimulated by preparations from yeast containing
311 products like red wine and animal feeds, but not breads or beers (Fig. 4a). Furthermore, P-*PUL71*
312 activity was increased in yeast-derived mannan prebiotic mannooligosaccharides (Bio MOS) but
313 absent in host- and most diet-derived mixtures (Fig. 4a). Furthermore, this glycan is an excellent
314 prebiotic candidate because a strain lacking PUL71 (*ΔBT3957-BT3965*) exhibited a 5-fold fitness
315 defect relative to *wild-type Bt* in mice supplied a 0.5% YEG solution in place of water (Fig. 5b).
316 *PUL71* includes a unique glycan sensor protein class, encoded by *BT3957*, that possesses a
317 periplasmic glycan binding domain and a cytoplasmic DNA-binding domain similar to HTCSs
318 (Extended Data Fig. 7a). We established that the putative sensor, BT3957, controls PUL71
319 transcription because a *sensor*$^{PUL71}$-deficient strain (*Δsensor*$^{PUL71}$) harboring P-*PUL71* failed to
320 exhibit increased activity in response to YEG (Extended Data Fig. 7b). Conversely, *sensor*$^{PUL36/68}$-
321 or *sensor*$^{PUL69}$-deficient strains harboring P-*PUL71* exhibited similar bioluminescence as *wild-type*
322 *Bt* supplied YEG (Extended Data Fig. 7b). PUL71 responds to a glycan distinct from other well
323 characterized yeast-derived glycans including α-mannan, which activates P-*PUL36/68/69*

324 (Extended Data Fig. 2b)[36] and β-1,6-glucan, which stimulates P-*PUL56* (Fig. 2a)[32], because
325 bioluminescence from P-*PUL71* was unchanged when supplied these purified glycans (Fig. 1e)
326 and *Δsensor*$^{PUL71}$ exhibited similar activity to *wild-type Bt* harboring P-*PUL36* or P-*PUL56* when
327 supplied YEG (Extended Data Fig. 7b), indicating that Sensor$^{PUL71}$ recognizes a unique yeast-
328 derived glycan that increases *PUL71* expression.

329 To characterize the PUL71 target glycan, we co-incubated YEG with recombinant
330 hexahistidine-tagged SGBP$^{PUL71}$ (BT3961) and recovered protein-bound glycans using metal
331 affinity chromatography[18]. The recovered glycans were enriched for the PUL71 target over other
332 MACs present in YEG because the co-purifying material significantly increased activity from P-
333 *PUL71,* but P-*PUL7* (α-arabinan), P-*PUL22* (fructans), P-*PUL29* (unknown), P-*PUL48* (dextran),
334 P-*PUL66* (glycogen), and P-*PUL72* (high mannose N-glycan[36]) were no longer detectable
335 (Extended Data Fig. 7c). However, P-*PUL36/68/69* and P-*PUL56* also exhibited increased activity
336 indicating that α-mannan and β-glucan co-purified with the PUL71 target (Extended Data Fig. 7c).
337 Significantly activated PULs contributed to accessing the SGBP$^{PUL71}$ co-purifying material
338 because the corresponding sensor-deficient strains exhibited reduced growth on this material as
339 a sole carbon source, with *Δsensor*$^{PUL71}$ exhibiting the largest reduction (Extended Data Fig. 7d).
340 We hypothesized that the BT3961-purified material contained covalently tethered carbohydrates
341 comprised of the PUL71-, PUL36/68/69-, and PUL56- target glycans because the yeast cell-wall
342 is an interconnected matrix comprised of chitin, β1,3-, and β1,6-glucans decorated with
343 mannoproteins through a GPI-anchor remnant glycan (Extended Data Fig. 2a). Consistent with
344 this notion, material co-purifying with the SGBP$^{PUL56}$, BT3313, incubated with YEG, also increased
345 bioluminescence from strains harboring P-*PUL56*, P-*PUL36/68*, and P-*PUL71*, although
346 bioluminescence from P-*PUL69* was not significantly increased possibly because the
347 corresponding ligand was below the limit of detection (Extended Data Fig. 7e). Accordingly, only
348 *Δsensor*$^{PUL56}$, which is required for growth on β1,6-glucan (Extended Data Fig. 7f), exhibited
349 substantially reduced growth when supplied SGBP$^{PUL56}$-isolated material (Extended Data Fig. 7g),

350 indicating the other target glycans are present in relatively low abundances. Collectively, these
351 data demonstrate that the PUL56 and PUL71 SGBPs isolate their corresponding target glycans
352 when assembled into polymeric structures.

353 To determine how the PUL71-target glycan is tethered to β1,6-glucan and α-mannan, we
354 treated YEG with the PUL56-encoded β1,6-glucosidase (BT3312), referred to here as $GH20^{PUL56}$,
355 which liberated glucose and β-gentibiose from YEG (Extended Data Fig. 7h). We determined that
356 $GH^{PUL56}$ treatment removed β1,6-glucan, but not α-mannan, from the PUL71 target because
357 $SGBP^{PUL71}$-isolated material no longer increased bioluminescence from P-*PUL56* but exhibited
358 significant increases from P-*PUL36/68/69* (Fig. 5c). Accordingly, *Δsensor*$^{PUL71}$ and
359 *Δsensors*$^{PUL36/68/69}$ each exhibited reduced growth on this material as a sole carbon source and
360 growth was completely abolished in a strain lacking all four sensors (*Δsensors*$^{PUL36/68/69/71}$) (Fig.
361 5d), whereas these strains grew on YEG with reduced growth maxima (Fig. 5e). P-PUL71 likely
362 responds to a distinct manno-oligosaccharide because introduction of various purified substrates
363 were unable to increase bioluminescence (Extended Data Fig. 7i). Furthermore, compositional
364 analysis demonstrated that $SGBP^{PUL71}$-isolated material from untreated YEG contained glucose
365 (Extended Data Fig. 8a), whereas this monosaccharide was no longer detectable when YEG was
366 pre-treated with $GH20^{PUL56}$ (Extended Data Fig. 8b). Consistent with this finding, HPAEC-PAD
367 analysis revealed that $SGBP^{PUL56}$-isolated glycans from untreated YEG contained mannose
368 (Extended Data Fig. 8c), further suggesting that yeast β1,6-glucan and α-mannan are tethered in
369 the absence of mannoproteins, which are eliminated by ORNG extraction. These data indicate
370 that yeast α-mannan, β-glucan, and PUL71-target glycan are covalently assembled into a
371 complex resistant to oxidation, and that the PUL71 target glycan decorates yeast-derived α-
372 mannan prepared using ORNG (Fig. 4a) but not alkaline extraction, which was used to generate
373 the α-mannan standard (Fig. 1e). PUL71 includes the mannosidases BT3962 and BT3965, which
374 respectively convert α1,2- and α1,4-mannobioses into mannose (Extended Data Fig. 9a, b), and
375 a putative mannosidase BT3963, which was unable to hydrolyze any mannobiose examined here

376 (Extended Data Fig. 9c-f), in agreement with previous results[46]. Collectively, these data suggest
377 that the PUL71-target glycan is an unknown yeast α-mannan decoration comprised of unique
378 glycosidic linkages that can function as a prebiotic.
379
380 **Discussion**
381 We have developed a microbial glycomics platform that harnesses *Bacteroides* PUL sensors
382 to detect, isolate, and characterize MACs from biologically derived mixtures (Fig. 1). We
383 demonstrated that a genome-wide *Bt* PUL-reporter library can indicate the monosaccharide
384 composition, glycosidic linkages and d.o.p. of purified *Bt* accessible- (Fig. 1e, 2b) and -
385 inaccessible carbohydrates (Fig. 2d). Furthermore, we demonstrated this toolkit can efficiently
386 characterize MAC content in semi-purified preparations (Fig. 1e, Extended Data Figs. 1, 2) and
387 various mixtures derived from the host mucosa and dietary material (Fig. 1e and Fig. 4a). We
388 establish that reporter responses largely correspond to the biological source of each mixture (Fig.
389 4b), indicating that MAC content reflects phylogeny. The exquisite specificity and sensitivity of
390 *Bacteroides* reporter strains highlight their utility as glycan biosensors that could be implemented
391 in product quality control to detect carbohydrate contaminants (Extended Data Fig. 1b, c) and co-
392 purifying moieties (Fig. 3). Furthermore, this approach could be applied to broader glycomics
393 applications including surveillance of different carbohydrate structures from patient samples, plant
394 cultivars, and microbial isolates (Fig. 4a). Finally, *Bt* PUL reporters can directly identify novel
395 MACs that could be translated into prebiotics to manipulate the gut microbiome composition and
396 products for therapeutic purposes (Fig. 5). Therefore, PUL reporters offer a powerful new toolkit
397 that can be expanded using additional *Bacteroides* species that possess unique PUL repertoires
398 (Fig. 4c, Extended Data Fig. 6).
399
400 **Materials and Methods**
401 ***Bacterial growth conditions***

