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 intestinal glycome contains carbohydrates that govern the composition and metabolism of the gut microbiome. Gut microbes differentially access intestinal glycans as a nutrient source thereby directing microbial abundance and metabolite synthesis. Thus, identifying microbial accessible carbohydrates (MACs) in the gut is critical to understanding microbial ecologies and developing novel prebiotics that manipulate gut microbial populations for beneficial outcomes. However, no methods exist to efficiently examine MACs in biologically-derived mixtures. Here, we present a high-throughput platform to examine MACs from plant, animal, and microbial sources by simultaneously measuring bioluminescence from a genome-wide library of engineered *Bacteroides thetaiotaomicron* strains that harness bacterial glycan detection machinery. We demonstrate that this platform exhibits specific and sensitive responses to glycan mixtures and characterize a previously unknown MAC present in *Saccharomyces cerevisiae*. Expanding this technology across gut *Bacteroides* species will yield a diverse and broadly applicable glycomics platform to identify novel 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 enzyme repertoires that
44 target specific glycosidic linkages[1, 8, 11, 12]. Characterizing MACs present in biologically-derived
45 mixtures inform how and why the gut microbiota changes due to host dietary choices, genetics,
46 or disease because chemically-distinct glycans support the expansion of defined microbial
47 subsets. Additionally, identifying MACs that can influence microbial abundance and metabolism
48 can promote engineered strain engraftment and gene expression to leverage the gut
49 microbiome as a therapeutic target. Glycan mixtures have traditionally been examined using
50 tandem chromatography and mass spectrometry, arrayed glycan binding aptamers, or lectins
51 that probe for a range of known carbohydrate structures[13-15]. Although these approaches can
52 characterize composite glycan content in various sample types, they are unable to directly
53 detect MACs.

54 Microbes access unique MACs by producing proteins that sequester, transport, degrade,
55 and detect target molecules in their environment. For example, members of the dominant
56 human intestinal phylum, *Bacteroidetes*, possess glycan sensors encoded within polysaccharide
57 utilization genetic loci (PULs) that direct rapid, dramatic, and predictable increases to
58 corresponding PUL gene transcripts following recognition of their target glycan epitopes[4, 6, 7, 16].
59 Ultimately, these sensors direct increased expression of surface glycan binding proteins
60 (SGBPs), SusCD transporters, and glycolytic enzymes for consumption of available MACs.
61 Different *Bacteroides* species encode vast and partially unique PUL repertoires, which endow
62 each with the ability to access various target glycans that establish distinct intra-intestinal
63 niches. We previously demonstrated that introduction of PUL promoters into a plasmid-borne

64 *Bacteroides*-optimized *lux* cassette, p*Bolux,* harness their glycan sensor proteins to convert
65 these organisms into biosensors of known and unknown target glycans present in biologically-
66 derived mixtures[17]. We reasoned that PUL reporters could harness the exquisite sensitivity and
67 specificity of their cognate sensor proteins to detect MACs in glycan mixtures and be
68 implemented to isolate and characterize those molecules in combination with corresponding
69 PUL SGBPs and glycolytic enzymes.

70 Here, we employ *Bacteroides* PUL sensors to generate a genome-wide *Bacteroides*
71 *thetaiotaomicron* (*Bt*) reporter library. We arrayed these PUL reporters into a scalable and
72 distributable high-throughput platform that sensitively indicates target glycans from biologically-
73 derived mixtures. We demonstrate that this approach distinguishes structurally distinct MACs
74 according to monosaccharide composition, glycosidic linkages, and in some instances
75 polysaccharide degree of polymerization. We use this technology to detect previously unknown
76 PUL target glycans and employ a recombinant SGBP to isolate and characterize an unknown
77 MAC in Baker's yeast (*Saccharomyces cerevisiae).* Collectively, these results establish
78 *Bacteroides* species as specific and sensitive toolkits to directly detect, isolate, and characterize
79 unknown PUL target glycans. We assert that expanding these tools into additional *Bacteroides*
80 species will support broad, untargeted glycomics applications to examine heterogenous glycan
81 mixtures and overcome current detection limitations. Ultimately, our platform can be used to
82 develop new prebiotics that manipulate gut microbial abundance and metabolism to benefit
83 human health.
84

85 **Results**

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

87 PUL sensors specifically direct target PUL gene transcription in engineered strains
88 expressing constitutively active sensor proteins from PULs 16, 19, 49, 53, and 57, resulting in
89 dramatic increases in corresponding transcript abundances (Fig. 1a)[18, 19]. These results

90 demonstrate that a single PUL sensor directs increases in the predicted PUL transcripts[20, 21]. To
91 examine whether sensor-directed PUL transcription can be faithfully reported by corresponding
92 transcriptional reporters on a genome-wide scale, we generated a complete library of *Bt* PUL
93 reporter plasmids by introducing putative promoters identified from RNAseq datasets[22] and
94 mapped transcription start sites[23, 24] from annotated *Bt* PULs[25, 26] into p*Bolux*[17]. Although *Bt*
95 possesses 88 PULs, we generated 91 total reporter strains because PULs 14, 52, 74, and 75
96 possess two divergently transcribed *susCD* pairs were assigned unique reporters (denoted a
97 and b) and PUL63 was omitted due to the absence of a *susD*-like gene (Fig. 1a). Following
98 cryopreservation of batch-produced 96-well plates, individual replicates were revived by the
99 addition of rich media and incubated overnight. Each strain was transferred to a 384-well plate
100 and supplied 4 distinct carbohydrate preparations, one of which was always galactose as a
101 negative control (Fig. 1b). Fold changes in bioluminescence generated over 18 hours (Fig. 1c)
102 were calculated from the area under each curve from strains supplied unique glycan
103 preparations and normalized to a strain harboring the promoter-less p*Bolux* plasmid and
104 cultures supplied galactose alone (Fig. 1d).
105
106 ***Bt* PUL reporters distinguish between glycosidic linkages and monosaccharide**
107 **composition.**
108 To examine the *Bt* PUL reporter array functionality, we examined changes in
109 bioluminescence from all strains following the introduction of various purified mono- (Extended
110 Data Fig. 1a) and poly-saccharide preparations (Fig. 1e). As expected, the monosaccharides,
111 fructose and ribose, increased bioluminescence from their respective corresponding PUL
112 reporters, P-*PUL22* and P-*PUL37*, which agree with previous findings[27]. However, ribose also
113 increased *PUL22* transcripts[27] and corresponding bioluminescence from P-*PUL22* (Extended
114 Data Fig. 1a, b). Ribose is likely contaminated with fructose because a *Bt* strain lacking
115 *sensor*^{PUL22} (*BT1754*) harboring P-*PUL22* exhibited no bioluminescence increases when

116 supplied ribose, whereas this strain harboring P-*PUL37* exhibited similar responses to *wild-type*
117 *Bt* when supplied ribose (Extended Data Fig. 1c). In contrast, a strain lacking *sensor*^{PUL37}
118 (*BT2802*), which is unable to increase *PUL37* transcripts or utilize ribose as a sole carbon
119 source[27], exhibited similar bioluminescence when harboring either P-*PUL37* or p*Bolux*
120 (Extended Data Fig. 1c). These results indicate that the Sensor^{PUL22}, which binds fructose with
121 high affinity[7], is responsible for increasing PUL22 expression when supplied ribose. P-*PUL22*
122 also exhibited increased bioluminescence when supplied mannose (Extended Data Fig. 1a),
123 however, treatment with fructose-eliminating enzymes reduced P-*PUL22* activity (Extended
124 Data Fig. 1d), suggesting that the mannose preparation examined here was also contaminated
125 with fructose. Ultimately, these data highlight the exquisite specificity of PUL reporters for their
126 target ligands and detection of non-target glycans reflect their extraordinary sensitivity even in
127 mixtures.

128 We established that *Bt* PUL reporters can distinguish between MACs based on glycosidic
129 linkages, monosaccharide composition, and degree of polymerization (d.o.p.). Distinct strains
130 exhibited increased bioluminescence according to the administration of various purified plant,
131 animal, and microbially-derived MACs consumed via known PULs[4, 6, 7, 11, 27, 28]. Strikingly, P-
132 *PUL66*, P-*PUL48*, and P-*PUL56* were specifically activated by amylose, dextran, and pustulan,
133 respectively, which are distinct glucose homopolymers differing by their α1,4-, α1,6-, and β1,6-
134 glycosidic linkages (Fig. 1e). Each PUL requires respective linked SusR-like sensor proteins,
135 encoded by BT3703 (Sensor^{PUL66}), BT3090 (Sensor^{PUL48}), and BT3309 (Sensor^{PUL56}), to
136 increase target PUL transcription and utilize these MACs as growth substrates. Sensor^{PUL66} and
137 Sensor^{PUL56} differentiate between glycosidic linkage because their corresponding PUL reporters
138 only exhibit increase bioluminescence when supplied α1,4- or β1,6- linked glucose
139 disaccharides (Fig. 2a). However, P-*PUL48* can also distinguish between d.o.p. because this
140 strain only exhibited increased bioluminescence when supplied isomaltooligosaccharides
141 greater than 3 α1,6-linked glucose units and requires at least 6 residues to reach activity similar

