A Novel Family of RNA-Binding Proteins Regulate Polysaccharide Metabolism in *Bacteroides thetaiotaomicron*

Amanda N.D. Adams[a], Muhammad S. Azam[a,b], Zachary A. Costliow[a,c], Xiangqian Ma[a], Patrick H. Degnan[d#] and Carin K. Vanderpool[a#]

[a]Department of Microbiology, University of Illinois Urbana-Champaign, Urbana, Illinois, USA

[b]Present address: Department of Microbiology, University of Chicago, Chicago, Illinois, USA

[c]Present address: The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

[d]Department of Microbiology and Plant Pathology, University of California-Riverside, Riverside, California, USA

[#]Correspondence to: Patrick H. Degnan, patrickd@ucr.edu; or Carin K. Vanderpool, cvanderp@illinois.edu

Running Title: *Bacteroides* RBPs regulate polysaccharide metabolism

Abstract

Human gut microbiome composition is constantly changing, and diet is a major driver of these changes. Gut microbial species that persist in mammalian hosts for long periods of time must possess mechanisms for sensing and adapting to nutrient shifts to avoid being outcompeted. Global regulatory mechanisms mediated by RNA-binding proteins (RBPs) that govern responses to nutrient shifts have been characterized in Proteobacteria and Firmicutes but remain undiscovered in the Bacteroidetes. Here we report the identification of RBPs that are broadly distributed across the Bacteroidetes, with many genomes encoding multiple copies. Genes encoding these RBPs are highly expressed in many *Bacteroides* species. A purified RBP, RbpB, from *Bacteroides thetaiotaomicron* binds to single-stranded RNA *in vitro* with an affinity similar to other characterized regulatory RBPs. *B. thetaiotaomicron* mutants lacking RBPs show dramatic shifts in expression of polysaccharide utilization and capsular polysaccharide loci, suggesting that these RBPs may act as global regulators of polysaccharide metabolism. A *B. thetaiotaomicron* Δ*rbpB* mutant shows a growth defect on dietary sugars belonging to the raffinose family of oligosaccharides (RFOs). The Δ*rbpB* mutant had reduced expression of *BT1871*, encoding a predicted RFO-degrading melibiase, compared to the wild-type strain. Mutation of *BT1871* confirmed that the enzyme it encodes is essential for growth on melibiose and promotes growth on the RFOs raffinose and stachyose. Our data reveal that RbpB is required for optimal expression of *BT1871* and other polysaccharide-related genes, suggesting that we have identified an important new family of global regulatory proteins in the Bacteroidetes.

39 **Importance**

40 The human colon houses hundreds of bacterial species, including many belonging to
41 the genus *Bacteroides,* that aid in breaking down our food to keep us healthy.
42 *Bacteroides* have many genes responsible for breaking down different dietary
43 carbohydrates and complex regulatory mechanisms ensure that specific genes are only
44 expressed when the right carbohydrates are available. In this study, we discovered that
45 *Bacteroides* use a family of RNA-binding proteins as global regulators to coordinate
46 expression of carbohydrate utilization genes. The ability to turn different carbohydrate
47 utilization genes on and off in response to changing nutrient conditions is critical for
48 *Bacteroides* to live successfully in the gut, and thus the new regulators we have
49 identified may be important for life in the host.

50

51 **Key words** RRM-1, RNA-binding protein, Hfq, melibiose, PUL, CPS

52

53

54 **Introduction**

55 The human gut microbiome is an important player in host health, with diet being
56 one of the principal drivers of gut microbial composition and function (1–3). Dietary
57 carbohydrates, including complex polysaccharides and oligosaccharides are not readily
58 absorbed by the host and reach the distal gut where they are broken down and
59 metabolized by a consortium of microbes with diverse enzymatic capabilities (4).
60 Members of the dominant bacterial phylum Bacteroidetes can readily switch between
61 carbohydrate types as they become available due to dozens of substrate-specific
62 polysaccharide utilization loci (PUL) that encode proteins responsible for sensing and
63 catabolizing diverse polysaccharides (5–7). Characterized PULs are tightly regulated by
64 several distinct families of transcriptional regulators so that they are only abundantly
65 expressed when their substrates are available (5, 8–17). However, accumulating
66 evidence suggests that post-transcriptional regulation also plays an important part in gut
67 colonization and preferential use of carbohydrates through control of PULs (18–20).

68 Post-transcriptional regulation can be mediated by multiple regulators. In
69 *Bacteroides* species, the roles of small RNA (sRNA) regulators (19, 21) and other RNA
70 regulatory elements like riboswitches (22–24) in control of carbohydrate and vitamin
71 metabolism are beginning to be recognized. In well-studied Firmicutes and
72 Proteobacteria, post-transcriptional regulation of carbon metabolism and other systems
73 often occurs through the actions of sRNAs and their helper RNA chaperones (25).
74 Three of the most well-studied RNA chaperones include Hfq, CsrA, and ProQ.
75 Collectively, these three RNA-binding proteins (RBPs) regulate the bulk of the RNA
76 regulatory interactome in organisms like *Escherichia coli* and *Salmonella enterica*, and

77 each RBP has its own distinct RNA targets (26, 27). Hfq in particular functions as a
78 global post-transcriptional regulator of gene expression (28, 29). It binds to both mRNAs
79 and sRNAs, facilitating their interactions through short stretches of complementarity (30,
80 31). These interactions result in a variety of different regulatory outcomes, primarily
81 resulting from changes in translation initiation or mRNA stability (28). In many
82 organisms, mutation of *hfq* causes global changes in gene expression and pleiotropic
83 phenotypes (32–34).