402 All bacteria were cultured as described previously[47]. Briefly, *Bacteroides* strains
403 (described in Supplementary Table 5) were cultured on brain-heart infusion agar (MilliporeSigma)
404 containing 5% horse blood (Hardy) under anaerobic conditions (85% $N_2$, 12.5% $CO_2$, 2.5% $H_2$).
405 Liquid cultures were inoculated from a single colony into TYG media and incubated under
406 anaerobic conditions before sub-culture into Bacteroides minimal media containing 100 mM
407 $KH_2PO_4$ (pH=7.2), 15 mM NaCl, 8.5 mM $(NH_4)_2SO_4$, 0.5 µg/ml of L-cysteine, 1.9 µM hematin,
408 200 µM L-histidine, 100 µM $MgCl_2$, 1.4 µM $FeSO_4$, 50 µM $CaCl_2$, 1 µg/ml of vitamin $K_3$, 5 ng/ml of
409 vitamin $B_{12}$ and individual carbon sources described in Supplementary Table 4. All bacterial
410 strains included the following antibiotics where appropriate: 100 µg/mL ampicillin, 200 µg/mL
411 gentamicin, 2 µg/mL tetracycline, or 25 µg/mL erythromycin (MilliporeSigma).
412
413 ***RNA-seq***
414 Strains engineered to express constitutively active PUL sensors or corresponding vector
415 control were cultured in triplicate in minimal media containing galactose as the sole carbon source
416 to mid-exponential phase ($OD_{600}$ = 0.45 - 0.7) before collection by centrifugation, frozen on dry
417 ice, and stored at -80°C. Cell pellets were treated with RNAprotect Bacteria Reagent (Qiagen)
418 and RNA was harvested using RNeasy Mini Kit with on-column DNase digestion (Qiagen)
419 according to the manufacturer's instructions. Purified RNA was treated with Turbo DNase
420 (Invitrogen) according to the manufacturer's instructions and re-purified using RNeasy Mini Kit
421 with on-column DNase digestion. RNA-sequencing was performed by SeqCoast Genomics using
422 a Nextseq 2000 (Illumina) following ribosomal RNA depletion using the Ribo-Zero Plus
423 Microbiome rRNA Depletion Kit (Illumina) to generate 12 million 150bp paired-end reads per
424 sample. Reads were aligned to the *B. thetaiotaomicron* genome (strain VPI-5482; GenBank
425 accession number NC_004663) in Galaxy using Bowtie2 (v2.5.3). The mapped reads were
426 quantified using featureCounts (v2.0.8) and differential gene expression was measured using

427 DEseq2 (v2.11.40.8) and values are listed in Supplementary Table 1. Raw files are available on
428 the NCBI Sequence Read Archive, PRJNA1298319.
429
430 ***Reporter construction***
431 Reporter plasmids were constructed by amplifying PUL promoter fragments
432 (Supplementary Table 2) using oligonucleotides (Supplementary Table 6) as previously
433 described[18]. Briefly, PUL promoters amplified using Q5 Master Mix (NEB) were combined with
434 pBolux[18] linearized with BamHI-HF and SpeI-HF (NEB) and NEBuilder Master Mix (NEB).
435 Reactions were incubated for 1 hour at 50°C before dialyzed on a 0.025 μM dialysis membrane
436 against water and subsequent electroporation into competent *E. coli* S17-1. All plasmids were
437 verified by Sanger sequencing.
438
439 ***Engineering Bt mutants***
440 Indicated *Bt* genomic deletions were generated using pEXCHANGE-*tdk* plasmids
441 (Supplementary Table 5) harboring flanking sequences, using amplicon-specific primers
442 (Supplementary Table 6), as previously described[48]. In short, pEXCHANGE constructs were
443 introduced via di-parental mating and chromosomal integration was validated by PCR. Parent
444 strains were counter-selected on solid media containing 200ng/mL 5-fluoro-2-deoxyuridine (DOT
445 Scientific). All strains were validated using linear amplicon sequencing.
446
447 ***Bt reporter array production***
448 All 91 reporter *Bt* strains described were cultured in TYG to stationary phase in a distinct
449 position of a 96 deep-well plate (Supplementary Table 2). Each strain was combined with glycerol
450 to a 10% final concentration, and 5 μL of the resulting mixture was dispensed into a sterile 96-
451 well plate using a Mini-96 (Integra), covered with adhesive sterile foil, and placed on dry ice before
452 ultimate storage at -80°C.

453

454 ***Bt reporter assays***

455 One replicate array plate was thawed and combined with pre-reduced TYG and cultured
456 overnight at 37°C under anaerobic conditions. The following day, each strain was diluted 50-fold
457 into minimal media containing 0.5% galactose and cultured to mid-logarithmic phase before
458 centrifugation at 2,204 x *g* to pellet cells, resuspension into 2x minimal media containing 0.2%
459 galactose and distribution into a sterile 384-well white, clear bottom microplate (Corning)
460 containing equal amounts of 0.8% carbon. Absorbance and bioluminescence were measured
461 using an Infinite M200 Plex (Tecan) every 15 minutes over 18 hours.

462

463 ***Biologically derived glycan mixture preparation***

464 Solid material was homogenized in 6% bleach (PureBright) and allowed to stir for 30
465 minutes at room temperature before addition of 1:100 volumes of formic acid were combined and
466 stirred for 5 additional minutes at room temperature. Reactions were centrifuged at 7200 x *g* for
467 10 minutes and the resulting supernatant was combined with bleach to a final concentration of
468 0.2% and stirred for 16 hours at room temperature before addition of equal volumes of formic acid
469 and stirring for an additional 5 minutes. Reactions were centrifuged at 7200 x *g* for 10 minutes
470 and the resulting supernatant was sterilized using a 0.22 µm vacuum filter (Millipore). The
471 resulting material was dialyzed with ultrapure $H_2O$ using 1KDa cutoff dialysis membrane for 3
472 days before lyophilization.

473

474 ***Human material collection***

475 The Pennsylvania State University College of Medicine Institutional Review Board
476 approved this study (IRB Protocol No. HY98-057EP-A). Prior to colectomy, patients gave
477 informed consent to have surgically resected tissue collected and banked into the Carlino Family
478 Inflammatory Bowel and Colorectal Disease Biobank. Diagnoses were confirmed using

479 colonoscopy and biopsy. Surgical specimens were immediately transferred from the operating
480 room to the surgical pathology lab, where the tissue was examined by a pathologist and surgeon.
481 The samples were then transported to the research laboratory on ice for further processing.
482 Mucosa was scraped from approximately 1 inch tissue sections and stored at -80°C until samples
483 were weighed, pooled, and subjected to glycan extraction.
484
485 ***Protein expression and purification***
486 SGBPs and GHs were recombinantly expressed as previously described[47]. Briefly, inserts
487 encoding each enzyme were amplified from *B. thetaiotaomicron* VPI-5482 genomic DNA and
488 cloned into pT7-7-N6H4A[18] linearized with NotI-HF and HindIII-HF using NEBuilder Hi-Fi Master
489 Mix (NEB). The resulting plasmids were introduced into *E. coli* S17-1 and verified by Sanger
490 sequencing. Plasmids were introduced into BL21 and plated on selective media. A single colony
491 was inoculated into Luria Bertani broth (BD) and cultured to mid-logarithmic phase before 50 μM
492 IPTG addition and subsequent incubation for 4 hours at 30°C. Pelleted cells were lysed in Lysis
493 Buffer (20 mM Tris (pH, 8) and 100 mM sodium chloride) and N-terminally hexa-histidine-tagged
494 proteins were recovered following co-incubation with $Ni^{2+}$-NTA resin (Thermo) and elution with
495 Lysis Buffer containing 25 mM Histidine. Recovered proteins were buffer exchanged in 10 mM
496 Tris (pH, 7.4) containing 10% glycerol using appropriate molecular weight cut-off centrifugal
497 concentrators (MilliporeSigma). Protein concentrations were determined using a BCA protein
498 quantification kit (Thermo).
499
500 ***SGBP-mediated glycan isolation***
501 Glycans were recovered from mixtures using recombinant SGBPs as previously
502 described[18]. Briefly, an *Ec* strain BL21 cell lysate expressing N-terminal hexa-his-tagged SGBP
503 was combined with 0.5% glycan and 1.0 mL $Ni^{2+}$-NTA resin (Thermo) and incubated overnight at
504 4°C with rocking. The mixture was packed into a 3 mL gravity flow column and washed with 20

505 mM Tris (pH=8) and 100mM Sodium Chloride 8. Protein-glycan complexes were eluted with 25
506 mM histidine (MilliporeSigma), treated with proteinase K before incubation at 65°C for 2 hours.
507 The reaction was combined with 3x volume of 100% ethanol incubated overnight at 4°C with
508 rocking and centrifuged for 15 minutes at 21,291 x *g*. The pelleted material was dried under filtered
509 air, resuspended in water and combined before lyophilization.
510
511 ***qPCR***
512 Stationary phase cultures were diluted 50-fold into pre-reduced minimal media containing the
513 0.5% galactose and incubated to mid-logarithmic phase ($OD_{600}$ = 0.45 - 0.65). 1.0 mL of culture
514 was pelleted by centrifugation, immediately placed on dry ice, and stored at -80°C Remaining
515 cultures were pelleted by centrifugation at 7,200 x *g* for 3 minutes and resuspended in pre-
516 reduced 2X minimal media containing no carbon. Equal amounts of resuspended culture and
517 carbon were combined for a final concentration of 0.5% carbon and incubated for 60-minutes
518 before 1.0 mL of culture was pelleted by centrifugation, immediately placed on dry ice, and stored
519 at -80°C. RNA was isolated using the RNeasy Mini Kit (QIAGEN), quantified by absorbance, and
520 1 µg was converted to cDNA following addition of the Superscript IV VILO Master Mix with
521 ezDNAse (Invitrogen) per the manufacturer's directions. Transcript abundances were measured
522 as previously described[47] using a QuantStudio5 (ThermoFisher) and PowerUp SYBR Green
523 Master Mix (ThermoFisher) with amplicon specific primers (Supplementary Table 3). The 16S
524 *rRNA, rrs,* was used as the reference gene for all qPCR experiments.
525
526 ***In vivo competitive fitness of B. thetaiotaomicron strains***
527 All animal experiments were performed in accordance with protocols approved by Penn State
528 Institutional Animal Care and Use Committee. 8-12 week old, mixed-gender, germ-free C57/BL6
529 mice were maintained in flexible plastic gnotobiotic isolators with a 12-hour light/dark cycle and
530 provided a standard, autoclaved mouse chow (LabDiet, 5021) *ad libitum*[47]. Mice were randomly