142 to dextran (Fig. 2b). These data demonstrate that SusR-like PUL reporters can detect distinct
143 glycosidic linkages and d.o.p.
144 Further examination of purified oligosaccharides yielded several unexpected outcomes.
145 First, monomeric glucose and various glucose-disaccharides, including maltose, isomaltose,
146 trehalose, and β-gentibiose increased bioluminescence from P-*PUL58* (Fig. 2a). Similarly,
147 supplying glucose-fructose heterodimers, turanose and palatinose, stimulated greater than 10-
148 fold increases from this strain, whereas sucrose did not, suggesting that this PUL reporter
149 responds to glucose- and fructose- containing disaccharides subsets. We hypothesized that
150 *PUL58* could be governed by the SusR-like sensor, encoded by *BT2160*, that is an important *Bt*
151 fitness determinant during growth in trehalose and palatinose[29]. Accordingly, $\Delta sensor^{PUL58}$
152 harboring P-*PUL58* did not exhibit increased bioluminescence in any condition, indicating that
153 this sensor facilitates PUL58 transcription in response to various disaccharides (Fig. 2c).
154 Curiously, this sensor also responds to glucosinolates derived from cruciferous vegetables and
155 is required for their conversion into isothiocyanates by increasing transcription of the linked
156 genes *BT2159-56*[30]. A deeper examination of how and why this distally encoded sensor protein
157 responds to these signals is necessary to understand this regulatory relationship. Unexpectedly,
158 P-*PUL46* detected a *Bt*-inaccessible carbohydrate because introduction of the β1,4-linked
159 glucose disaccharide, cellobiose, increased bioluminescence from this strain (Fig. 2d). This
160 response was d.o.p.-specific because there was no change when supplied cellulose-derived
161 tetra- or hexa-saccharides (Fig. 2d). Because *Bt* cannot utilize cellulose or its breakdown
162 products (PBP) including cellobiose[31], these results suggest that cellulose-derived PBPs serve
163 as signals in the intestine.
164
165 ***Bt* PUL reporters detect discreet carbohydrate structures**
166 The *Bt* PUL array can also distinguish between various microbial-, animal-, and plant
167 glycans covalently tethered into polymeric complexes. For example, *Bt* utilization of yeast cell

168 wall mannans (Extended Data Fig. 2a) requires three PULs: 36, 68, and 69[32, 33], which increase
169 activity from the reporter strains, P-*PUL36*, P-*PUL68,* and P-*PUL69* by the addition of purified α-
170 mannan (Fig. 1e). The corresponding sensor proteins, BT2628 ($Sensor^{PUL36}$) and BT3786
171 ($Sensor^{PUL68}$) facilitate mannan-responsive PUL36/68 transcription because bioluminescence
172 from strains harboring either P-*PUL36* or P-*PUL68* is not apparent in a strain lacking both
173 sensors (*Δsensors*$^{PUL36/68}$; Extended Data Fig. 2b), consistent with previous transcript
174 measurements[32]. In contrast, PUL69 transcription is facilitated by $Sensor^{PUL69}$, encoded by
175 *BT3853*[32], which is required for increased bioluminescence from P-*PUL69* when supplied α-
176 mannan (Extended Data Fig. 2b). $Sensors^{PUL36/68/69}$ are collectively required for mannan
177 utilization as a sole carbon source (Extended Data Fig. 2c) but detect discreet carbohydrate
178 structures because the branched α1,3-[α1,6] trisaccharides increases P-*PUL36/68*
179 bioluminescence but not from P-*PUL69* (Extended Data Fig. 2d, e). Conversely, P-*PUL69* did
180 not exhibit increased bioluminescence in response to any di- or tri-saccharide examined,
181 suggesting it responds to a different oligo-mannan structure (Extended Data Fig. 2f). These data
182 highlight that PUL sensor proteins direct cognate PUL expression following recognition of
183 distinct glycan moieties assembled as larger polymers.
184 *Bt* possess 3 PULs that respond to glycosaminoglycans, which are intestinally abundant
185 heteropolymers comprised of amine sugars and uronic acids, including previously characterized
186 PULs 57[6, 10] and 85[34], and the uncharacterized PUL20 (Fig. 3a). Strains harboring P-*PUL57*
187 exhibited increased bioluminescence when supplied chondroitin sulfate (CS) or dermatan
188 sulfate (DS), which are variably sulfated linear polymers of repeating, disaccharide units
189 comprised of glucuronic acid (gluA) and β1,3-linked N-acetylgalactosamine acid (GalNAc). P-
190 *PUL57* also responds to hyaluronic acid[17], an unsulfated GAG whose disaccharide unit is
191 comprised of gluA and β1,3-linked N-acetylglucosamine (GlcNAc) because the $Sensor^{PUL57}$
192 protein, BT3334, cannot distinguish between either CS-, DS- or HA-derived disaccharide
193 subunits (Fig. 3b)[6]. Introduction of CS, DS, and HA also increased bioluminescence from P-

194 *PUL20* (Fig. 1e), consistent with previous transcriptional measurements[3], suggesting that both
195 GAGs may also activate an additional PUL. However, Sensor^{PUL57} is not responsible for
196 increasing bioluminescence from P-*PUL20* because: 1.) CS- a HA-derived disaccharides
197 increased activity from P-*PUL57* but not P-*PUL20* (Fig. 3b), 2.) heterologous expression of a
198 constitutively active Sensor^{PUL57} (BT3334*) increased *PUL57* but not *PUL20* transcripts (Fig.
199 1a), and 3.) BT3334* increased bioluminescence from strains harboring P-*PUL57* but not P-
200 *PUL20* (Fig. 3c). Furthermore, a distinct GAG, heparan sulfate (HS), also increased
201 bioluminescence from P-*PUL20* and P-*PUL85* (Fig. 1e), whereas the composite disaccharide
202 comprised of alternating N-acetylglucosamine (GlcNAc) and uronic acid increased activity from
203 P-*PUL85* but not P-*PUL20* (Fig. 3b). We propose that PUL20 targets the common GAG
204 tetrasaccharide linker, GlcA-[β1-3]-Gal-[β1-3]-Galβ-[1-4]-Xyl2P-β-Ser (Fig. 3d) because CS from
205 shark cartilage stimulates increased bioluminescence from P-*PUL20* and P-*PUL57* but no other
206 contaminating glycans following co-incubation with SGBP^{PUL57} (Fig. 3e)[17], indicating that the
207 PUL20-target glycan is covalently tethered to CS.

208 Plant cell wall glycans called pectins and hemicelluloses, are frequently assembled into
209 complex polymeric structures (Extended Data Fig. 3a) that are targeted by distinct *Bt* PULs[1, 4, 9,
210 11]. This is exemplified by P-*PUL7*, which exhibited 68-fold increased bioluminescence when
211 supplied arabinan (Fig. 1e). However, arabinan also elicited increased bioluminescence from
212 strains harboring P-*PUL5*, P-*PUL13*, P-*PUL22*, P-*PUL23,* P-*PUL47*, P-*PUL48*, P-*PUL65*, P-
213 *PUL77*, or P-*PUL86*, indicating this preparation also contains arabinogalactan (PUL5), and II
214 (PULs 13 & 21), fructose (PUL22), homogalacturonan (PUL75), and rhamnogalacturonan I
215 (PUL77). Reporter activity faithfully reflects sensor-dependent activity because the Sensor^{PUL7}
216 (BT0366)[35], which recognizes arabino-oligomers of at least 6 α1,5-linked arabinose residues[4],
217 was only required to stimulate P-*PUL7* when supplied arabinan (Extended Data Fig. 3b).
218 Consistent with this notion, purified polygalacturonic acid (PGA) oligosaccharides increased
219 bioluminescence from its corresponding reporter, P-*PUL75a&b* (Fig. 2a), whereas highly

220 polymeric PGA also increased activity from P-*PUL7*, P-*PUL13*, P-*PUL22*, P-*PUL48*, and P-
221 *PUL66*, and P-*PUL86* (Fig. 1e), demonstrating that PUL reporters respond predictably to
222 complex MACs comprised of multiple ligands[1, 4, 7, 9, 11]. Finally, the ubiquitous food additives
223 carrageenan, xantham gum, locust bean gum, karaya gum, and gum arabic are complex
224 heteropolymers of largely *Bt*-inaccessible carbohydrates[31] but contain small amounts of
225 accessible structures because each elicits increased bioluminescence from at least one PUL
226 reporter (Fig. 1e). For example, P-*PUL5* is dramatically activated by purified arabinogalactan
227 and exhibits significant increases when supplied karaya gum and gum arabic (Fig. 1e).
228 Collectively, these data demonstrate the exquisite specificity of *Bt* PUL reporters for complex
229 multi-valent MACs, facilitated by the cognate sensor proteins.
230
231 **The Bt PUL reporter array can faithfully survey glycan mixtures.**
232 *Bt* PUL reporters can simultaneously examine MACs present in heterogeneous glycan
233 mixtures because P-*PUL5*, P-*PUL22*, P-*PUL57*, and P-*PUL48* exhibited increased
234 bioluminescence when supplied an equal mixture of purified arabinogalactan (PUL5), levan
235 (PUL22), hyaluronan (PUL57), and dextran (PUL48) (Fig. 1e). Additionally, this mixture
236 increased activity from P-*PUL20* and P-*PUL47*, which were both increased by HA (Fig. 1c-e),
237 displaying faithful reporting of lab-prepared mixtures. To characterize biologically derived
238 mixtures, we measured reporter responses to PMOGs, a mixture of porcine mucosal-derived O-
239 glycans that increases many *Bt* PUL transcripts[3, 17, 36]. When supplied PMOG, we detected
240 increased activity in GAG-, and O-glycan-specific PUL reporters, consistent with previous
241 transcript measurements[3, 17, 36] indicating that the reporter strains can faithfully examine
242 biologically-derived mixtures (Fig. 4a). To prepare other host-, diet-, and microbially derived
243 mixtures, we adapted the oxidative-release of natural glycans (ORNG) method that can liberate
244 glycans from a wide variety of biological materials[37]. We determined that a preparation of
245 porcine mucosal glycans using ORNG (PMG), yielded similar reporter profiles as PMOG (Fig.