84 Though regulatory RNA chaperones have not been characterized in the
85 Bacteroidetes, post-transcriptional regulation has been implicated in control of gene
86 expression in *Bacteroides* species (21, 35), including regulation of PUL expression (19).
87 In particular, the cis-antisense PUL-associated sRNA DonS in *B. fragilis* as well as
88 several PUL-associated cis-antisense sRNAs in *B. thetaiotaomicron* (19, 21) have been
89 implicated in modulation of PUL function through repression of carbohydrate transporter
90 gene expression. A recent study identified dozens of sRNAs encoded throughout the
91 genome of *B. thetaiotaomicron* with many being PUL-associated, suggesting that
92 sRNA-mediated regulation of PUL function may be a common phenomenon in
93 *Bacteroides* (21). Additionally, there are a growing number of examples of regulatory
94 effects in *Bacteroides* mediated by sequences in mRNA untranslated regions (UTRs)
95 (18, 36) and these may be mediated by as yet unidentified sRNAs or RNA chaperones.
96 To better understand the scope of RNA-mediated regulatory mechanisms in the
97 Bacteroidetes, we sought to identify and characterize RBPs that may act as regulatory
98 RNA chaperones.

99 Here we report the identification of a family of genes commonly found in
100 Bacteroidetes genomes, which encode RBPs with a single RNA Recognition Motif 1
101 (RRM-1) domain. These genes are conserved, often exist in multiple copies per
102 genome, and are highly expressed in many human gut *Bacteroides* isolates. We
103 demonstrate that a member of this family, RbpB, is a single-stranded (ss) RNA-binding
104 protein that binds with some specificity and affinities similar to other characterized RNA
105 chaperones. *B. thetaiotaomicron* mutants lacking one or more of these RBPs have
106 large-scale changes to their transcriptomes compared to the wild-type strain, with genes
107 belonging to PUL and capsular polysaccharide (CPS) loci being the most differentially
108 regulated. *B. thetaiotaomicron rbpB* mutants have growth defects on the common
109 dietary plant sugars raffinose family oligosaccharides due to decreased expression of
110 *BT1871*, an essential melibiase encoded in PUL24. Our findings suggest that this family
111 of RBPs play an important role in global regulation of polysaccharide metabolism in
112 *Bacteroides*.
113

114 **Results**

115 **Identification of a conserved family of RNA-binding proteins in the phylum**
116 **Bacteroidetes**

117 To identify putative RNA-binding proteins that may act as global regulators in the
118 Bacteroidetes, we compared a set of 313 human gut-associated microbial genomes
119 representing major phyla commonly found in gut microbial communities (37). We first
120 searched for canonical RNA chaperones – Hfq, ProQ, and CsrA, which are involved in
121 post-transcriptional regulation of gene expression in Proteobacteria. Using hidden

122 Markov models (HMMs) with trusted cutoffs, we identified Hfq in 23% (72/313) of the
123 genomes (a total of 79 Hfq homologs) mostly in the Proteobacteria, although there were
124 some identified in Firmicutes genomes (Fig.1A and Dataset S1). ProQ homologs (a total
125 of 40 across 36 genomes) were entirely restricted to the Proteobacteria, with the
126 majority being found in γ-proteobacterial genomes. CsrA homologs were identified in
127 25% (79/313) of genomes with a total of 93 CsrA homologs distributed across the
128 Proteobacteria and Firmicutes. We did not identify any Hfq, ProQ, or CsrA homologs
129 among Bacteroidetes genomes suggesting that if RNA chaperone regulators are
130 present in this phylum, they do not belong to these canonical families.

131 To identify other putative RNA chaperones in our model organism, we searched
132 the *B. thetaiotaomicron* VPI-5482 genome for proteins with conserved RNA-binding
133 domains. This yielded three intriguing candidates comprised of a single RNA
134 Recognition Motif 1 (RRM-1; PF00076) domain, here named RbpA (BT0784), RbpB
135 (BT1887), and RbpC (BT3840). The small, ~70 amino acid RRM-1 domain is one of the
136 most common RNA-binding domains in eukaryotes where it is typically found in
137 multidomain proteins involved in post-transcriptional RNA processing events including
138 regulation of RNA stability, translation, and turnover (38). Though poorly characterized,
139 many bacterial genomes appear to encode RRM-1 domain-containing proteins (39, 40).
140 Given the characterized roles of RRM-1 domain-containing proteins in post-
141 transcriptional RNA regulatory processes, we chose to focus on these homologs for
142 further characterization.

143 Expanding our search for RRM-1 domain-containing proteins to our larger set of
144 gut microbial genomes identified homologs of *B. thetaiotaomicron* RBPs in 69 of 313