531 distributed into 2 groups, which were supplied either autoclaved drinking water or 0.5% yeast
532 extract glycans, 2 days prior to gavage with $10^8$ CFU of each indicated strains suspended in
533 200 µL of phosphate-buffered saline. Input (day 0) abundance of each strain was determined by
534 counting colony forming units following plating on solid media. Fresh fecal pellets were collected
535 at the indicated days and genomic DNA was extracted as described previously[23]. The abundance
536 of each strain was measured by qPCR, using barcode-specific primers (Supplemental Table 2)
537 as described previously[47].
538
539 ***Monosaccharide Composition Analysis***
540 Trifluoroacetic acid was added to the indicated glycan preparations to a final concentration of 2N
541 and heated to 100°C for 4 hours. Reactions were cooled to room temperature, dried under filtered
542 air, and washed twice with 50% isopropanol before being dried to completion. Reactions were
543 resuspended in 0.2 mL of deionized water and analyzed by HPAEC-PAD.
544
545 ***CAZyme activity***
546 YEG was combined with 200 nM $GH20^{PUL56}$ to a final concentration of 1.0% in 20 mM Tris (pH=8),
547 incubated for 4 hours at 37 °C, and analyzed by HPAEC-PAD. Mannobioses were combined with
548 200 nM of each PUL71 GH in 20 mM Tris (pH=8) and 1 µM calcium chloride.
549
550 ***Software and statistics***
551 Data collection and curation was done in Microsoft Excel. All data was plotted in GraphPad Prism,
552 except for heatmaps and PCoA, which were generated using ggplot2 in RStudio. Statistical
553 analyses were calculated in Prism. For all experiments, N denotes individual biological replicates
554 across at least two independent experiments.
555
556 ***Resource Availability***

557 All PUL reporter plasmids are available on addgene.com and a complete *Bt* reporter array will be
558 supplied upon request. All other *Bt* and *Ec* strains used in this work are available upon request.
559
560 **Supplementary Information**
561 **Supplementary Table 1. RNA-seq analysis of *Bt* strains expressing constitutively active**
562 **PUL sensors.**
563 **Supplementary Table 2. *Bt* PUL reporter plasmids.**
564 **Supplementary Table 3. PUL similarity across *Bacteroides* species.**
565 **Supplementary Table 4. Carbohydrates examined in this study.**
566 **Supplementary Table 5. Strains and plasmids used in this study.**
567 **Supplementary Table 6. Oligonucleotides used in this study.**
568
569 **Acknowledgements**
570 We thank Walter Koltun, Leonard Harris, Troy Ott, Wesley Raup-Konsavage, and Hannah Valensi
571 for assistance producing glycan extracts. We thank Biswa Chourdry and Richard Helm for
572 assistance with carbohydrate compositional assistance. Additionally, we thank Jordan Bisanz for
573 assistance with animal experiments and useful discussions. This project is funded, in-part by
574 National Institutes of Health Grants GM147178 and DK132711 to G.E.T. and a grant with the
575 Pennsylvania Department of Health using Tobacco CURE Funds. The PA Department of Health
576 specifically disclaims responsibility for any analyses, interpretations or conclusions. This work
577 was supported by the Carlino fund to The Penn State Colorectal Disease Biobank.
578

579 **References**


580 1. Luis, A.S. et al. Dietary pectic glycans are degraded by coordinated enzyme pathways in
581 human colonic Bacteroides. *Nat Microbiol* **3**, 210-219 (2018).
582 2. Marcobal, A. et al. Bacteroides in the infant gut consume milk oligosaccharides via
583 mucus-utilization pathways. *Cell Host Microbe* **10**, 507-514 (2011).
584 3. Martens, E.C., Chiang, H.C. & Gordon, J.I. Mucosal glycan foraging enhances fitness
585 and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**,
586 447-457 (2008).
587 4. Martens, E.C. et al. Recognition and degradation of plant cell wall polysaccharides by
588 two human gut symbionts. *PLoS Biol* **9**, e1001221 (2011).
589 5. Salyers, A.A., Vercellotti, J.R., West, S.E. & Wilkins, T.D. Fermentation of mucin and
590 plant polysaccharides by strains of Bacteroides from the human colon. *Appl Environ*
591 *Microbiol* **33**, 319-322 (1977).
592 6. Raghavan, V., Lowe, E.C., Townsend, G.E., 2nd, Bolam, D.N. & Groisman, E.A. Tuning
593 transcription of nutrient utilization genes to catabolic rate promotes growth in a gut
594 bacterium. *Mol Microbiol* **93**, 1010-1025 (2014).
595 7. Sonnenburg, E.D. et al. Specificity of polysaccharide use in intestinal bacteroides
596 species determines diet-induced microbiota alterations. *Cell* **141**, 1241-1252 (2010).
597 8. Bakshani, C.R. et al. Carbohydrate-active enzymes from Akkermansia muciniphila break
598 down mucin O-glycans to completion. *Nat Microbiol* **10**, 585-598 (2025).
599 9. Cartmell, A. et al. A surface endogalactanase in Bacteroides thetaiotaomicron confers
600 keystone status for arabinogalactan degradation. *Nat Microbiol* **3**, 1314-1326 (2018).
601 10. Ndeh, D. et al. Metabolism of multiple glycosaminoglycans by Bacteroides
602 thetaiotaomicron is orchestrated by a versatile core genetic locus. *Nat Commun* **11**, 646
603 (2020).
604 11. Ndeh, D. et al. Complex pectin metabolism by gut bacteria reveals novel catalytic
605 functions. *Nature* **544**, 65-70 (2017).
606 12. McNulty, N.P. et al. Effects of diet on resource utilization by a model human gut
607 microbiota containing Bacteroides cellulosilyticus WH2, a symbiont with an extensive
608 glycobiome. *PLoS Biol* **11**, e1001637 (2013).
609 13. Whitaker, W.R. et al. Controlled colonization of the human gut with a genetically
610 engineered microbial therapeutic. *Science* **389**, 303-308 (2025).
611 14. Paulson, J.C., Blixt, O. & Collins, B.E. Sweet spots in functional glycomics. *Nat Chem*
612 *Biol* **2**, 238-248 (2006).
613 15. Heimburg-Molinaro, J., Mehta, A.Y., Tilton, C.A. & Cummings, R.D. Insights Into
614 Glycobiology and the Protein-Glycan Interactome Using Glycan Microarray
615 Technologies. *Mol Cell Proteomics* **23**, 100844 (2024).
616 16. Ruhaak, L.R., Xu, G., Li, Q., Goonatilleke, E. & Lebrilla, C.B. Mass Spectrometry
617 Approaches to Glycomic and Glycoproteomic Analyses. *Chem Rev* **118**, 7886-7930
618 (2018).
619 17. Raghavan, V. & Groisman, E.A. Species-specific dynamic responses of gut bacteria to a
620 mammalian glycan. *J Bacteriol* **197**, 1538-1548 (2015).
621 18. Modesto, J.L., Pearce, V.H. & Townsend, G.E., 2nd Harnessing gut microbes for glycan
622 detection and quantification. *Nat Commun* **14**, 275 (2023).
623 19. Glowacki, R.W.P. & Martens, E.C. If you eat it, or secrete it, they will grow: the
624 expanding list of nutrients utilized by human gut bacteria. *J Bacteriol* **203** (2020).
625 20. Schwalm, N.D., 3rd & Groisman, E.A. Navigating the Gut Buffet: Control of
626 Polysaccharide Utilization in Bacteroides spp. *Trends Microbiol* **25**, 1005-1015 (2017).