246 4a), but did display a 4-fold decrease in bioluminescence from a strain harboring P-*PUL15* and
247 a 5-fold increase in signal from P-*PUL34*, indicating glycan subsets are differentially extracted or
248 detected when this approach is applied (Extended Data Fig. 4a).

249 To determine how gut microbial responses differ between animal and human mucosal
250 material, we compared human-derived intestinal glycan mixtures collected from various patient
251 intestinal samples recovered following surgery (HIG) and porcine cecal mixtures from freshly
252 slaughtered boars (BCG). Human mucosal glycans elicited similar patterns of reporter activity
253 as PMOG and fresh porcine cecal glycans, suggesting that mammalian hosts largely supply gut
254 microbes with similar growth substrates (Fig. 4a). For example, HIG and BCG elicited similar
255 bioluminescence from P-*PUL14*a&b, P-*PUL20*, P-*PUL38*, P-*PUL57*, P-*PUL66*, and P-*PUL72*
256 (Extended Data Fig. 4b). However, the following distinctions were readily apparent: on one
257 hand, human material elicited reduced bioluminescence from P-*PUL31*, P-*PUL52b*, and P-
258 *PUL85*, and on the other hand they elicited increased responses from P-*PUL75b*, P-*PUL79, P-*
259 *PUL80, and P-PUL81* (Extended Data Fig. 4b). Furthermore, human milk glycans (HMG)
260 elicited similar responses to HIG, however, signals were improved when HMG were prepared by
261 defatting and proteinase K treatment (Extended Data Fig. 4c), suggesting that ORNG can result
262 in glycan peeling from some source materials[38].

263 We used ORNG to prepare various animal, plant, and fungal material that are abundant
264 components of the human diet (Fig. 4b) and examined *Bt* reporter responses. As expected,
265 plant-, animal-, and microbially derived material occupied distinct space following PCoA analysis
266 indicating that MAC distribution reflects source phylogeny at the phylum level (Fig. 4c). For
267 example, dietary plant products emanating from both mono- and di-cot species distinctly
268 increased bioluminescence from strains harboring P-*PUL5* (arabinogalactan), P-*PUL7*
269 (arabinan), P-*PUL13* and P-*PUL21* (RGII), P-*PUL22* (fructan), P-*PUL66* (starch), P-*PUL75*
270 (PGA), and P-*PUL86* (RGI), which is consistent with common plant fibers. Conversely,
271 mammalian derived mixtures stimulated P-*PUL14b* (N-glycans), P-*PUL20* (GAG associated), P-

272 *PUL38* (unknown mucin *O*-glycan), P-*PUL57* (CS/HA), P-*PUL66* (glycogen), and P-*PUL76*
273 (unknown mucin *O*-glycan). Finally, fungal products consistently increased activity from strains
274 containing P-*PUL29* (unknown), P-*PUL36/68* (α-mannan), P-*PUL56* (β-glucan), P-*PUL66*
275 (glycogen), and P-*PUL69* (α-mannan). Accordingly, processed foods containing both plant and
276 fungal products, including breads and beer, contained reporter increases consistent with both
277 grains and yeast (Fig. 4a).

278 Unexpectedly, several PUL reporters did not exhibit bioluminescence increases when
279 supplied any purified carbohydrate or mixture examined in this study, suggesting that either
280 these target glycans are not present in the mixtures prepared for this work or the corresponding
281 reporter plasmid does not contain the appropriate regulatory sequence to facilitate increased
282 bioluminescence. We determined that the former is the most likely scenario because qPCR
283 analysis for *susC* transcripts from PULs whose corresponding reporters failed to increase
284 bioluminescence were not increased when supplied various mixtures (Extended Data Fig. 5a).
285 Furthermore, expression of constitutively active sensor proteins from PULs 16, 19, 49, 53, and
286 57, which increased target PUL transcription (Fig. 1a)[18, 19], stimulated bioluminescence
287 increases when these strains contained the corresponding PUL reporter (Extended Data Fig.
288 5b). A notable exception is PUL2, which targets an unknown glycan exhibits 180-fold increased
289 transcription in yeast extract and red-wine derived glycan mixtures (Extended Data Fig. 5a), but
290 whose corresponding reporter did not (Fig. 4a). We engineered an additional reporter that
291 includes an internal transcription start site that exhibited a significantly increased
292 bioluminescence when supplied red wine and yeast glycans, however, these responses were
293 not reflective of the fold increases in PUL transcripts (Extended Data Fig. 5a). Thus, the majority
294 of PULs can be converted into bioluminescent glycan biosensors with a small minority that are
295 unable to be adapted using p*Bolux*. Finally, establishing the functionality of various PUL
296 reporters with unknown target glycans (Extended Data Fig. 5b) that are not activated across the

297 common foodstuffs examined in this study, indicate that future investigations should focus on
298 other glycan sources including co-resident microbes and fermented products.
299
300 **PUL reporter construction across species**
301 This study demonstrates how *Bt* PUL reporters can be readily constructed and deployed
302 to detect MAC content in various glycan preparations. However, other *Bacteroides* species
303 possess unique PUL repertoires that putatively confer distinct glycan accessibility, indicating
304 they possess PUL sensors that recognize a partially unique subset of carbohydrate structures
305 (Fig. 4c). For example, *B. fragilis* possesses PULs that facilitate its growth in host-derived
306 glycan preparations including those distinct from *Bt* target glycans[19], whereas *B. ovatus*[39] and *B.*
307 *uniformis*[40] possess PUL repertoires that target *Bt*-inaccessible hemicelluloses. We previously
308 demonstrated that p*Bolux* can report promoter-dependent transcription in *B. ovatus*[17] and *B.*
309 *fragilis*[41], and this is consistent in *B. uniformis* and *B. cellulosilyticus* because introduction of
310 their corresponding *rpoD* promoters in this plasmid increase bioluminescence from all five
311 species (Extended Data Fig. 6). Therefore, construction of future PUL reporter libraries in these
312 species will expand the breadth of glycan detection capabilities to comprehensively characterize
313 MAC content in host- and dietary- components.
314
315 **Isolation and characterization of a *Saccharomyces cerevisiae* glycan consumed by *Bt.***
316 To explore the utility of PULs as glycomics tools, we characterized an unknown glycan
317 detected in a biologically derived mixture. We selected PUL71 (Fig. 5a) because yeast extract
318 glycans (YEG) elicited 32-fold increased bioluminescence from strains harboring P-*PUL71*,
319 which was also stimulated by preparations from yeast containing products like red wine and
320 animal feeds but not breads or beers (Fig. 4a). Furthermore, P-*PUL71* activity was increased in
321 yeast-derived mannan prebiotic mannooligosaccharides (MOS) but absent in host- and most
322 diet-derived mixtures, although increased in grape juice possibly from yeast contamination[42]

323 (Fig. 4a). Furthermore, this glycan is an excellent prebiotic candidate because a strain lacking
324 PUL71 (*ΔBT3957-BT3965*) exhibited a 5-fold fitness defect relative to *wild-type Bt* in mice
325 supplied a 0.5% YEG solution rather than water (Fig. 5b). Increased P-*PUL71* activity requires a
326 unique glycan sensor protein class, encoded by *BT3957*, that possesses a periplasmic glycan
327 binding domain and a cytoplasmic DNA-binding domain similar to HTCSs but lacks conserved
328 histidine kinase and response regulator receiver domains (Extended Data Fig. 7a). We
329 established that the putative sensor, BT3957, controls PUL71 transcription because a
330 *sensor*^{PUL71}-deficient strain (*Δsensor*^{PUL71}) harboring P-*PUL71* failed to exhibit increased activity
331 in response to YEG (Extended Data Fig. 7b). Conversely, *sensor*$^{PUL36/68}$- or *sensor*^{PUL69}-deficient
332 strains harboring P-*PUL71* exhibited similar bioluminescence as *wild-type Bt* supplied YEG
333 (Extended Data Fig. 7b). PUL71 responds to a glycan distinct from other well characterized
334 yeast-derived glycans including α-mannan, which activates P-*PUL36*/*68/69* (Extended Data Fig.
335 2b)[32] and β-1,6-glucan, which stimulates P-*PUL56* (Fig. 2a)[43], because bioluminescence from
336 P-*PUL71* was unchanged when supplied these purified glycans (Fig. 1e) and *Δsensor*^{PUL71}
337 exhibited similar activity to *wild-type Bt* harboring P-*PUL36* or P-*PUL56* when supplied YEG
338 (Extended Data Fig. 7b). Collectively, these data indicate that Sensor^{PUL71} recognizes a unique
339 yeast-derived glycan that increases *PUL71* expression.
340 To characterize the PUL71 target glycan, we co-incubated YEG with recombinant
341 SGBP^{PUL71} (BT3961), and recovered protein-bound glycans using metal affinity
342 chromatography[17]. The recovered glycans were enriched for the PUL71 target over other MACs
343 present in YEG because the co-purifying material significantly increased activity from P-*PUL71*
344 but P-*PUL7* (α-arabinan), P-*PUL22* (fructans), P-*PUL29* (unknown), P-*PUL48* (dextran), P-
345 *PUL66*, and P-*PUL72* (HMNG) were no longer detectable (Extended Data Fig. 7c). However, P-
346 *PUL36/68/69* and P-*PUL56* also exhibited increased activity indicating that α-mannan and β-
347 glucan co-purified with the PUL71 target (Extended Data Fig. 7c). Significantly activated PULs
348 contributed to accessing the SGBP^{PUL71} co-purifying material because the corresponding