145 genomes (Fig. 1A and Dataset S1). These proteins were widely distributed among
146 Bacteroidetes genomes accounting for 86% (149/174) of the total number of RRM-1
147 domain proteins identified. We also identified homologs in a small subset of
148 Proteobacterial genomes (Fig. 1A, Dataset S1). In contrast to eukaryotes, the bacterial
149 RRM-1 proteins we identified are small, single-domain proteins ranging in size from 60-
150 132 amino acids (aa), with the majority being 80-100 aa. Each protein contains a single
151 ferredoxin-like fold RRM-1 motif followed by a predicted disordered C-terminus of
152 varying lengths. This structure is reminiscent of the disordered C-termini of Hfq and
153 other RNA chaperones that plays a role in RNA-binding and cycling among various
154 binding partners (41–45). CsrA, Hfq, and ProQ homologs were largely encoded in single
155 copy with only a few instances of more than one copy in individual genomes (Dataset
156 S1). In contrast, RRM-1 genes frequently occurred in multiple copies per genome in
157 Bacteroidetes genomes (Table S1, Fig. 1B). Of the Bacteroidetes genomes we
158 analyzed, 50 of 58 contained one to four copies of genes encoding RRM-1 domain
159 proteins, with the majority of genomes containing three (Fig. 1B).
160 To analyze phylogenetic relationships among novel RRM-1 domain proteins
161 found in Bacteroidetes genomes, we used MCL (Markov cluster algorithm) with a 70%
162 amino acid identity cutoff and compared the resulting clusters to a species phylogeny of
163 the Bacteroidetes (Dataset S1). Clustering was chosen because the extent of
164 divergence among the homologs and their short lengths makes phylogenetic
165 reconstruction unreliable. The clustering revealed a complicated history of divergence
166 and duplication resulting in 10 clusters designated *rbpA–rbpJ.* It is notable that the three
167 loci represented in *B. thetaiotaomicron*, *rbpA, rbpB,* and *rbpC* represent the most

168 widespread clusters. Even within these clusters we identified evidence of likely
169 duplications or horizontal gene transfer among particular lineages, resulting in genomes
170 that encode two genes belonging to a single cluster. For example, the *Bacteroides*
171 *fragilis* 3_1_2 genome contains *rbpA* and *rbpA'* which share 84% amino acid identity.
172 The well-characterized RNA chaperone Hfq is an abundant protein in
173 Proteobacteria (46). To determine whether genes encoding *Bacteroides* RBPs show
174 similarly high levels of expression, we analyzed available RNA-seq data for *B.*
175 *thetaiotaomicron* (generated by us (see Materials and Methods) and others (23)) and
176 eight additional species – *Bacteroides caccae* (47), *Bacteroides cellulosyliticus* (48),
177 *Bacteroides dorei* (49), *Bacteroides massilensis* (18), *Bacteroides uniformis* (24),
178 *Bacteroides vulgatus* (24), and *Bacteroides xylanisolvens* (50) that were grown under a
179 variety of *in vitro* conditions. Virtually all of the genes encoding RBP homologs were
180 highly expressed in these datasets. Most *Bacteroides rbp* genes (represented by
181 colored triangles, Fig. 1C, 1D) were expressed at levels placing them among the top
182 10% of most highly expressed genes (represented by grey dots, Fig. 1C, 1D). Looking
183 specifically at *B. thetaiotaomicron rbpA, rbpB,* and *rbpC*, we observed that these genes
184 are highly expressed both *in vitro* (minimal medium with glucose and TYG medium) and
185 *in vivo* in monocolonized mice (51) or mice colonized with a synthetic consortium (47)
186 (Fig. 1D).

187 ***B. thetaiotaomicron* RbpB is a single-stranded RNA-binding protein**

188 To test the RNA-binding activity of a representative of this family of *Bacteroides*
189 RBPs, we conducted electrophoretic mobility shift assays (EMSAs). We overexpressed
190 and purified *B. thetaiotaomicron* RbpB and tested binding to a series of *in vitro*

191 transcribed single-stranded RNA (ssRNA) “pentaprobes” (52). The twelve ssRNA
192 pentaprobes are each 100-nucleotides (nts) in length and collectively contain all
193 possible 5-nt sequence combinations (Table S1). RbpB shifted ten of the twelve probes
194 to varying degrees, indicating that RbpB binds ssRNA *in vitro* and suggesting that it
195 does so with some degree of sequence specificity (Figs. 2A and S1A). RbpB showed no
196 evidence of binding to two of the pentaprobes PP6 and PP12 (Figs. S1A and 2B).
197 Probes PP2 and PP3 shifted at lower concentrations of RbpB compared with other
198 probes (Figs. 2A and S1A). The calculated K_D of RbpB binding to PP3 is 10.5 μM (Fig.
199 2C), a dissociation constant similar to those previously reported for RRM domains (38,
200 53). RRM domains can interact with a variable number of nts in the binding pocket, with
201 binding motifs that are typically 5-8 nts in length (54). To identify candidate RbpB
202 binding motifs in the pentaprobes, we used MEME motif discovery tool (55) to identify
203 sequence motifs (<9-nt in length) that occurred in RbpB-binding pentaprobes but were
204 absent in non-binding pentaprobes (Fig. S1B). MEME identified 10 such motifs in RbpB-
205 binding pentaprobe sequences. Motif 1, comprised of G, U, and A residues was the
206 most common motif found exclusively in the bound pentaprobes (Figs. 2D and S1B). A
207 C/U-rich motif, motif 9 (Fig. 2E and S1B), was present in probes PP2 and PP3, which
208 bound RbpB with higher affinity than other pentaprobes (Fig. 2A and C and S1A). To
209 test whether RbpB would bind specifically to motif 1 (5’-GUAGGAUA-3’) or motif 9
210 (5’UCCUGUGC-3’), we conducted EMSAs using new RNA oligonucleotide probes
211 containing three repeats of each motif. RbpB shifted the probe containing three copies
212 of motif 1 with a K_D of 5.1 μM (Fig. 2D). In contrast, RbpB did not shift a probe
213 containing three copies of motif 9 (Fig. 2E). Overall, these results demonstrate that