627 21. Townsend, G.E., 2nd, Raghavan, V., Zwir, I. & Groisman, E.A. Intramolecular
628 arrangement of sensor and regulator overcomes relaxed specificity in hybrid two-
629 component systems. *Proc Natl Acad Sci U S A* **110**, E161-169 (2013).
630 22. Cao, Y., Rocha, E.R. & Smith, C.J. Efficient utilization of complex N-linked glycans is a
631 selective advantage for Bacteroides fragilis in extraintestinal infections. *Proc Natl Acad*
632 *Sci U S A* **111**, 12901-12906 (2014).
633 23. Townsend, G.E., 2nd et al. A Master Regulator of Bacteroides thetaiotaomicron Gut
634 Colonization Controls Carbohydrate Utilization and an Alternative Protein Synthesis
635 Factor. *mBio* **11** (2020).
636 24. Ryan, D. et al. An integrated transcriptomics-functional genomics approach reveals a
637 small RNA that modulates Bacteroides thetaiotaomicron sensitivity to tetracyclines.
638 *bioRxiv* (2023).
639 25. Ryan, D., Jenniches, L., Reichardt, S., Barquist, L. & Westermann, A.J. A high-resolution
640 transcriptome map identifies small RNA regulation of metabolism in the gut microbe
641 Bacteroides thetaiotaomicron. *Nat Commun* **11**, 3557 (2020).
642 26. Terrapon, N. et al. PULDB: the expanded database of Polysaccharide Utilization Loci.
643 *Nucleic Acids Res* **46**, D677-D683 (2018).
644 27. Terrapon, N., Lombard, V., Gilbert, H.J. & Henrissat, B. Automatic prediction of
645 polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**, 647-655
646 (2015).
647 28. Glowacki, R.W.P. et al. A Ribose-Scavenging System Confers Colonization Fitness on
648 the Human Gut Symbiont Bacteroides thetaiotaomicron in a Diet-Specific Manner. *Cell*
649 *Host Microbe* **27**, 79-92 e79 (2020).
650 29. Reeves, A.R., D'Elia, J.N., Frias, J. & Salyers, A.A. A Bacteroides thetaiotaomicron outer
651 membrane protein that is essential for utilization of maltooligosaccharides and starch. *J*
652 *Bacteriol* **178**, 823-830 (1996).
653 30. D'Elia, J.N. & Salyers, A.A. Effect of regulatory protein levels on utilization of starch by
654 Bacteroides thetaiotaomicron. *J Bacteriol* **178**, 7180-7186 (1996).
655 31. Jones, D.R. et al. Engineering dual-glycan responsive expression systems for tunable
656 production of heterologous proteins in Bacteroides thetaiotaomicron. *Sci Rep* **9**, 17400
657 (2019).
658 32. Temple, M.J. et al. A Bacteroidetes locus dedicated to fungal 1,6-beta-glucan
659 degradation: Unique substrate conformation drives specificity of the key endo-1,6-beta-
660 glucanase. *J Biol Chem* **292**, 10639-10650 (2017).
661 33. Liu, H. et al. Functional genetics of human gut commensal Bacteroides thetaiotaomicron
662 reveals metabolic requirements for growth across environments. *Cell Rep* **34**, 108789
663 (2021).
664 34. Liou, C.S. et al. A Metabolic Pathway for Activation of Dietary Glucosinolates by a
665 Human Gut Symbiont. *Cell* **180**, 717-728 e719 (2020).
666 35. Pudlo, N.A. et al. Phenotypic and Genomic Diversification in Complex Carbohydrate-
667 Degrading Human Gut Bacteria. *mSystems* **7**, e0094721 (2022).
668 36. Cuskin, F. et al. Human gut Bacteroidetes can utilize yeast mannan through a selfish
669 mechanism. *Nature* **517**, 165-169 (2015).
670 37. Abbott, D.W., Martens, E.C., Gilbert, H.J., Cuskin, F. & Lowe, E.C. Coevolution of yeast
671 mannan digestion: Convergence of the civilized human diet, distal gut microbiome, and
672 host immunity. *Gut Microbes* **6**, 334-339 (2015).
673 38. Lowe, E.C., Basle, A., Czjzek, M., Firbank, S.J. & Bolam, D.N. A scissor blade-like
674 closing mechanism implicated in transmembrane signaling in a Bacteroides hybrid two-
675 component system. *Proc Natl Acad Sci U S A* **109**, 7298-7303 (2012).

39. Schwalm, N.D., 3rd, Townsend, G.E., 2nd & Groisman, E.A. Multiple Signals Govern Utilization of a Polysaccharide in the Gut Bacterium Bacteroides thetaiotaomicron. *mBio* **7** (2016).
40. Ryan, D. et al. An expanded transcriptome atlas for Bacteroides thetaiotaomicron reveals a small RNA that modulates tetracycline sensitivity. *Nat Microbiol* **9**, 1130-1144 (2024).
41. Song, X. et al. Oxidative release of natural glycans for functional glycomics. *Nat Methods* **13**, 528-534 (2016).
42. Vos, G.M. et al. Oxidative Release of O-Glycans under Neutral Conditions for Analysis of Glycoconjugates Having Base-Sensitive Substituents. *Anal Chem* **95**, 8825-8833 (2023).
43. Wu, M. et al. Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut Bacteroides. *Science* **350**, aac5992 (2015).
44. Feng, J. et al. Polysaccharide utilization loci in Bacteroides determine population fitness and community-level interactions. *Cell Host Microbe* **30**, 200-215 e212 (2022).
45. Kabonick, S.G. et al. Hierarchical glycolytic pathways control the carbohydrate utilization regulator in human gut Bacteroides. *Nat Commun* **16**, 4488 (2025).
46. Zhu, Y. et al. Mechanistic insights into a Ca2+-dependent family of alpha-mannosidases in a human gut symbiont. *Nat Chem Biol* **6**, 125-132 (2010).
47. Pearce, V.H., Groisman, E.A. & Townsend, G.E., 2nd Dietary sugars silence the master regulator of carbohydrate utilization in human gut Bacteroides species. *Gut Microbes* **15**, 2221484 (2023).
48. Koropatkin, N.M., Martens, E.C., Gordon, J.I. & Smith, T.J. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**, 1105-1115 (2008).

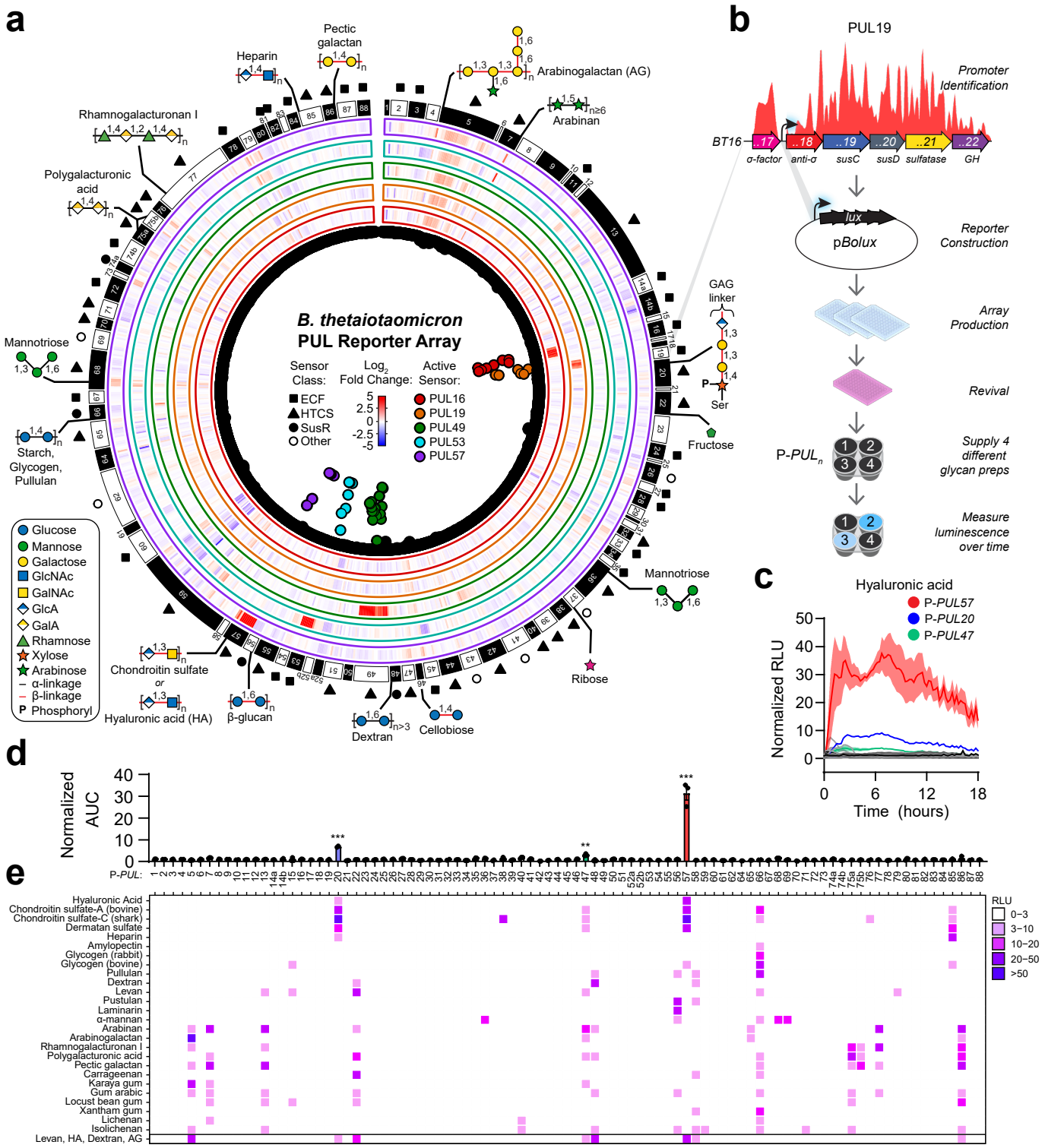



**Fig. 1. Construction of a genome-wide *Bt* PUL reporter library array. a,** Cartoon depicting 88 *Bt* PULs denoted by individual genes (innermost black circles). Dots emanating inwards represent significantly increased PUL transcripts measured by RNA-seq from strains engineered to produce constitutively active sensor proteins from PUL16 (red), PUL19 (orange), PUL49 (green), PUL53 (cyan), or PUL57 (purple) and color-corresponding heatmaps are positioned centrally. The outermost symbols represent distinct classes of known or predicted PUL sensor proteins (filled square: ECF, extracytoplasmic sigma factor; filled triangle: HTCS, hybrid two-component system; filled circle: SusR, starch utilization system regular; open circle: other classes) and corresponding glycan structures that activate them emanate radially. **b,** schematic depicting PUL reporter construction and glycan analysis using the arrayed *Bt* library. **c-d,** Bioluminescence (**c**) from all 91 *Bt* reporters over 18 hours following introduction of hyaluronic acid and normalized by identical strains supplied galactose and corresponding (**d**) AUC normalized by responses from a strain containing p*Bolux*. All significantly increased reporters have corresponding colors and insignificant increases are displayed in grayscale. N=3, error is SEM. **e,** Heatmap of AUC over 18 hours from all reporters supplied mixtures containing galactose and various purified polysaccharides. N≥2. All statistics were calculated using 2-way ANOVA with Fisher's LSD test and *** indicates *P*-values < 0.001, and ** < 0.01.