349 sensor-deficient strains exhibited reduced growth on this material as a sole carbon source, with
350 *Δsensor*^{PUL71} exhibiting the largest reduction (Extended Data Fig. 7d). We hypothesized that the
351 BT3961-purified material contained covalently tethered carbohydrates comprised of the PUL71-,
352 PUL36/68/69-, and PUL56- target glycans because the yeast cell-wall is an interconnected
353 matrix comprised of chitin, β1,3-, and β1,6-glucans decorated with mannoproteins through a
354 GPI-anchor remnant glycan (Fig. 5d). Consistent with this notion, material co-purifying with the
355 SGBP^{PUL56}, BT3313, incubated with YEG, also increased bioluminescence from strains
356 harboring P-*PUL56*, P-*PUL36/68*, and P-*PUL71*, although bioluminescence from P-*PUL69* was
357 not significantly increased possibly because the corresponding ligand was below the limit of
358 detection (Extended Data Fig. 7e). Accordingly, only a *Δsensor*^{PUL56} exhibited substantially
359 reduced growth when supplied SGBP^{PUL56}-isolated material, indicating the other target glycans
360 are present in relatively low abundances (Extended Data Fig. 7f). Collectively, these data
361 demonstrate that the PUL56 and PUL71 SGBPs isolate their corresponding target glycans
362 assembled into interconnected polymeric structures.
363 To determine how the PUL71-target glycan is tethered to β1,6-glucan and α-mannan, we
364 treated YEG with the PUL56-encoded β1,6-glucosidase (BT3312), referred to here as
365 GH20^{PUL56}, which liberated glucose and β-gentibiose from YEG (Extended Data Fig. 7i). We
366 determined that GH^{PUL56} treatment removed β1,6-glucan but not α-mannan from the PUL71
367 target because SGBP^{PUL71}-isolated material no longer increased bioluminescence from P-
368 *PUL56* but exhibited significant increases from P-*PUL36/68/69* (Fig. 5c). Accordingly,
369 *Δsensor*^{PUL71} and *Δsensors*$^{PUL36/68/69}$ each exhibited reduced growth on this material as a sole
370 carbon source and growth was completely abolished in a strain lacking all four sensors
371 (*Δsensors*$^{PUL36/68/69/71}$) (Fig. 5e). Furthermore, compositional analysis demonstrated that
372 SGBP^{PUL71}-isolated material from untreated YEG contained glucose (Extended Data Fig. 8a),
373 whereas this monosaccharide was no longer detectable when YEG was pre-treated with
374 GH20^{PUL56} (Extended Data Fig. 8b). Consistent with this finding, HPAEC-PAD analysis revealed

375 that SGBP^{PUL56}-isolated glycans from untreated YEG contained mannose (Extended Data Fig.
376 8c), further suggesting that yeast β1,6-glucan and α-mannan are tethered in the absence of
377 mannoproteins, which are eliminated by ORNG extraction. These data indicate that yeast α-
378 mannan, β-glucan, and PUL71-target glycan are covalently assembled into a complex resistant
379 to oxidation, and that the PUL71 target glycan decorates yeast-derived α-mannan prepared
380 using ORNG (Fig. 4a, Fig. 5c) but not alkaline extraction (Fig. 1e). PUL71 includes the
381 mannosidases BT3962 and BT3965, which respectively convert α1,2- and α1,4-mannobioses
382 into mannose (Extended Data Fig. 9a, b), and a putative mannosidase BT3963, which was
383 unable to hydrolyze any mannobiose examined here (Extended Data Fig. 9c-e), in agreement
384 with previous results[44]. Collectively, these data suggest that the PUL71-target glycan is an
385 unknown yeast α-mannan decoration comprised of unique glycosidic linkages that could be a
386 useful prebiotic.
387
388 **Discussion**
389 We have developed a microbial glycomics platform that harnesses *Bacteroides* PUL sensors
390 to detect, isolate, and characterize MACs from biologically derived mixtures (Fig. 1). We
391 demonstrated that a genome-wide *Bt* PUL-reporter library can indicate the monosaccharide
392 composition, glycosidic linkages and d.o.p. of purified MACs (Fig. 1e ,2b). Furthermore, we
393 demonstrated this toolkit can efficiently characterize MAC content in semi-purified preparations
394 (Fig. 1e, Extended Data Figs. 1, 2) and various mixtures (Fig. 1e and Fig. 4a) derived from the
395 host mucosa and dietary material (Fig. 4b). We establish that reporter responses largely
396 correspond to the biological source of each mixture (Fig. 4c), indicating that MAC content
397 reflects phylogeny. The exquisite specificity and sensitivity of *Bacteroides* reporter strains
398 highlight their utility as glycan biosensors that could be implemented in product quality control to
399 detect carbohydrate contaminants (Extended Data Fig. 1d) and co-purifying moieties (Fig. 3).
400 Furthermore, this approach could be applied to broader glycomics applications including

401 surveillance of different carbohydrate structures from patient samples, plant cultivars, and
402 microbial isolates (Fig. 4a). Finally, *Bt* PUL reporters can directly identify novel MACs that could
403 be translated into prebiotics to manipulate the gut microbiome composition and products for
404 therapeutic purposes (Fig. 5). Therefore, PUL reporters offer a powerful new toolkit that can be
405 expanded using additional *Bacteroides* species that possess unique PUL repertoires (Extended
406 Data Fig. 9).
407
408 **Materials and Methods**
409 **Bacterial growth conditions**
410 All bacteria were cultured as described previously[45]. Briefly, *Bacteroides* strains were
411 cultured on brain-heart infusion agar (MilliporeSigma) containing 5% horse blood (Hardy) under
412 anaerobic conditions (85% N_2, 12.5% CO_2, 2.5% H_2). Liquid cultures were inoculated from a
413 single colony into TYG media and incubated under anaerobic conditions. All bacterial strains
414 included the following antibiotics where appropriate: 100 µg/mL ampicillin, 200 µg/mL
415 gentamicin, 2 µg/mL tetracycline, or 25 µg/mL erythromycin (MilliporeSigma).
416
417 ***Reporter construction***
418 Reporter plasmids (Supplementary Table 1) were constructed by amplifying PUL
419 promoter fragments using oligonucleotides (Supplementary Table 3) as previously described[17].
420 Briefly, PUL promoters amplified using Q5 Master Mix (NEB) were combined with pBolux[17]
421 linearized with BamHI and SpeI (NEB) and NEBuilder Master Mix (NEB). Reactions were
422 incubated for 1 hour at 50°C before dialyzed on a 0.025 µM dialysis membrane against water
423 and subsequent electroporation into competent *E. coli* S17-1. All plasmids were verified by
424 Sanger sequencing.
425
426 ***Engineering Bt mutants***

427 Indicated *Bt* genomic deletions were generated using pEXCHANGE-*tdk* plasmids
428 (Supplementary Table 2) harboring flanking sequences, using amplicon-specific primers
429 (Supplementary Table 3), as previously described[46]. In short, pEXCHANGE constructs were
430 introduced via di-parental mating and chromosomal integration was validated by PCR. Parent
431 strains were counter-selected on solid media containing 200ng/mL 5-fluoro-2-deoxyuridine
432 (DOT Scientific). All strains were validated using linear amplicon sequencing.
433
434 ***Bt reporter array production***
435 All 91 reporter *Bt* strains described were cultured in TYG to stationary phase in a distinct
436 position of a 96 deep-well plate (Supplementary Table 1). Each strain was combined with
437 glycerol to a 10% final concentration, and 5 µL of the resulting mixture was dispensed into a
438 sterile 96-well plate using a Mini-96 (Integra), covered with adhesive sterile foil, and placed on
439 dry ice before ultimate storage at -80°C.
440
441 ***Bt reporter assays***
442 One replicate array plate was thawed and combined with pre-reduced TYG and cultured
443 overnight at 37°C under anaerobic conditions. The following day, each strain was diluted 50-fold
444 into minimal media containing 0.5% galactose and cultured to mid-logarithmic phase before
445 centrifugation at 2,204 x *g* to pellet cells, resuspension into 2x minimal media containing 0.2%
446 galactose and distribution into a sterile 384-well white, clear bottom microplate (Corning)
447 containing equal amounts of 0.8% carbon. Absorbance and bioluminescence were measured
448 using an Infinite M200 Plex (Tecan) every 15 minutes over 18 hours.
449
450 ***Biologically derived glycan mixture preparation***
451 Solid material was homogenized in 6% bleach (PureBright) and allowed to stir for 30
452 minutes at room temperature before addition of 1:100 volumes of formic acid were combined