214 RbpB binds ssRNA with some degree of specificity at affinities comparable to known
215 RNA-binding proteins (56–58).

216 **Loss of RBPs leads to altered expression of PUL and CPS loci**

217 To assess possible functions of RBPs in *B. thetaiotaomicron*, we made mutant
218 strains lacking *rbpA* and *rbpB*. We generated three strains: Δ*rbpA (BT0784)*, Δ*rbpB*
219 *(BT1887),* and Δ*rbpA*Δ*rbpB.* We were unable to generate a Δ*rbpC* mutant. We
220 performed RNA-seq on RNA samples from Δ*rbpA,* Δ*rbpB,* and Δ*rbpA*Δ*rbpB* strains
221 grown to mid-log or stationary phase in rich medium (TYG). Among protein coding
222 genes, 12.3% (587/4,778) were significantly differentially regulated (q-value <0.06, $\log_2$
223 fold-change of ≥+1 or ≤-1) in at least one condition (Dataset S2A), with Δ*rbpA*Δ*rbpB*
224 having the greatest number of differentially regulated genes among the three mutants.
225 To identify functional classes of differentially expressed genes, Gene Set Enrichment
226 Analysis (GSEA) (59) was used with *B. thetaiotaomicron*-specific custom gene sets (see
227 Materials and Methods). Differentially regulated genes that were not categorized in
228 GSEA were further grouped according to Gene Ontology (Dataset S2A and Materials
229 and Methods). Considering all differentially regulated genes across all strains, enriched
230 functional groups included CPS loci, PULs, hypothetical proteins, transmembrane
231 transport, redox activities, B-vitamin metabolism, transcription, translation, and a variety
232 of other metabolic pathways (Fig. 3A). The largest functional group of differentially
233 regulated genes was CPS genes, accounting for 17% (98/587) of differentially regulated
234 genes across all six wild-type to mutant comparisons (Fig. 3A, B). CPS loci encode
235 functions that produce the polysaccharide coats that surround the *Bacteroides* cell
236 surface (10, 17, 60). Of the eight CPS loci in *B. thetaiotaomicron* VPI-5482, five were

237 differentially regulated across the three mutant strains including CPS2, CPS4, CPS5,
238 CPS6, and CPS7 loci (Fig. 3B). CPS4 and 6 loci were downregulated in all three
239 mutants compared to wild-type in both conditions, whereas CPS2, 5, and 7 loci were
240 upregulated in some mutants compared to wild-type in a subset of conditions. CPS2
241 was upregulated in both Δ*rbpA* and Δ*rbpB* mutants in mid-log and stationary phase, but
242 was unchanged in the Δ*rbpA*Δ*rbpB* mutant in either growth condition, implying a genetic
243 interaction between *rbpA* and *rbpB* in the regulation of the CPS2 locus. Expression
244 patterns for all mutants were similar between mid-log and stationary phase conditions,
245 except for the CPS7 locus. CPS7 was upregulated in Δ*rbpA* and Δ*rbpA*Δ*rbpB* mutants
246 compared to wild-type in mid-log cells but only the Δ*rbpA*Δ*rbpB* mutant showed a
247 difference with wild-type in stationary phase.

248 PULs were the second most abundantly represented functional group among
249 differentially regulated genes (Fig. 3A). Of the 88 annotated PULs (5), 29 (33%) had at
250 least one differentially regulated gene in *rbp* mutant strains compared to wild-type,
251 accounting for 29% (75/263) of genes across the 29 PULs (Fig. 3C). In contrast to CPS
252 expression, PUL expression differences in mutant strains frequently varied according to
253 growth phase. PUL56 was downregulated in all three mutant strains exclusively during
254 stationary phase. In contrast, PUL71 was downregulated in all three mutant strains
255 during stationary phase but in mid-log was only downregulated in single mutants Δ*rbpA*
256 and Δ*rbpB* (Fig. 3C). Similar to the CPS loci, several PULs demonstrated expression
257 patterns indicative of a genetic interaction between *rbpA* and *rbpB*, including PUL22.
258 PUL22 was upregulated in the Δ*rbpB* mutant but downregulated in Δ*rbpA*Δ*rbpB* double
259 mutant during mid-log growth. In contrast, in stationary phase, PUL22 was upregulated

260 in Δ*rbpA* and downregulated in Δ*rbpB* and Δ*rbpA*Δ*rbpB* mutant strains. Several PULs
261 including PUL08, 10, 14, 51, 59, 75, 80, and 81, were differentially regulated in specific
262 single mutant strains but not differentially regulated in the Δ*rbpA*Δ*rbpB* double mutant.
263 We also saw some expression patterns that may be indicative of redundant regulation
264 by RbpA and RbpB – PULs 36, 45, 54, 72, 73, and 82 were not differentially expressed
265 in the single deletion mutants but were differentially expressed in the Δ*rbpA*Δ*rbpB*
266 double mutant. Collectively, these results suggest that RbpA and RbpB play global roles
267 in *B. thetaiotaomicron* gene expression, and in particular suggest that they coordinate
268 capsular polysaccharide production and carbohydrate utilization through control of CPS
269 and PUL genes, respectively.