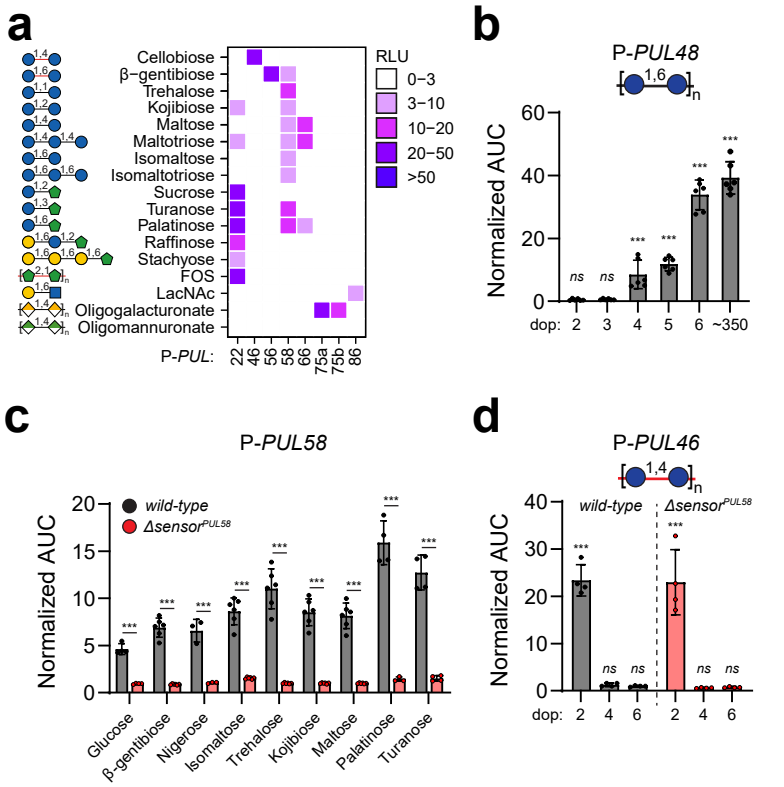



**Fig. 2. *Bt* PUL reporters distinguish between monosaccharide composition, d.o.p., and glycosidic linkage. a,** A heatmap displaying normalized AUC following the introduction of various oligosaccharides to the *Bt* PUL reporter array. n=2. **b,** Normalized AUC from P-*PUL48* supplied isomaltooligosaccharides with increasing d.o.p. N=6, error is SEM. **c,** Normalized AUC from *wild-type* (black) or *sensor-PUL58*-deficient (*ΔBT2160*, red) strains harboring P-*PUL58* supplied the indicated di- and tri-saccharides. N≥3, error is SEM. **d,** Normalized AUC from *wild-type* (black) or *ΔBT2160* (red) strains harboring P-*PUL46* supplied the indicated cello-oligosaccharides. N=4, error is SEM. For panels **b-d**, *P*-values were computed using 2-way ANOVA. *** indicates values < 0.001 and *ns* indicates values > 0.05.

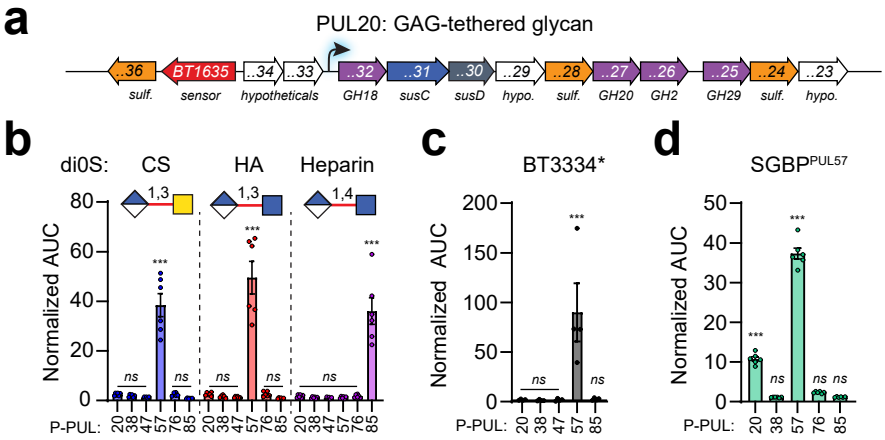



**Fig. 3. PUL 20 responds to a unique GAG-tethered glycan. a,** cartoon representing PUL20 activated by GAGs. **b**, Normalized bioluminescence from *Bt* strains harboring P-*PUL20,* P-*PUL38,* P-*PUL47,* P-*PUL57,* or P-*PUL85* supplied unsaturated disaccharides (di0S) derived from CS (blue), HA (red), or heparin (purple) and normalized by responses from a strain harboring p*Bolux*. N=6. **c,** Normalized bioluminescence from a strain expressing BT3334* harboring the reporters described in (**b**) and normalized by identical strains harboring the empty vector cultured in galactose as a sole carbon source. N=4. **d,** Normalized bioluminescence from *Bt* strains harboring P-*PUL20,* P-*PUL38,* P-*PUL47,* P-*PUL57,* or P-*PUL85* supplied SGBP$^{PUL57}$ co-purified material following incubation with shark CS and normalized by responses from a strain harboring p*Bolux*. N=6. For panels **b-d**, error is SEM and *P*-values were calculated using 2-way ANOVA with Fisher's LSD test. *** indicates values < 0.001 and *ns* > 0.05.