453 and stirred for 5 additional minutes at room temperature. Reactions were centrifuged at 7200 x
454 g for 10 minutes and the resulting supernatant was combined with bleach to a final
455 concentration of 0.2% and stirred for 16 hours at room temperature before addition of equal
456 volumes of formic acid and stirring for an additional 5 minutes. Reactions were centrifuged at
457 7200 x g for 10 minutes and the resulting supernatant was sterilized using a 0.22 µm vacuum
458 filter (Millipore). The resulting material was dialyzed with ultrapure H_2O using 1KDa cutoff
459 dialysis membrane for 3 days before lyophilization.
460
461 ***Human tissue collection***
462 The Pennsylvania State University College of Medicine Institutional Review Board
463 approved this study (IRB Protocol No. HY98-057EP-A). Prior to colectomy, patients gave
464 informed consent to have surgically resected tissue collected and banked into the Carlino
465 Family Inflammatory Bowel and Colorectal Disease Biobank. Diagnoses were confirmed using
466 colonoscopy and biopsy. Surgical specimens were immediately transferred from the operating
467 room to the surgical pathology lab, where the tissue was examined by a pathologist and
468 surgeon. The samples were then transported to the research laboratory on ice for further
469 processing. Mucosa was scraped from approximately 1 inch tissue sections and stored at -80°C
470 until samples were weighed, pooled, and subjected to glycan extraction.
471
472 ***Protein expression and purification***
473 SGBPs and GHs were recombinantly expressed as previously described[45]. Briefly,
474 inserts encoding each enzyme were amplified from *B. thetaiotaomicron* VPI-5482 genomic DNA
475 and cloned into pT7-7-N6H4A[17] linearized with NotI-HF and HindIII-HF using NEBuilder Hi-Fi
476 Master Mix (NEB). The resulting plasmids were introduced into *E. coli* S17-1 and verified by
477 Sanger sequencing. Plasmids were introduced into BL21 and plated on selective media. A
478 single colony was inoculated into Luria Bertani broth (BD) and cultured to mid-logarithmic phase

479 before 50 μM IPTG addition and subsequent incubation for 4 hours at 30°C. Pelleted cells were
480 lysed in Lysis Buffer (20 mM Tris (pH, 8) and 100 mM sodium chloride) and N-terminally hexa-
481 histidine-tagged proteins were recovered following co-incubation with Ni^{2+}-NTA resin (Thermo)
482 and elution with Lysis Buffer containing 25 mM Histidine. Recovered proteins were buffer
483 exchanged in 10 mM Tris (pH, 7.4) containing 10% glycerol using appropriate molecular weight
484 cut-off centrifugal concentrators (MilliporeSigma). Protein concentrations were determined using
485 a BCA protein quantification kit (Thermo).
486
487 ***SGBP-mediated glycan isolation***
488 Glycans were recovered from mixtures using recombinant SGBPs as previously
489 described[17]. Briefly, an *Ec* strain BL21 cell lysate expressing N-terminal hexa-his-tagged SGBP
490 was combined with 0.5% glycan and 1.0 mL Ni^{2+}-NTA resin (Thermo) and incubated overnight
491 at 4°C with rocking. The mixture was packed into a 3 mL gravity flow column and washed with
492 20 mM Tris (pH, 8) and 100mM Sodium Chloride 8. Protein-glycan complexes were eluted with
493 25 mM histidine (MilliporeSigma), treated with proteinase K before incubation at 65°C for 2
494 hours. The reaction was combined with 3x volume of 100% ethanol incubated overnight at 4°C
495 with rocking and centrifuged for 15 minutes at 21,291 x *g*. The pelleted material was dried under
496 filtered air, resuspended in water and combined before lyophilization.
497
498 ***qPCR***
499 Stationary phase cultures were diluted 50-fold into pre-reduced minimal media containing the
500 0.5% galactose and incubated to mid-logarithmic phase (OD_{600} = 0.45 - 0.65). 1.0 mL of culture
501 was pelleted by centrifugation, immediately placed on dry ice, and stored at -80°C Remaining
502 cultures were pelleted by centrifugation at 7200 x g for 3 minutes and resuspended in pre-
503 reduced 2X minimal media containing no carbon. Equal amounts of resuspended culture and
504 carbon were combined for a final concentration of 0.5% carbon and incubated for 60-minutes

505 before 1.0 mL of culture was pelleted by centrifugation, immediately placed on dry ice, and
506 stored at -80°C. RNA was isolated using the RNeasy Mini Kit (QIAGEN), quantified by
507 absorbance, and 1 µg was converted to cDNA following addition of the Superscript IV VILO
508 Master Mix with ezDNAse (Invitrogen) per the manufacturer's directions. Transcript abundances
509 were measured as previously described[45] using a QuantStudio5 (ThermoFisher) and PowerUp
510 SYBR Green Master Mix (ThermoFisher) with amplicon specific primers (Supplementary Table
511 3). The 16S *rRNA, rrs,* was used as the reference gene for all qPCR experiments.
512
513 ***In vivo competitive fitness of B. thetaiotaomicron strains***
514 All animal experiments were performed in accordance with protocols approved by Penn State
515 Institutional Animal Care and Use Committee. Germ-free C57/BL6 mice were maintained in
516 flexible plastic gnotobiotic isolators with a 12-hour light/dark cycle and provided a standard,
517 autoclaved mouse chow (LabDiet, 5021) *ad libitum*[45]. Mice were randomly distributed into 2
518 groups, which were supplied either autoclaved drinking water or 0.5% yeast extract glycans, 2
519 days prior to gavage with 10^8 CFU of each indicated strains suspended in 200 µL of phosphate-
520 buffered saline. Input (day 0) abundance of each strain was determined by counting colony
521 forming units following plating on solid media. Fresh fecal pellets were collected at the indicated
522 days and genomic DNA was extracted as described previously[22]. The abundance of each strain
523 was measured by qPCR, using barcode-specific primers (Supplemental Table 2) as described
524 previously[45].
525
526 ***Monosaccharide Composition Analysis***
527 Trifluoroacetic acid was added to the indicated glycan preparations to a final concentration of
528 2N and heated to 100°C for 4 hours. Reactions were cooled to room temperature, dried under
529 filtered air, and washed twice with 50% isopropanol before being dried to completion. Reactions
530 were resuspended in 0.2 mL of deionized water and analyzed by HPAEC-PAD.

531

532 ***Software and statistics***

533 Data collection and curation was done in Microsoft Excel. All data was plotted in GraphPad
534 Prism, except for heatmaps and PCoA, which were generated using ggplot2 in RStudio.
535 Statistical analyses were calculated in Prism. For all experiments, N denotes individual
536 biological replicates across at least two independent experiments.

537

538 ***CAZyme activity***

539 YEG was combined with 200 nM *GH20*^{PUL56} to a final concentration of 1.0% in 20 mM Tris (pH,
540 8), incubated for 4 hours at 37 °C, and analyzed by HPAEC-PAD. Mannobioses were combined
541 with 200 nM of each PUL71 GH in 20 mM Tris (pH=8) and 1 μM calcium chloride.

542

543 **Supplementary Information**

544 **Supplementary Table 1. *Bt* PUL reporter plasmids**

545 **Supplementary Table 2. Strains and plasmids used in this study**

546 **Supplementary Table 3. Oligonucleotides used in this study**

547

548 **Acknowledgements**

549 We thank Walter Koltun, Leonard Harris, Troy Ott, Wesley Raup-Konsavage, and Hannah
550 Valensi for assistance producing glycan extracts. We thank Biswa Chourdry and Richard Helm
551 for assistance with carbohydrate compositional assistance. Additionally, we thank Jordan
552 Bisanz for assistance with animal experiments and useful discussions. This project is funded, in-
553 part by National Institutes of Health Grants GM147178 and DK132711 to G.E.T. and a grant
554 with the Pennsylvania Department of Health using Tobacco CURE Funds. The PA Department
555 of Health specifically disclaims responsibility for any analyses, interpretations or conclusions.
556 This work was supported by the Carlino fund to The Penn State Colorectal Disease Biobank.