270 **The Δ*rbpB* mutant is defective for growth on raffinose family oligosaccharides**

271 To determine whether *B. thetaiotaomicron* RBPs are required for growth on
272 specific carbohydrates, we carried out an initial screen of Δ*rbpA*, Δ*rbpB*, and
273 Δ*rbpA*Δ*rbpB* strains for growth defects on Biolog plates containing a variety of carbon
274 sources (Fig. S2 and Dataset S3). All three strains were defective for utilization of a
275 number of dietary and host-associated glycans (Fig. S2 and Dataset S3). As observed
276 for the transcriptome, phenotypes for the double mutant Δ*rbpA*Δ*rbpB* strain did not
277 recapitulate all growth defects observed in single mutant Δ*rbpA* or Δ*rbpB* strains,
278 implying a genetic interaction between *rbpA* and *rbpB*. For example, the Δ*rbpA* strain
279 showed faster growth than the wild-type strain on maltotriose, α-methyl-D-galactoside,
280 α-D-lactose, and lactulose while showing slower growth on sucrose, D-trehalose,
281 turanose, D-mannose, and palatinose as compared to wild-type. Defects on turanose,
282 D-trehalose, and palatinose were recapitulated in the Δ*rbpA*Δ*rbpB* strain, but the other

283 growth changes seen in the Δ*rbpA* strain were not observed in the Δ*rbpA*Δ*rbpB* strain.
284 The Δ*rbpB* strain grew slower than wild-type on D-melibiose, β-methyl-D-galactoside,
285 palatinose, and mannan, and defects on β-methyl-D-galactoside, palatinose, and
286 mannan were also observed for the Δ*rbpA*Δ*rbpB* strain. Interestingly, all three strains
287 were defective for growth on palatinose and a methylated galactoside. Unique to
288 Δ*rbpA*Δ*rbpB* was slow growth on gentiobiose and N-acetyl-D-galactosamine. Overall,
289 these results are consistent with transcriptome results that suggest that both *rbpA* and
290 *rbpB* play a role in regulation of carbohydrate utilization.
291 One of the carbohydrates on which the Δ*rbpB* mutant alone had substantial
292 growth defects was D-melibiose, a subunit of the raffinose family oligosaccharides
293 (RFOs) (Fig. S2). Since RFOs are prevalent in the human diet and are available to
294 organisms that can metabolize them in the distal gut, we chose this phenotype for
295 further evaluation. RFOs consist of the disaccharide sucrose (α-1,2-glucose-fructose)
296 bound to repeating α-1,6-galactosyl residues producing the trisaccharide raffinose and
297 the tetrasaccharide stachyose (Fig. 4A), along with the pentasaccharide verbascose. In
298 addition to sucrose, the α-1,6-galactose-glucose disaccharide melibiose is an RFO
299 subunit. When grown in minimal medium with RFOs or their subunits as the sole carbon
300 source, the Δ*rbpB* strain displayed growth defects on melibiose, raffinose, and
301 stachyose (Fig. 4B). The doubling time of the wild-type strain on minimal medium with
302 melibiose as the sole carbon source was 2.86 hours compared to twice that for the
303 Δ*rbpB* strain (5.76 hours). The Δ*rbpA* and Δ*rbpA*Δ*rbpB* strains did not have growth
304 defects on these substrates, again consistent with possible genetic interactions between
305 RbpA and RbpB with respect to growth on RFOs. Δ*rbpA*, Δ*rbpB*, and Δ*rbpA*Δ*rbpB*

306 strains showed no growth defects on monosaccharide subunits of RFOs, including
307 glucose, galactose, and fructose, or on the disaccharide sucrose (Fig. 4B-C),
308 suggesting that the Δ*rbpB* growth defect is due to the inability of this strain to utilize
309 sugars containing the α-1,6-galactose-glucose linkage.

310 Complementation of the Δ*rbpB* strain was attempted with two different constructs
311 (Fig. S3A). Neither complementation construct restored growth of the Δ*rbpB* mutant on
312 RFOs (Fig. S3B). We measured *rbpB* mRNA levels from wild-type (*rbpB*$^+$), Δ*rbpB,* and
313 both complementation strains and found that levels of *rbpB* mRNA in complementation
314 strains were significantly lower than in the wild-type strain (Fig. S3C), which may
315 account for the inability to restore growth on melibiose.

316 Upon further inspection of the *rbpB* (*BT1887*) native locus in our TYG RNA-seq
317 data, we noticed reduced expression of the immediately adjacent genes *BT1886*,
318 *BT1885*, and *BT1884* in the Δ*rbpB* strain, especially in mid-log phase, suggesting a
319 possible polar effect of the *rbpB* mutation on *BT1886-BT1884* (Fig. S4A). We also
320 observed a single transcription start site upstream of *rbpB-BT1884* in TYG (21), a
321 terminator prediction after the *rbpB* ORF, and a terminator prediction after *BT1884* (21),
322 suggesting *rbpB* may be expressed as both a monocistronic mRNA and polycistronic
323 with *BT1886-BT1884* (Fig. S4B). To determine if we could detect *rbpB* co-transcription
324 with *BT1886-BT1884* under conditions relevant to the *rbpB* mutant phenotype we
325 conducted RT-PCR on RNA samples harvested from wild-type (*rbpB*$^+$) cells grown to
326 mid-log phase on minimal media with glucose or melibiose. Primer sets spanning
327 junctions between each gene in the putative operon yielded PCR products (Fig. S4C),
328 suggesting that *rbpB* and *BT1886-BT1884* are expressed as an operon. *BT1886*,

329 *BT1885*, and *BT1884* encode a putative RhlE DEAD-box RNA helicase, a hypothetical
330 protein, and a cold shock domain containing protein, respectively. The operon structure
331 suggests the functions of these proteins are linked. One other possibility that may
332 explain the inability to complement the *rbpB* mutant melibiose growth phenotype is that
333 the appropriate stoichiometry of these proteins was not restored by the
334 complementation constructs.