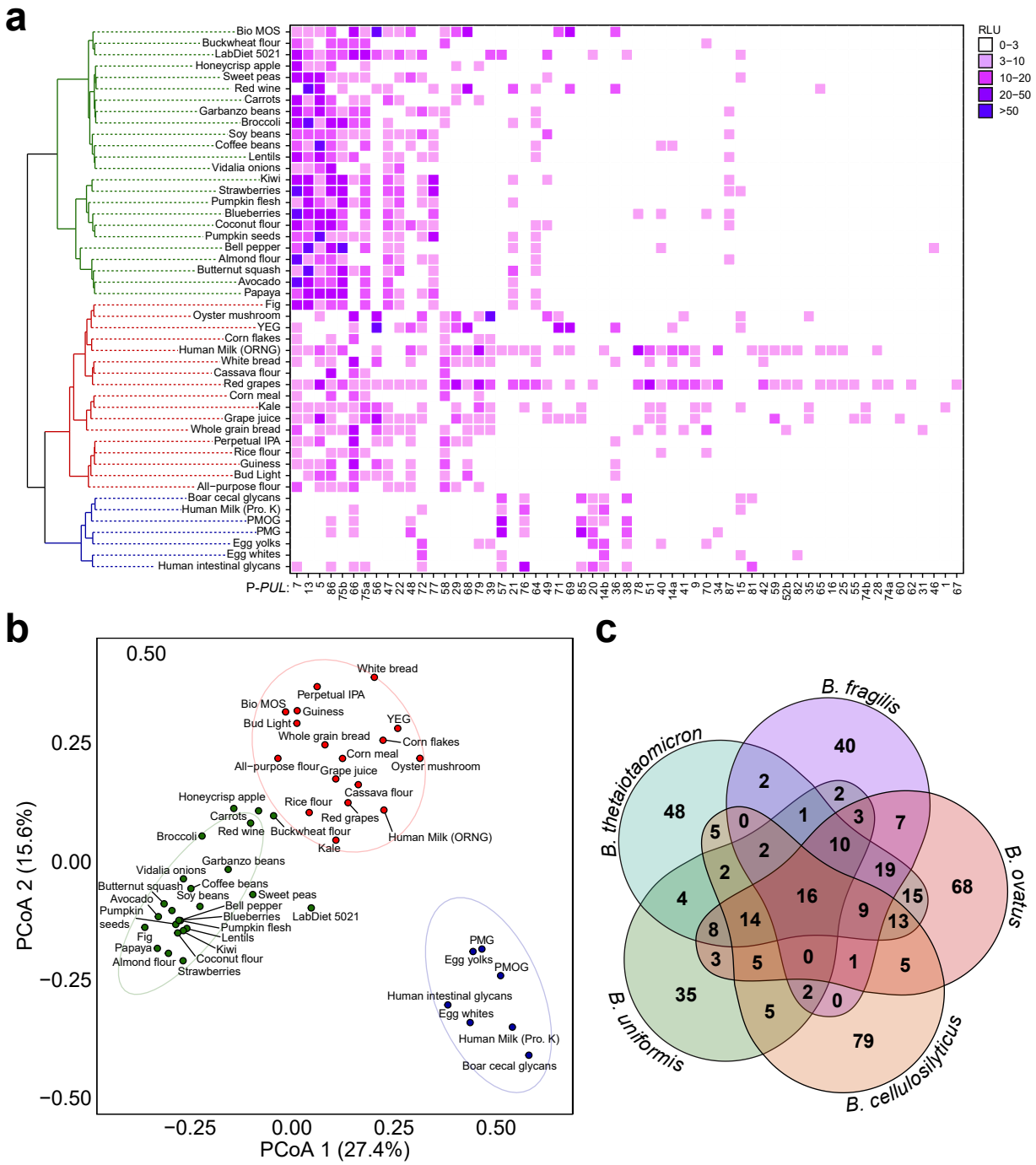



**Fig. 4. Arrayed *Bt* PUL reporter strains differentiate MAC sources. a,** Heatmap representing *Bt* PUL array responses to the indicated biologically-derived mixtures clustered by reporter activation similarity. n≥2. **b,** PCoA of *Bt* reporter responses to different MACs from (**a**) using the Jaccard similarity index and Ward D2 distance calculations. **c,** Distribution of unique and shared predicted PULs across the indicated *Bacteroides* species.

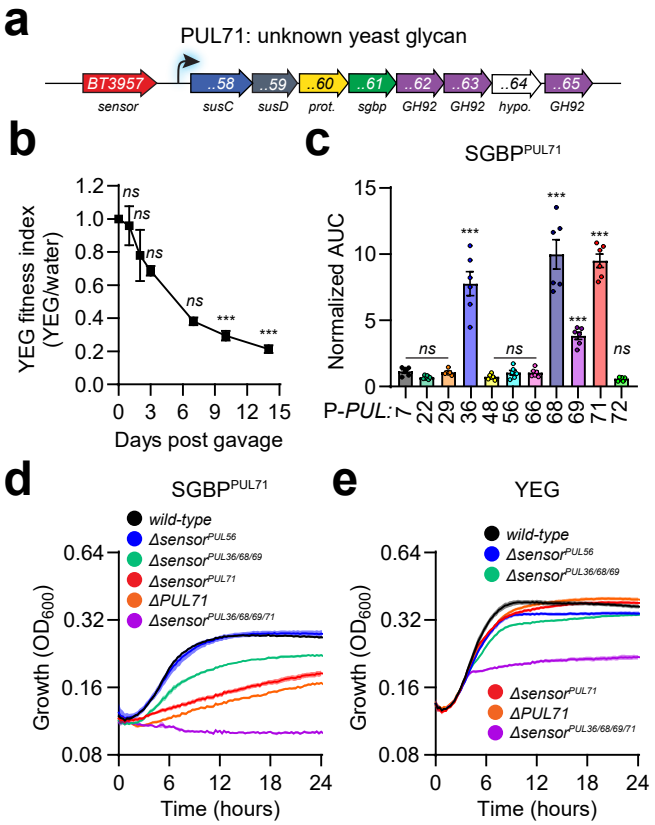
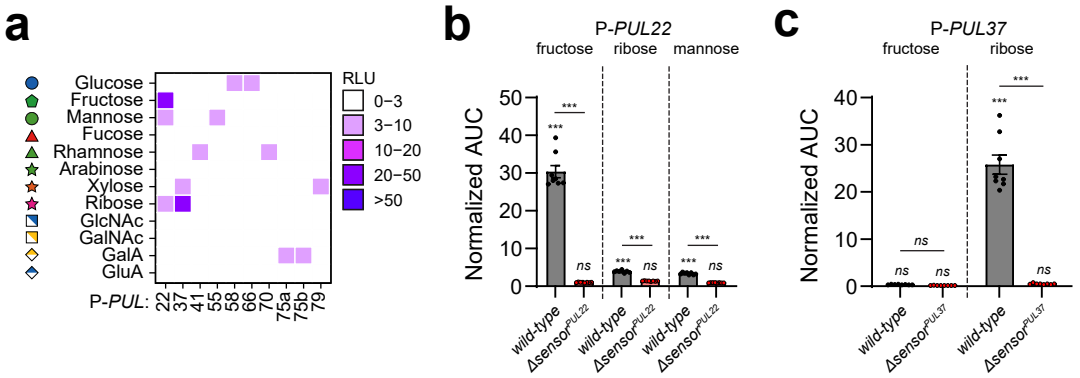
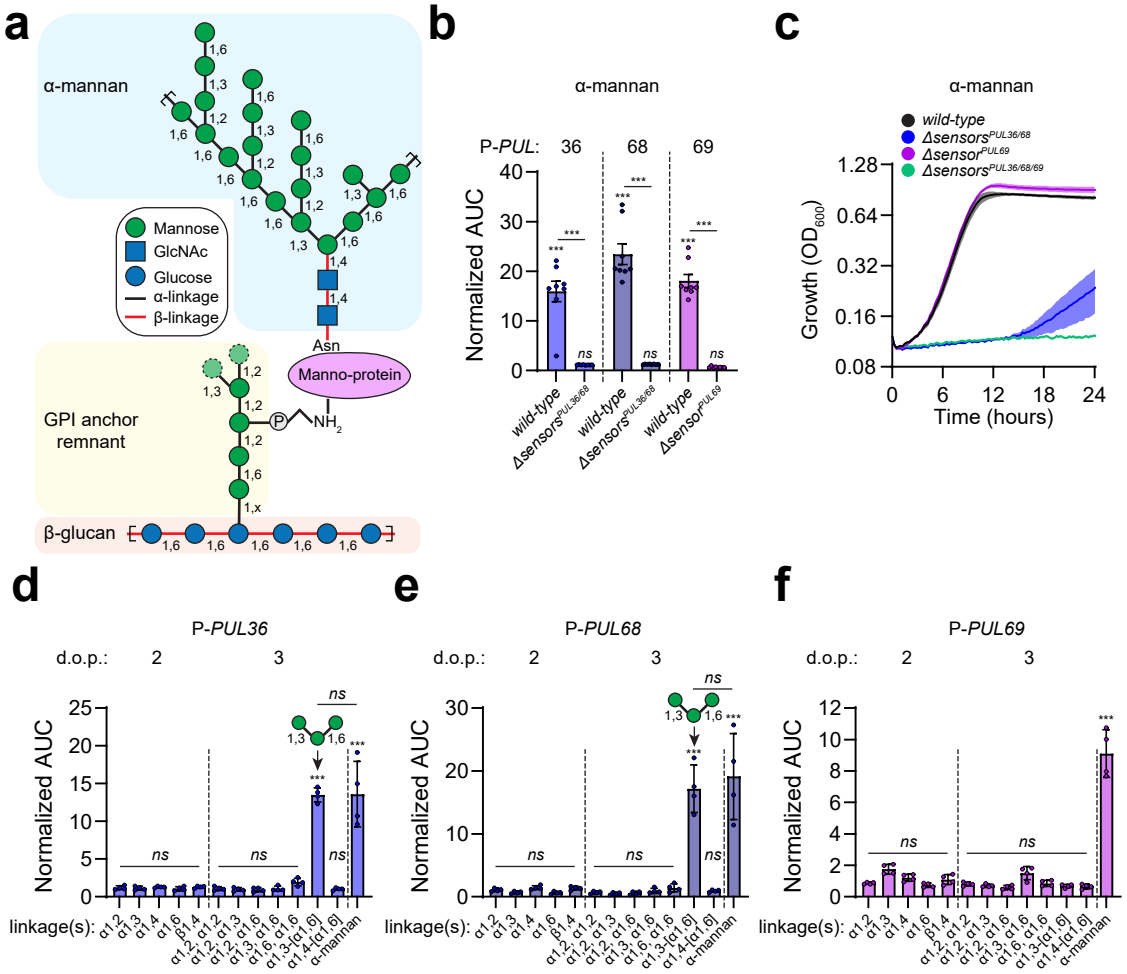


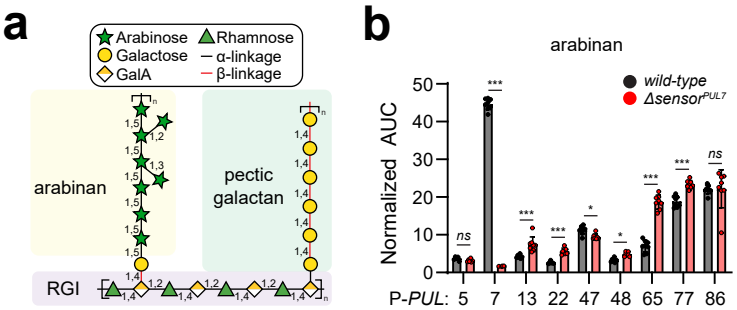

**Fig. 5. PUL71 consumes an unknown yeast cell wall glycan. a,** Cartoon schematic of PUL71, which responds to yeast glycans. **b,** Fitness index of a *PUL71*-deficient strain relative to *wild-type Bt* in mice supplied a 0.5% YEG solution compared to those supplied water. N=5 and error is SEM. *P*-values were calculated using 2-way ANOVA with Bonferroni correction. *** indicates values < 0.001 and *ns* > 0.05. **c,** Normalized bioluminescence of YEG-induced reporter strains supplied SGBP$^{PUL71}$-isolated glycans from GH30$^{PUL56}$-treated YEG. N=6, error is SEM. *P*-values were calculated using 2-way ANOVA with Fisher's LSD test. *** indicates values < 0.001 and *ns* > 0.05. **d&e,** Growth of *wild-type* and the indicated sensor-deficient *Bt* strains supplied (**d**) SGBP$^{PUL71}$-co-purifying material following incubation with GH30$^{PUL56}$-treated YEG or (**e**) untreated YEG. N=6, error is SEM.