557

558 **References**

559 1. Luis, A.S. et al. Dietary pectic glycans are degraded by coordinated enzyme pathways in
560 human colonic Bacteroides. *Nat Microbiol* **3**, 210-219 (2018).
561 2. Marcobal, A. et al. Bacteroides in the infant gut consume milk oligosaccharides via
562 mucus-utilization pathways. *Cell Host Microbe* **10**, 507-514 (2011).
563 3. Martens, E.C., Chiang, H.C. & Gordon, J.I. Mucosal glycan foraging enhances fitness
564 and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**,
565 447-457 (2008).
566 4. Martens, E.C. et al. Recognition and degradation of plant cell wall polysaccharides by
567 two human gut symbionts. *PLoS Biol* **9**, e1001221 (2011).
568 5. Salyers, A.A., Vercellotti, J.R., West, S.E. & Wilkins, T.D. Fermentation of mucin and
569 plant polysaccharides by strains of Bacteroides from the human colon. *Appl Environ*
570 *Microbiol* **33**, 319-322 (1977).
571 6. Raghavan, V., Lowe, E.C., Townsend, G.E., 2nd, Bolam, D.N. & Groisman, E.A. Tuning
572 transcription of nutrient utilization genes to catabolic rate promotes growth in a gut
573 bacterium. *Mol Microbiol* **93**, 1010-1025 (2014).
574 7. Sonnenburg, E.D. et al. Specificity of polysaccharide use in intestinal bacteroides
575 species determines diet-induced microbiota alterations. *Cell* **141**, 1241-1252 (2010).
576 8. Bakshani, C.R. et al. Carbohydrate-active enzymes from Akkermansia muciniphila break
577 down mucin O-glycans to completion. *Nat Microbiol* **10**, 585-598 (2025).
578 9. Cartmell, A. et al. A surface endogalactanase in Bacteroides thetaiotaomicron confers
579 keystone status for arabinogalactan degradation. *Nat Microbiol* **3**, 1314-1326 (2018).
580 10. Ndeh, D. et al. Metabolism of multiple glycosaminoglycans by Bacteroides
581 thetaiotaomicron is orchestrated by a versatile core genetic locus. *Nat Commun* **11**, 646
582 (2020).
583 11. Ndeh, D. et al. Complex pectin metabolism by gut bacteria reveals novel catalytic
584 functions. *Nature* **544**, 65-70 (2017).
585 12. McNulty, N.P. et al. Effects of diet on resource utilization by a model human gut
586 microbiota containing Bacteroides cellulosilyticus WH2, a symbiont with an extensive
587 glycobiome. *PLoS Biol* **11**, e1001637 (2013).
588 13. Paulson, J.C., Blixt, O. & Collins, B.E. Sweet spots in functional glycomics. *Nat Chem*
589 *Biol* **2**, 238-248 (2006).
590 14. Heimburg-Molinaro, J., Mehta, A.Y., Tilton, C.A. & Cummings, R.D. Insights Into
591 Glycobiology and the Protein-Glycan Interactome Using Glycan Microarray
592 Technologies. *Mol Cell Proteomics* **23**, 100844 (2024).
593 15. Ruhaak, L.R., Xu, G., Li, Q., Goonatilleke, E. & Lebrilla, C.B. Mass Spectrometry
594 Approaches to Glycomic and Glycoproteomic Analyses. *Chem Rev* **118**, 7886-7930
595 (2018).
596 16. Raghavan, V. & Groisman, E.A. Species-specific dynamic responses of gut bacteria to a
597 mammalian glycan. *J Bacteriol* **197**, 1538-1548 (2015).
598 17. Modesto, J.L., Pearce, V.H. & Townsend, G.E., 2nd Harnessing gut microbes for glycan
599 detection and quantification. *Nat Commun* **14**, 275 (2023).
600 18. Townsend, G.E., 2nd, Raghavan, V., Zwir, I. & Groisman, E.A. Intramolecular
601 arrangement of sensor and regulator overcomes relaxed specificity in hybrid two-
602 component systems. *Proc Natl Acad Sci U S A* **110**, E161-169 (2013).
603 19. Cao, Y., Rocha, E.R. & Smith, C.J. Efficient utilization of complex N-linked glycans is a
604 selective advantage for Bacteroides fragilis in extraintestinal infections. *Proc Natl Acad*
605 *Sci U S A* **111**, 12901-12906 (2014).

606 20. Glowacki, R.W.P. & Martens, E.C. If you eat it, or secrete it, they will grow: the
607 expanding list of nutrients utilized by human gut bacteria. *J Bacteriol* **203** (2020).
608 21. Schwalm, N.D., 3rd & Groisman, E.A. Navigating the Gut Buffet: Control of
609 Polysaccharide Utilization in Bacteroides spp. *Trends Microbiol* **25**, 1005-1015 (2017).
610 22. Townsend, G.E., 2nd et al. A Master Regulator of Bacteroides thetaiotaomicron Gut
611 Colonization Controls Carbohydrate Utilization and an Alternative Protein Synthesis
612 Factor. *mBio* **11** (2020).
613 23. Ryan, D. et al. An integrated transcriptomics-functional genomics approach reveals a
614 small RNA that modulates Bacteroides thetaiotaomicron sensitivity to tetracyclines.
615 *bioRxiv* (2023).
616 24. Ryan, D., Jenniches, L., Reichardt, S., Barquist, L. & Westermann, A.J. A high-resolution
617 transcriptome map identifies small RNA regulation of metabolism in the gut microbe
618 Bacteroides thetaiotaomicron. *Nat Commun* **11**, 3557 (2020).
619 25. Terrapon, N. et al. PULDB: the expanded database of Polysaccharide Utilization Loci.
620 *Nucleic Acids Res* **46**, D677-D683 (2018).
621 26. Terrapon, N., Lombard, V., Gilbert, H.J. & Henrissat, B. Automatic prediction of
622 polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**, 647-655
623 (2015).
624 27. Glowacki, R.W.P. et al. A Ribose-Scavenging System Confers Colonization Fitness on
625 the Human Gut Symbiont Bacteroides thetaiotaomicron in a Diet-Specific Manner. *Cell*
626 *Host Microbe* **27**, 79-92 e79 (2020).
627 28. Reeves, A.R., D'Elia, J.N., Frias, J. & Salyers, A.A. A Bacteroides thetaiotaomicron outer
628 membrane protein that is essential for utilization of maltooligosaccharides and starch. *J*
629 *Bacteriol* **178**, 823-830 (1996).
630 29. Liu, H. et al. Functional genetics of human gut commensal Bacteroides thetaiotaomicron
631 reveals metabolic requirements for growth across environments. *Cell Rep* **34**, 108789
632 (2021).
633 30. Liou, C.S. et al. A Metabolic Pathway for Activation of Dietary Glucosinolates by a
634 Human Gut Symbiont. *Cell* **180**, 717-728 e719 (2020).
635 31. Pudlo, N.A. et al. Phenotypic and Genomic Diversification in Complex Carbohydrate-
636 Degrading Human Gut Bacteria. *mSystems* **7**, e0094721 (2022).
637 32. Cuskin, F. et al. Human gut Bacteroidetes can utilize yeast mannan through a selfish
638 mechanism. *Nature* **517**, 165-169 (2015).
639 33. Abbott, D.W., Martens, E.C., Gilbert, H.J., Cuskin, F. & Lowe, E.C. Coevolution of yeast
640 mannan digestion: Convergence of the civilized human diet, distal gut microbiome, and
641 host immunity. *Gut Microbes* **6**, 334-339 (2015).
642 34. Lowe, E.C., Basle, A., Czjzek, M., Firbank, S.J. & Bolam, D.N. A scissor blade-like
643 closing mechanism implicated in transmembrane signaling in a Bacteroides hybrid two-
644 component system. *Proc Natl Acad Sci U S A* **109**, 7298-7303 (2012).
645 35. Schwalm, N.D., 3rd, Townsend, G.E., 2nd & Groisman, E.A. Multiple Signals Govern
646 Utilization of a Polysaccharide in the Gut Bacterium Bacteroides thetaiotaomicron. *mBio*
647 **7** (2016).
648 36. Ryan, D. et al. An expanded transcriptome atlas for Bacteroides thetaiotaomicron
649 reveals a small RNA that modulates tetracycline sensitivity. *Nat Microbiol* **9**, 1130-1144
650 (2024).
651 37. Song, X. et al. Oxidative release of natural glycans for functional glycomics. *Nat*
652 *Methods* **13**, 528-534 (2016).
653 38. Vos, G.M. et al. Oxidative Release of O-Glycans under Neutral Conditions for Analysis
654 of Glycoconjugates Having Base-Sensitive Substituents. *Anal Chem* **95**, 8825-8833
655 (2023).

656 39. Wu, M. et al. Genetic determinants of in vivo fitness and diet responsiveness in multiple
657 human gut Bacteroides. *Science* **350**, aac5992 (2015).
658 40. Feng, J. et al. Polysaccharide utilization loci in Bacteroides determine population fitness
659 and community-level interactions. *Cell Host Microbe* **30**, 200-215 e212 (2022).
660 41. Kabonick, S.G. et al. Hierarchical glycolytic pathways control the carbohydrate utilization
661 regulator in human gut Bacteroides. *bioRxiv* (2025).
662 42. Deak, T. & Beuchat, L.R. Yeasts Associated with Fruit Juice Concentrates. *J Food Prot*
663 **56**, 777-782 (1993).
664 43. Temple, M.J. et al. A Bacteroidetes locus dedicated to fungal 1,6-beta-glucan
665 degradation: Unique substrate conformation drives specificity of the key endo-1,6-beta-
666 glucanase. *J Biol Chem* **292**, 10639-10650 (2017).
667 44. Zhu, Y. et al. Mechanistic insights into a Ca2+-dependent family of alpha-mannosidases
668 in a human gut symbiont. *Nat Chem Biol* **6**, 125-132 (2010).
669 45. Pearce, V.H., Groisman, E.A. & Townsend, G.E., 2nd Dietary sugars silence the master
670 regulator of carbohydrate utilization in human gut Bacteroides species. *Gut Microbes* **15**,
671 2221484 (2023).
672 46. Koropatkin, N.M., Martens, E.C., Gordon, J.I. & Smith, T.J. Starch catabolism by a
673 prominent human gut symbiont is directed by the recognition of amylose helices.
674 *Structure* **16**, 1105-1115 (2008).