335 **Loss of RbpB leads to decreased expression of an essential melibiase in PUL24**

336 Given the inability of Δ*rbpB* to utilize α-1,6 linked RFOs, we hypothesized that a
337 gene (or genes) encoding an α-galactosidase would be differentially regulated in the
338 Δ*rbpB* mutant strain compared to wild-type. Although there are several α-galactosidases
339 annotated in the genome (7), none of them were significantly differentially regulated in
340 our TYG RNA-seq, suggesting differential regulation may be specific to growth in
341 minimal medium with melibiose. We therefore performed more RNA-seq to identify
342 candidate genes responsible for this phenotype. We compared transcriptome profiles of
343 wild-type and Δ*rbpB* strains grown in minimal media with glucose or melibiose. To
344 identify genes that are uniquely transcriptionally responsive to the α-1,6 linkage in
345 melibiose, we also compared the glucose- and melibiose-grown cells' transcriptomes to
346 that of cells grown in minimal medium with a 1:1 mixture of glucose and galactose, the
347 monosaccharides that makeup melibiose.

348 Comparing wild-type and Δ*rbpB* transcriptomes in all three media, we identified
349 genes in PUL24 that were strongly differentially regulated (Fig. 5A and Dataset S2B).
350 PUL24 (genes *BT1871-BT1878*) contains a SusC/D-like pair (BT1874-BT1875), a
351 σ/anti-σ factor pair (BT1876-BT1877), and four putative glycosyl hydrolases belonging

352 to families GH3 (BT1872), GH43 (BT1873), GH76 (BT1878), and GH97 (BT1871).
353 Genes *BT1873-BT1878* were expressed at very low levels in wild-type and Δ*rbpB*
354 strains in all of the conditions we tested (Fig. 5A, Dataset S2B) suggesting that these
355 genes are not involved in glucose, galactose, or melibiose metabolism. However,
356 *BT1871* and *BT1872* were highly expressed in the wild-type strain growing in glucose
357 and melibiose (Fig. 5A) (and the glucose-galactose mixture, Dataset S2B) but were
358 expressed at barely detectable levels in the Δ*rbpB* strain. In contrast, *BT1871* and
359 *BT1872* were not differentially expressed between wild-type and Δ*rbpB* strains grown in
360 TYG (Dataset S2A).
361 Previous work showed that BT1871 has *in vitro* melibiase activity (61), and
362 transposon insertions in *BT1871* led to decreased fitness in melibiose in a carbohydrate
363 utilization screen (62). To confirm that BT1871 was important for *B. thetaiotaomicron*
364 utilization of RFOs including melibiose as a sole carbon source, we deleted *BT1871* and
365 cultured the Δ*BT1871* and wild-type strains in minimal media with melibiose, raffinose,
366 stachyose, and sucrose (Fig. 5B). The Δ*BT1871* strain showed no growth defect on
367 sucrose compared to the wild-type strain, which is expected based on its predicted
368 melibiase activity. In contrast, the Δ*BT1871* mutant could not grow on melibiose,
369 indicating *BT1871* is essential for melibiose utilization. Additionally, the Δ*BT1871* mutant
370 showed reduced growth on raffinose and stachyose compared to wild-type (Fig. 5B),
371 indicating that *BT1871* is required for metabolism of RFOs in general. Residual growth
372 of the Δ*BT1871* mutant on raffinose and stachyose is presumably due to the ability to
373 utilize fructose from the α-1,2 sucrose linkage.

374 We constructed three different complementation strains to confirm that BT1871 is
375 responsible for the melibiose growth defect. There is a single predicted promoter
376 upstream of *BT1872* (21) and the *BT1872* and *BT1871* open reading frames are
377 separated by only 32 bp suggesting that they are co-expressed. Complementation
378 strain 1 (compl1, Fig. 5C) carried the native promoter upstream of *BT1872* followed by a
379 deletion of the *BT1872* ORF and the intact *BT1871* gene. Complementation strain 2
380 (compl2, Fig. 5C) carried the intact promoter and *BT1872* and *BT1871* genes.
381 Complementation strain 3 (compl3, Fig. 5C) carried the promoter and *BT1872* only. The
382 compl1 construct partially restored growth on melibiose, raffinose, and stachyose (Fig.
383 5D). The compl2 construct improved growth on RFOs compared to the compl1
384 construct, whereas the compl3 construct (*BT1872* alone) failed to complement (Fig. 5D)
385 Taken together these results indicate that *BT1871* is an essential melibiase required for
386 RFO utilization and the decrease in *BT1871* mRNA in the Δ*rbpB* strain is responsible for
387 the melibiose growth defect.
388

389 **Discussion**

390 Though it is well established that rapid nutrient shifts affect the composition and
391 metabolic activities of gut microbes (47, 63–65), the regulatory mechanisms that allow
392 them to sense and rapidly adapt to use of different nutrient sources are poorly
393 understood. Canonical mechanisms for global transcriptional regulation of carbon
394 source utilization in model organisms from the phyla Proteobacteria and Firmicutes(66)
395 are absent in the Bacteroidetes (67–69). Likewise, RNA chaperones and RNA-mediated
396 post-transcriptional regulatory mechanisms that coordinate metabolism and responses