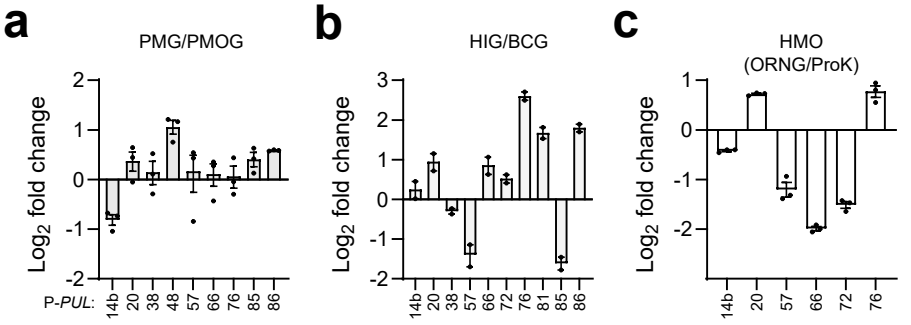

**Extended Data Fig. 1. Monosaccharide-responsive PUL reporters exhibit extraordinary sensitivity. a,** A heatmap displaying normalized bioluminescence responses to monosaccharides. N≥2. **b-c**, Normalized bioluminescence from *wild-type* (black) or corresponding *sensor*-deficient (red) *Bt* strains harboring (**b**) P-*PUL22* or (**c**) P-*PUL37* supplied the indicated monosaccharides. N=8, error is SEM. *P*-values were calculated using 2-way ANOVA with Fisher's LSD test. *** indicates values < 0.001 and *ns* > 0.05.



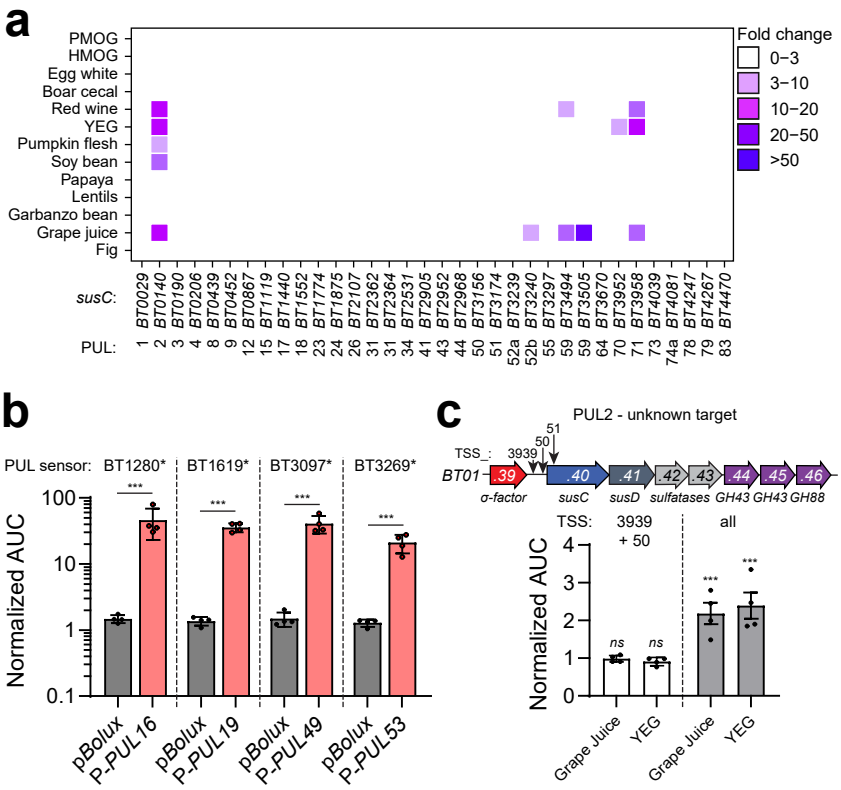

**Extended Data Fig. 2. *Bt* mannan PUL reporters detect structurally distinct manno-oligosaccharides. a,** Cartoon representation of yeast cell wall glycan structures containing mannoproteins decorated with α-mannan tethered to β 1,6-glucan by a GPI-anchor remnant glycan. **b**, Normalized bioluminescence from *wild-type*, *sensor*$^{PUL36/68}$-, or *sensor*-$^{PUL69}$- deficient *Bt* strains harboring P-*PUL36, P-PUL68*, *or* P-*PUL69* supplied α-mannan and normalized by responses from isogenic strains harboring p*Bolux*. N=8, error is SEM. **c,** Growth of strains described in (**b**) supplied α-mannan as a sole carbon source. N=8, error is SEM. **d-f,** strains harboring (**d**) P-*PUL36*, (**e**) P-*PUL68*, or (**f**) P-*PUL69* were supplied the indicated manno-di-, tri-saccharides, or poly-saccharides. n=4, error is SD. For panels **b,d-f,** *P*-values were calculated using 2-way ANOVA with Fisher's LSD test. *** indicates values < 0.001 and *ns* > 0.05.



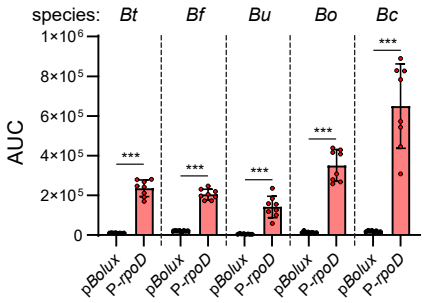

**Extended Data Fig. 3. *Bt* PUL reporters recognize discreet glycan moieties from plant pectins. a,** Cartoon representation of plant pectin structures including arabinan (PUL7), rhamnogalacturonan I (RGI, PUL77), and pectic galactan (PUL86). **b**, Normalized bioluminescence from *wild-type or Δsensor*$^{PUL7}$ *Bt* strains harboring the indicated PUL reporters supplied arabinan and normalized by responses from a strain harboring p*Bolux*. n=8, error is SEM. *P*-values were calculated using 2-way ANOVA with Fisher's LSD test and * indicates values <0.05, *** < 0.001, and *ns* > 0.05.