675

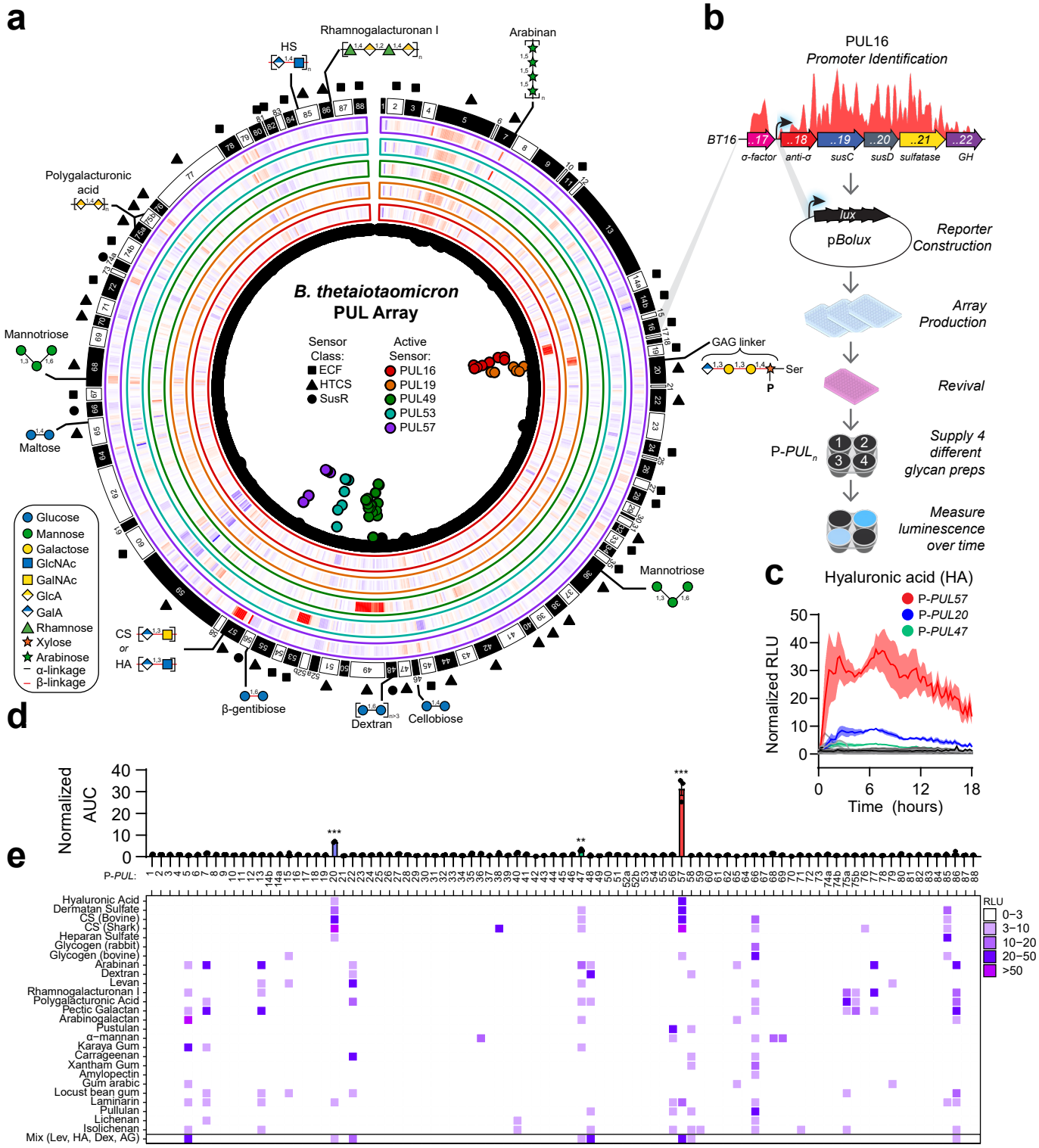

Fig. 1. Construction of a genome-wide *Bt* PUL reporter library array. a, Cartoon depicting 88 *Bt* PULs denoted by individual genes (innermost black circles). Dots emanating from PULs represent significantly increased transcripts measured by RNAseq from strains engineered to produce constitutively active PUL sensors from PUL16 (red), PUL19 (orange), PUL49 (green), PUL53 (cyan), or PUL57 (purple) and color-corresponding heatmaps. The outermost symbols represent distinct classes of known PUL sensor proteins and glycan structures that specifically activate various PULs are indicated. **b,** schematic depicting PUL reporter construction and glycan analysis using the arrayed library. **c-d,** Bioluminescence (**c**) from all 92 *Bt* reporters following supply of hyaluronic acid normalized by identical strains supplied galactose and (**d**) AUC over 18 hours from each strain. All significantly increased reporters have corresponding colors and insignificantly 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 and *** indicates P-values <0.001, **<0.01, and *<0.05.

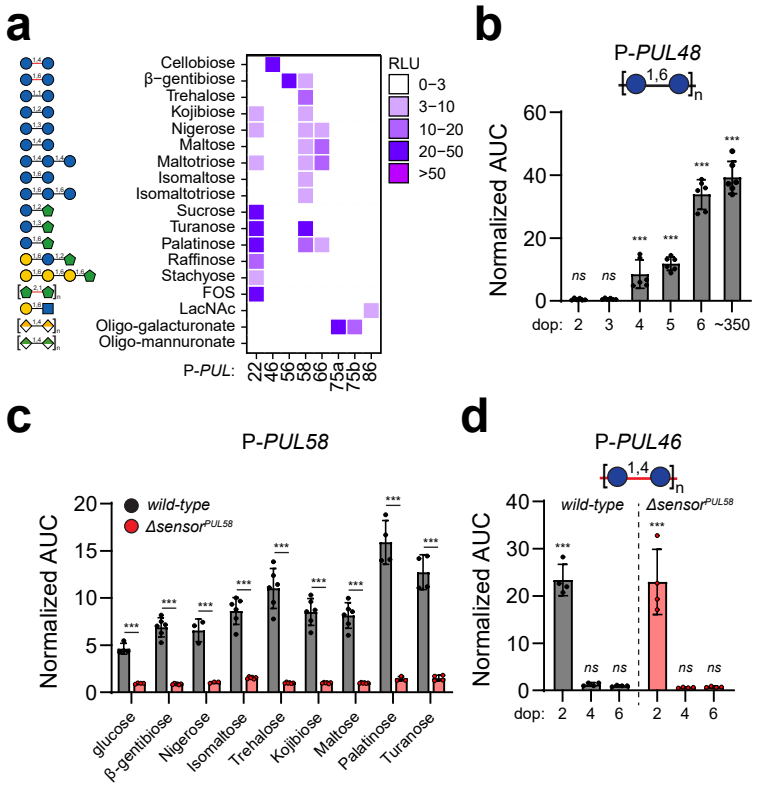

Fig. 2. *Bt* PUL reporters distinguish between monosaccharide composition, dop, and glycosidic linkage. a, A heatmap displaying normalized AUC following the introduction of various oligosaccharides to the *Bt* PUL reporter array. n=2. **b,** Normalized AUC from P-*PUL48* supplied isomaltooligosaccharides with increasing dop. N=6, error is SEM. **c,** Normalized AUC from *wild-type* (black) or *sensor*^{PUL58}-deficient (*BT2160*, 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 *sensor*^{PUL58}-deficient (*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.

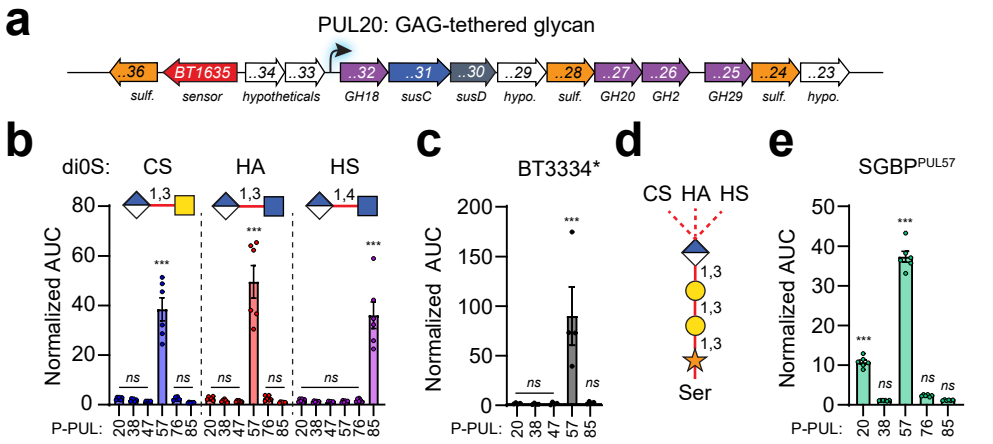

Fig. 3. PUL 20 responds to a unique GAG-tethered glycan. a, cartoon representing PUL20 activated by GAGs. **b**, Normalized bioluminescence from *wild-type Bt* strains harboring P-*PUL20,* P-*PUL38,* P-*PUL47,* P-*PUL57,* or P-*PUL85* supplied unsaturated disaccharides (di0S) derived from CS (blue), HA (red), or HS (purple) and normalized by responses from a strain harboring p*Bolux*. n=4, error is SEM. P-values were calculated using 2-way ANOVA with Fisher's LSD test. *** indicates values < 0.001. **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. **d,** Cartoon representation of the common GAG core linker glycan. **e,** Normalized bioluminescence from *wild-type 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=4, error is SEM.