397 to changing environmental conditions (28, 70–73) are commonly found in
398 Proteobacteria and Firmicutes but have not been described in Bacteroidetes. In this
399 study, we identify a family of conserved RNA-binding proteins that is broadly distributed
400 among members of the Bacteroidetes and some Proteobacteria that lack canonical
401 RNA chaperone regulators. These RBPs occur in multiple copies in a given genome
402 and are highly expressed in a number of *Bacteroides* species from the human gut in
403 culture and in mouse models. At least one of these proteins from *B. thetaiotaomicron*,
404 RbpB, is able to bind ssRNA *in vitro* in a sequence specific manner. Deletion of *rbpA*
405 and *rbpB* in *B. thetaiotaomicron* leads to global dysregulation of CPS loci and PULs and
406 perturbed growth on a variety of carbohydrate sources. Overall, these results suggest
407 that this family of RBPs may play global regulatory roles in carbohydrate metabolism in
408 the *Bacteroides*.
409 While our study provides strong evidence for the importance of these RBPs in
410 global regulation of gene expression, the mechanisms by which RBPs mediate these
411 effects are still unknown. We hypothesize that RBPs act as RNA chaperones that
412 control mRNA stability and translation by binding to target mRNAs and modulating
413 ribosome association or access of RNases (Fig. 6). RBP modulation of mRNA
414 translation or stability may be through direct interaction of RBPs with target mRNAs
415 (Fig. 6A) or through facilitating base pairing of sRNAs to mRNAs (Fig. 6B), either of
416 which could result in changes to mRNA structure that alter accessibility to ribosomes or
417 RNases. Little is known about RNA chaperone function in the Bacteroidetes, but in the
418 case of RbpB, its role in RNA metabolism is supported by its genomic location.
419 Annotations for BT1885 (DEAD-box RNA helicase) and BT1884 (cold-shock protein)

420 suggest that *rbpB-BT1884* may be an RNA metabolism operon. To date, we do not
421 have evidence supporting or refuting a role for RBPs in modulation of sRNA function.
422 However, recent literature suggests that sRNAs may play an important role in
423 modulation of carbohydrate metabolism in *Bacteroides* (19, 21). One recent study
424 described an N-acetyl-D-glucosamine-inducible sRNA called GibS that binds *in vitro* to
425 mRNAs involved in carbohydrate metabolism. Mutant strains lacking GibS had nine
426 differentially-regulated genes compared to the wild-type parent strain. Two of these
427 were *BT1871-BT1872,* where expression was reduced in the Δ*gibS* compared to the
428 wild-type strain. GibS binding to *BT1871* mRNA was predicted *in silico* but could not be
429 demonstrated *in vitro*. The authors speculated that GibS binding to *BT1871* mRNA
430 required an unidentified RNA chaperone. To test whether RbpB facilitates RFO
431 utilization by a GibS-dependent mechanism, we generated deletion mutants Δ*gibS* and
432 Δ*rbpB*Δ*gibS* and grew these strains alongside wild-type in the presence of melibiose
433 (Fig. S5A). The Δ*gibS* mutant grew similarly to wild-type and the Δ*rbpB*Δ*gibS* mutant
434 grew similarly to the Δ*rbpB* parent strain on melibiose (Fig. S5A). RT-qPCR showed that
435 levels of *BT1871* mRNA were also similar between wild-type and Δ*gibS* strains (Fig.
436 S5B), suggesting that GibS does not play a major role in modulating *BT1871* mRNA
437 levels in our growth conditions. Overall, these results suggest that under our growth
438 conditions, RbpB regulates *BT1871* independently of GibS.
439 We have yet to explore the role of RBPs in helping *B. thetaiotaomicron* colonize
440 or be maintained in the host gut. In a recent study (62) screening transposon (Tn)
441 mutants for a wide variety of *in vitro* and *in vivo* phenotypes, there were no reported
442 insertions in *rbpA* or *rbpB*. Insertions in *rbpC* led to reduced fitness on glucose-

443 containing media (62), possibly explaining our inability to generate Δ*rbpC* mutants in our
444 standard glucose-rich media. The *rbpC* mutants also had an increased growth on
445 melibiose suggesting that *rbpC* also plays a role in utilization of RFOs. In the same
446 study (62), colonization of germ-free mice fed a plant polysaccharide-rich diet with the
447 *B. thetaiotaomicron* Tn-mutant pool led to increased fitness of *rbpC* mutants. *BT1871*
448 mutants showed decreased fitness over time, whereas Tn-insertions into several other
449 PUL24 genes led to increased fitness *in vivo*. Overall, these data indicate that RBPs
450 and the PULs they regulate may be important for *in vivo* fitness.
451 Regulation by RBPs may represent a critical mechanism for coordination of
452 carbohydrate utilization and production of cell surface capsular polysaccharides.
453 Differential regulation of PULs and CPS loci in *rbp* deletion strains is consistent with
454 several reports indicating a regulatory link between these polysaccharide metabolic
455 processes in *B. thetaiotaomicron* (17, 74, 75). Our RNA-seq data showed that deletion
456 of RBPs leads to reduced expression of CSP4 and CPS6 and increased expression of
457 CPS5. In *B. thetaiotaomicron,* CPS4 is normally the most highly expressed locus *in vitro*
458 and in mouse models when dietary glycans are present (64, 75, 76), and disruption of
459 CPS4 expression leads to decreased fitness in mouse competitions (77, 78). In a study
460 monitoring CPS expression in a mouse model over time, it was observed that even
461 when the *B. thetaiotaomicron* inoculum expressed one dominant CPS locus, expression
462 over time varied between mice (75). While CPS4 was most often highly expressed in
463 mice fed a high-fiber diet, mice on fiber-free diets typically expressed CPS5 or CPS6. *B.*
464 *thetaiotaomicron* mutants that could only express a single CPS locus had a decreased
465 ability to recover from antibiotic-induced stress (75). These studies along with our

466 present work collectively suggest that the ability to shift among different CPS types is
467 advantageous in the host and that this regulation may be mediated in part by RBPs.
468 Further characterization of the RBPs and their regulatory mechanisms may provide
469 critical insight into how *Bacteroides* coordinately control carbohydrate availability with
470 cell surface properties. This could reveal key principles governing mechanisms in host
471 dynamics.
472
473
474