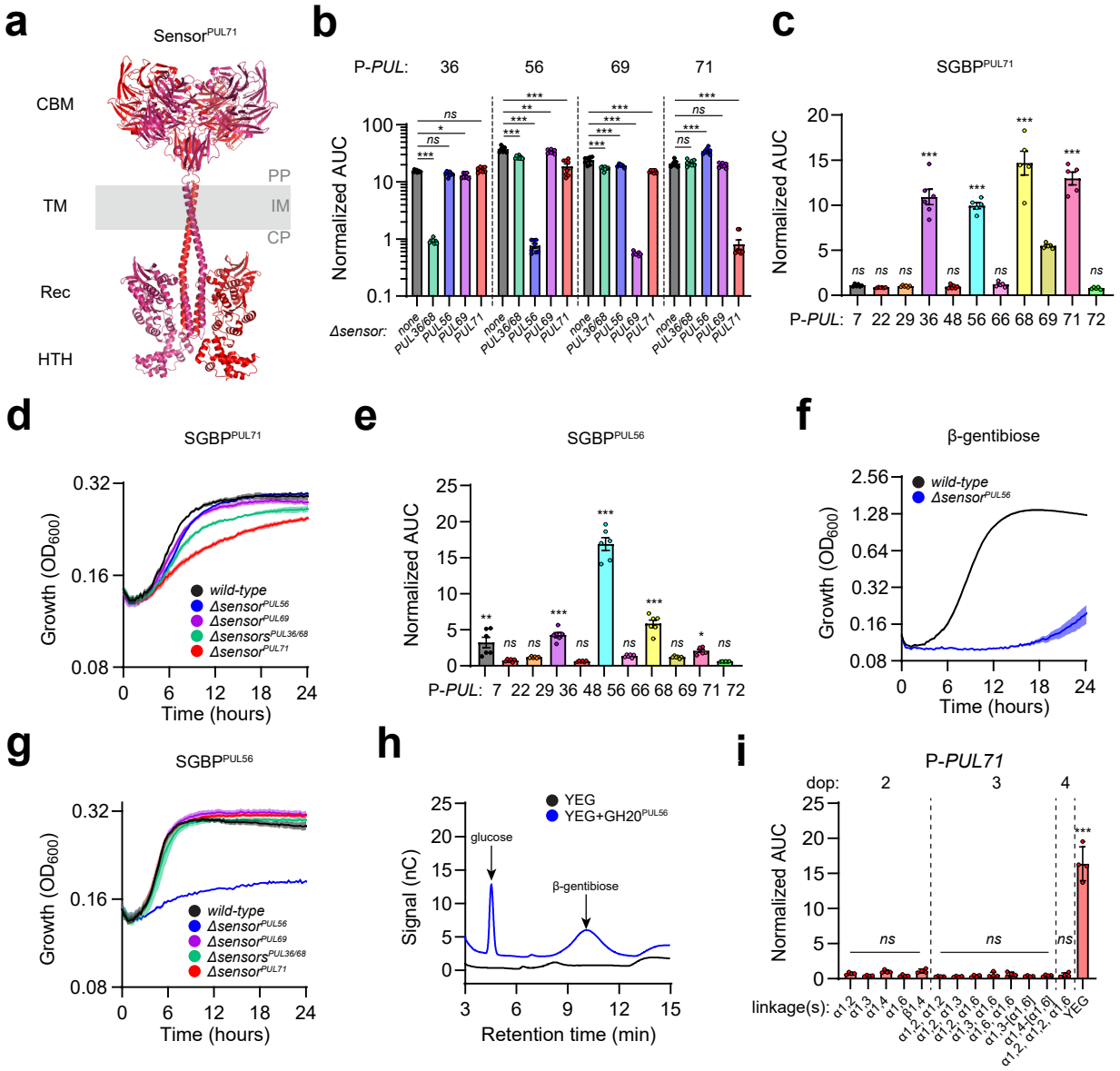

**Extended Data Fig. 4. Extraction- and species-specific reporter responses. a-c,** Log2 fold differences between reporter responses from relevant strains supplied (**a**) porcine gastric mucin prepared using ORNG (PMG) or β-elimination (PMOG), (**b**) fresh human (HIG) and boar (BCG), and (**c**) human milk oligosaccharides (HMO) prepared using ORNG or enzymatic treatment (ProK). N≥2, error is SD.



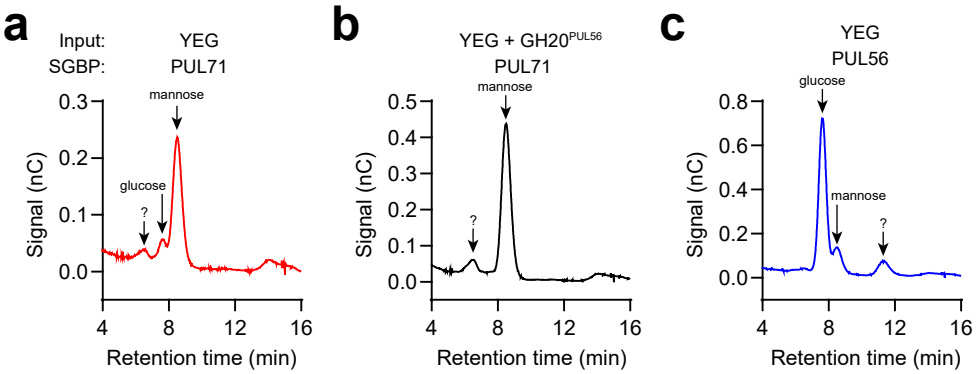

**Extended Data Fig. 5. PUL expression is faithfully reported. a,** Heatmap depicting the indicated *susC* transcript measurements from *Bt* cultures supplied the indicated glycan mixtures. N=3. **b,** Bioluminescence from strains expressing the indicated constitutively active PUL sensor and either p*Bolux* (black) or the corresponding PUL reporter (red) normalized by strains harboring an empty vector. N=4, error is SEM. **c,** Cartoon of PUL2 (top) and normalized bioluminescence (bottom) from 2 different PUL reporters containing distinct TSS combinations supplied grape juice or yeast extract glycans (YEG). N=4, error is SEM. For panels **b,c**, *P*-values were calculated using 2-way ANOVA with Fisher's LSD test. *** indicates values < 0.001 and *ns* > 0.05.



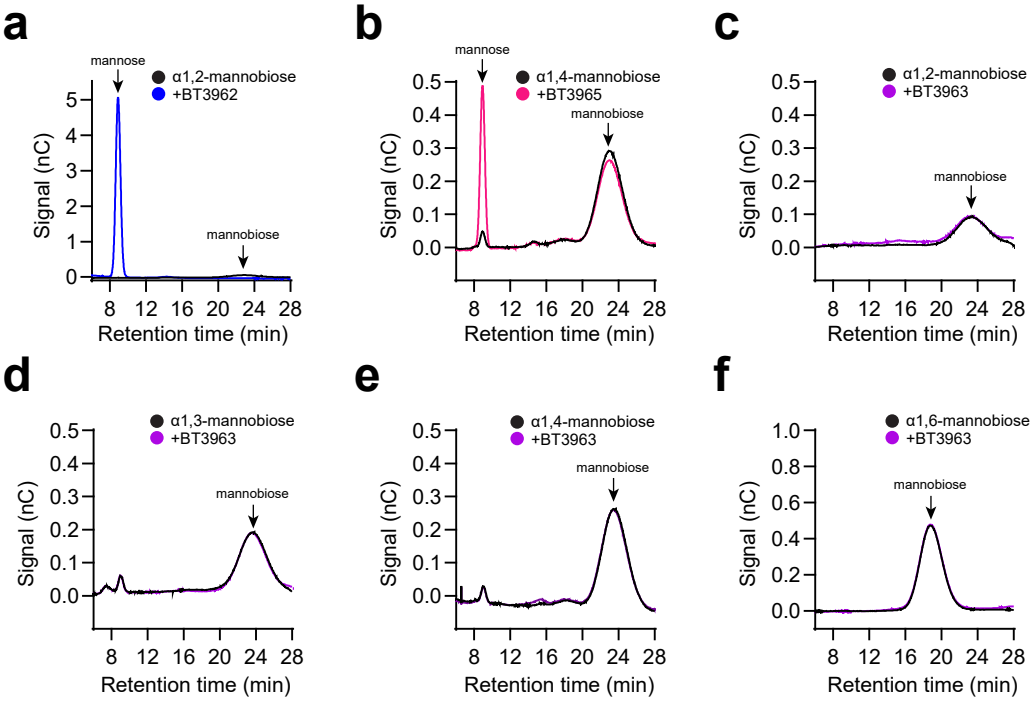

**Extended Data Fig. 6. Various *Bacteroides* species can be converted into glycan biosensors.** Normalized bioluminescence values from *Bt*, *B. fragilis* (*Bf*), *B. uniformis* (*Bu*), *B. ovatus* (*Bo*), and *B. cellulosilyticus* (*Bc*) type strains harboring either p*Bolux* (black) or a plasmid containing the corresponding *rpoD* promoter (red). N=8, error is SEM. *P*-values were calculated using 2-way ANOVA with Fisher's LSD test. *** indicates values < 0.001.




**Extended Data Fig. 7. PUL reporters demonstrate a unique assemblage of distinct yeast glycans. a,** Alpha-fold structure of Sensor$^{PUL71}$ (BT3957) with labels indicating the periplasmic (PP) carbohydrate-binding module (CBM), the transmembrane (TM) region in the inner membrane (IM), and the cytoplasmic (CP) receiver (Rec) and helix-turn-helix (HTH) domains. **b,** Normalized AUC from *wild-type Bt* or the indicated sensor-deficient strains harboring P-*PUL36*, P-*PUL56*, P-*PUL69*, or P-*PUL71* supplied YEG. **c&e,** Normalized AUC from YEG-activated reporters supplied co-purified material following incubation of YEG with recombinant (**c**) SGBP$^{PUL71}$ or (**e**) SGBP$^{PUL56}$. N=6, error is SEM. **d&g,** Growth of *wild-type Bt* or the indicated sensor-deficient strains supplied co-purified material following incubation of YEG with recombinant (**d**) SGBP$^{PUL71}$ or (**g**) SGBP$^{PUL56}$. N=3, error is SEM. **f,** Growth of *wild-type Bt* or *Δsensor$^{PUL56}$* supplied β-gentibiose. N=8, error is SEM. **h,** HPAEC-PAD analysis of YEG treated with GH20$^{PUL56}$. **i,** Normalized bioluminescence from a strain harboring P-*PUL71* supplied various di- and oligo-mannosaccharides or YEG. N=4, error is SEM. For panels **b**, **c**, **e**, and **i**, *P*-values were calculated using 2-way ANOVA with Fisher's LSD test. * indicates values < 0.05, ** < 0.01, *** < 0.001 and *ns* > 0.05.




**Extended Data Fig. 8. Monosaccharide composition analysis of SBGP-isolated glycans. a-b,** HPAEC-PAD analysis of monosaccharides released from glycans that co-purified with SGBP$^{PUL71}$ (**a**) before or (**b**) after YEG treatment with GH20$^{PUL56}$. **c,** HPAEC-PAD analysis of monosaccharides released from glycans that co-purified with SGBP$^{PUL56}$ following incubation with YEG.




**Extended Data Fig. 9. PUL71 GH specificity indicates novel glycan structure. a-f,** HPAEC-PAD analysis of reactions containing (**a**) α1,2-mannobiose alone (black) or incubated with BT3962 (blue), (**b**) α1,4-mannobiose alone (black) or incubated with BT3965 (pink), or (**c**) α1,2-mannobiose, (**d**) α1,3-mannobiose, (**e**) α 1,4-mannobiose, or (**f**) α1,6-mannobiose alone (black) or incubated with BT3963 (purple).