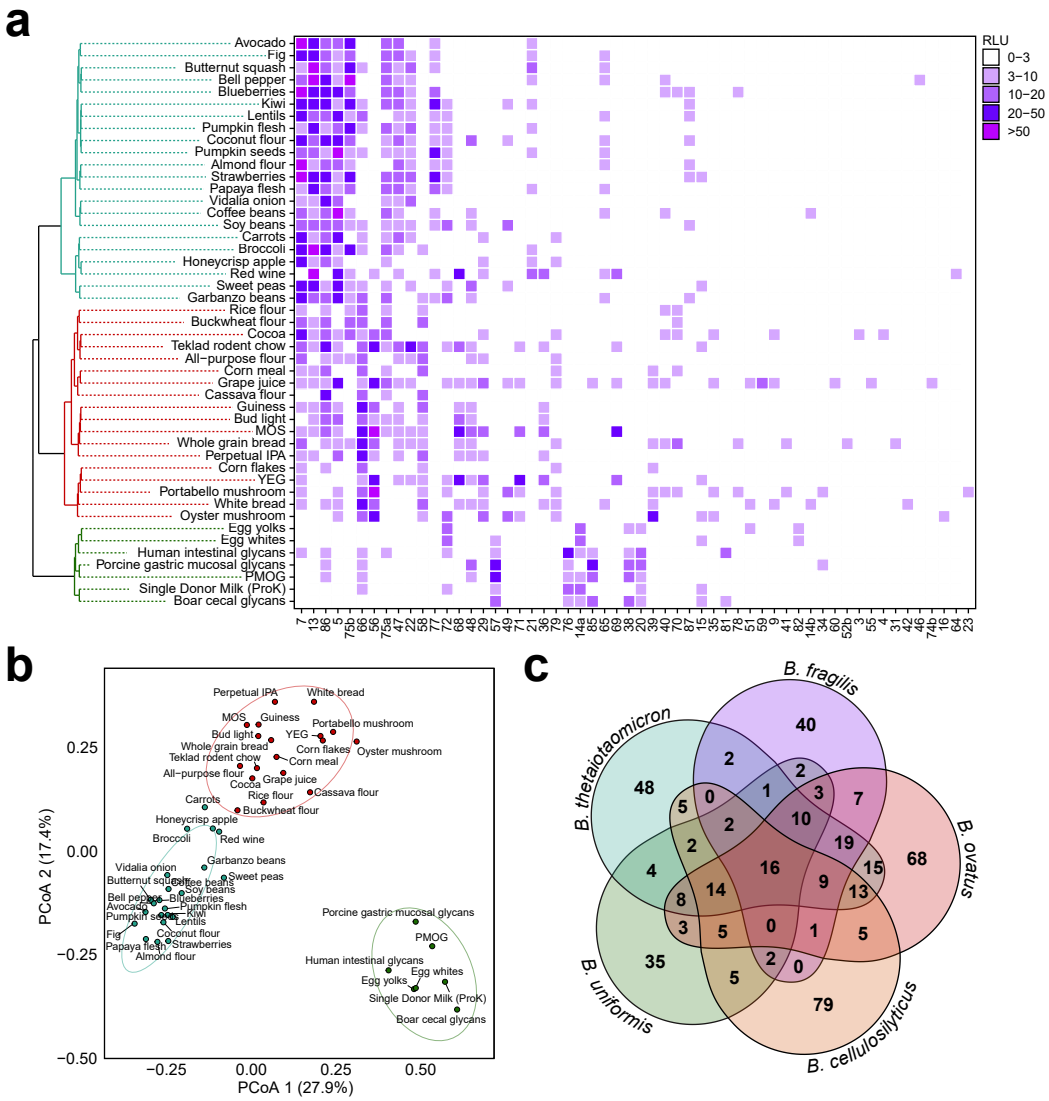

Figure 4. Arrayed *Bt* PUL reporter strains differentiate MAC sources. a, Heatmap representing Bt PUL array responses to the indicated biologically-derived mixtures clustered by similar reporter activation. n>2. **b,** PCoA of *Bt* reporter responses to different MACs from (**a**) using the Jaccard similarity index and Ward D2 distance calculations. **c,** Distribution of literature-derived and predicted PULs across *Bacteroides* species. Similarities were calculated using a previously described modularity equation.

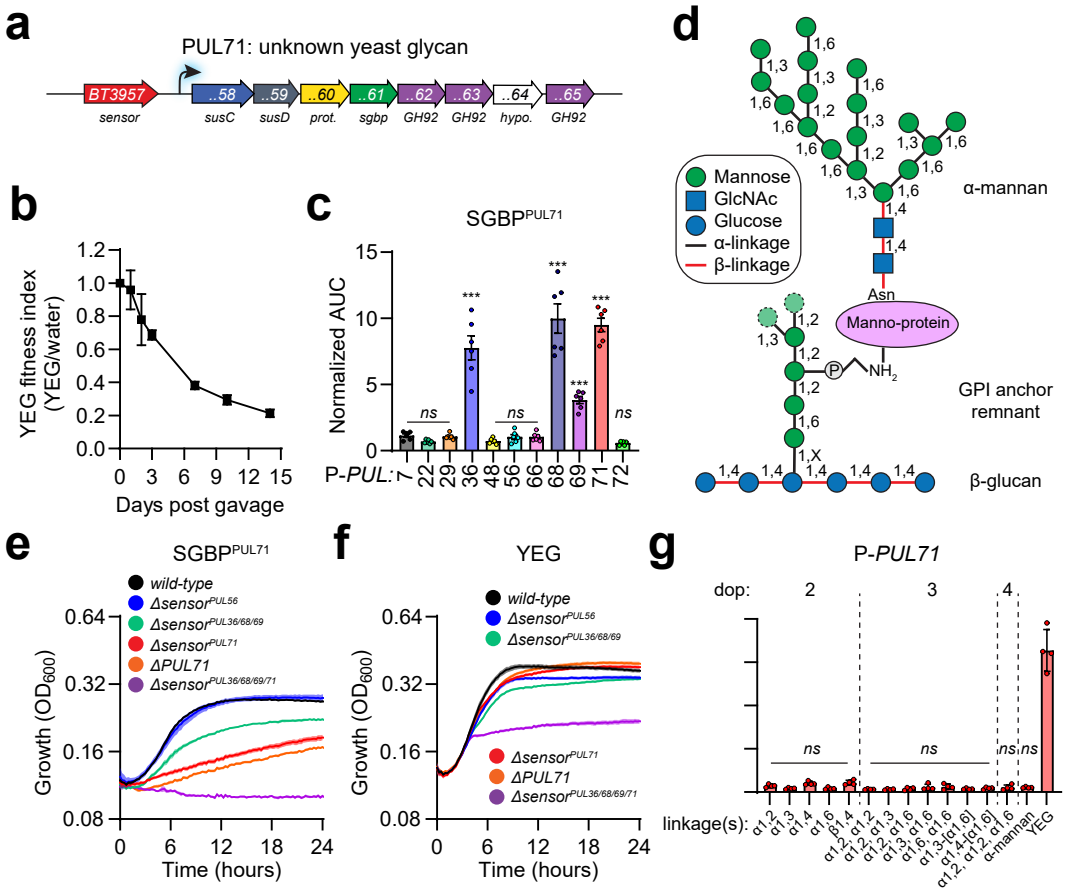

Fig. 5. PUL71 consumes an unknown yeast cell wall glycan. a, Cartoon schematic of PUL71, which is increased in response to yeast extract glycans. **b,** Fitness index of *PUL71*-deficient strain relative to *wild-type Bt* in mice supplied a 0.5% YEG solution compared to those supplied water. **c,** Normalized bioluminescence of YEG-induced reporter strains supplied BT3961-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. **d,** cartoon schematic of *Bt*-accessible yeast cell wall glycans, α-mannan and β1,6-glycan tethered by the GPI-anchored mannoprotein. **e-f,** Growth of *wild-type* and the indicated sensor-deficient *Bt* strains supplied (**e**) SGBP-^{PUL71}-co-purifying material following co-incubation with GH30^{PUL56}-treated or (**f**) untreated YEG. N=6, error is SEM. **g,** Normalized bioluminescence from a strain harboring P-*PUL71* supplied various di- and oligo-mannosaccharides, yeast α-mannan, or YEG. N=4, error is SEM. P-values were calculated using 2-way ANOVA with Fisher's LSD test. *** indicates values < 0.001 and ns > 0.05.