475 **Materials and Methods**

476 **Bacterial culturing and genetic manipulation**

477 *B. thetaiotaomicron* VPI-5482 strains were grown anaerobically in a Coy
478 Laboratory Products vinyl anaerobic chamber with an input gas of 20% CO_2, 10% H_2,
479 70% N_2 balance. Routine culturing of *B. thetaiotaomicron* was done in Tryptone-Yeast-
480 Extract-Glucose (TYG) (79) broth and on Difco Brain Heart Infusion (BHI) agar plates
481 with 10% defibrinated Horse Blood (HB) (Quad Five) at 37˚C. *Escherichia coli* strains
482 were grown aerobically at 37˚C on BHI-10%HB for conjugations and Luria Broth for all
483 other applications. Minimal medium (22) was supplemented with B_{12} [3.75nM final]
484 (Sigma) and carbohydrates as needed at the following final w/v concentrations unless
485 otherwise indicated: 4.0% stachyose (Sigma), 2.0% D-(+)-raffinose (Sigma), 0.5% D-
486 (+)-melibiose (Sigma), 0.5% α-D-glucose (Sigma), 0.5% D-(+)-galactose (Sigma), 0.5%
487 β-D-(-)-fructose (MP Biomedicals), and 0.5% sucrose (MP Biomedicals). When needed,
488 antibiotics were added at the following final concentrations: 100 µg/ml ampicillin
489 (Sigma), 200 µg/ml gentamicin (Goldbio), 25 µg/ml erythromycin (VWR), 200µg/ml 5'-
490 fluoro-2'-deoxyuridine (VWR), 100ng/ml anhydrotetracycline (Sigma), 25 µg/ml
491 kanamycin (Fisher). All strains, vectors, and primers are listed in Table S1. For all
492 experiments, wild-type *B. thetaiotaomicron* is Δ*tdk* (strain AA0014 in Table S1).

493 Markerless deletions were made in *B. thetaiotaomicron* using the
494 pExchange_*bla_tdk_ermGb* (80) and the pLGB13_*bla_ermG* (81) suicide vector-based
495 allelic exchange methods. Upstream and downstream regions of the gene to be deleted
496 were amplified using Kappa HiFi (Kappa Biosystems) and cloned into

497 pExchange_*bla_tdk_ermGb* using standard restriction digest and ligation methods and
498 splicing by overlap exchange (SOE) (22). Alternatively, inserts were cloned into Q5
499 (NEB) amplified pExchange_*bla_tdk_ermGb* using restriction digest and ligation of a
500 gBlock insert (IDT). GibS flanks were Q5 amplified and ligated to restriction-digested
501 pLGB13_*bla_ermG.* Complete vectors were conjugated into *B. thetaiotaomicron* with *E.*
502 *coli* S17 λ-pir using established methods (22). Complementation pNBU2_*bla_ermGb*
503 vectors (5, 82) were cloned using standard restriction digest and ligation methods and
504 conjugated into *B. thetaiotaomicron* as done with pExchange. pNBU2_*bla_ermGb*
505 vectors were PCR screened for insertion into a single attachment site as done
506 previously (5). The pET-28a-*rbpB* protein expression vector was generated by inserting
507 the *rbpB* (*BT1887*) ORF 5' to the thrombin cleavage site and 6xHis tag in the pET28a
508 backbone. pET-28a-*rbpB* was cloned using Q5 PCR amplification and NEBuilder
509 assembly (NEB) in *E. coli* XL10-Gold competent cells (Agilent) before being moved into
510 *E. coli* BL21 (DE3) for protein expression.

511 **Computational identification of RNA regulators in human gut-associated**
512 **microbial genomes**

513 To identify genomes containing CsrA, ProQ, RRM-1, and Hfq in the human gut
514 microbiome, we utilized a custom database of 313 human gut-associated microbial
515 genomes containing a single representative genome for a species (22, 37). Candidate
516 RNA regulator genes were identified using hmmer v3.3 (hmmer.org) with trusted cutoffs
517 and the individual hidden Markov model from each protein queried: Hfq, PF17209.4;
518 CsrA, PF02599.17; ProQ, PF04352.14; RRM-1, PF00076.23 (39). The resulting gene

519 list was then run against Pfam-A.hmm version 33.1 using hmmer to verify that the query
520 PFAM was the top hit for the target domain using trusted cutoff values. ProQ
521 PF04352.14 gene hits that also contained an N-terminal FinO_N domain (PF12602.9)
522 were removed from the final annotation list. RRM-1 PF00076.23 gene hit list was limited
523 to less than 150 amino acids to remove a few genes containing transmembrane
524 domains. The maximum likelihood phylogenetic tree in Fig. 1A was built using a
525 multisequence alignment of 13 conserved core genes (AspS, Ffh, FusA, GltX, InfB,
526 LeuS, RplB, RpsE, RpsH, RpsK, TopA, TufA, RpoB) identified and described previously
527 (83). Briefly, protein sequences for each group of orthologs were individually aligned
528 with MUSCLE (84), concatenated, and subjected to phylogenetic reconstruction with
529 RAxML(85). The phylogeny was visualized using FigTree
530 (http://tree.bio.ed.ac.uk/software/figtree/).
531

532 **RBP expression in publicly available RNA-seq datasets**

533 Publicly available RNA-seq datasets were downloaded from NCBI (see Table S2)
534 for sample IDs. RNA-seq reads were quality filtered with Trimmomatic v0.36 (86). Read
535 mapping and sample normalization was calculated with Rockhopper v2.03 (87, 88) and
536 normalized expression values were graphed using JMP v15 (89).

537 **RbpB EMSAs and motif identification**

538 **Purification of RbpB.** *E. coli* BL21 (DE3) cells with the pET-*rbpB* vector were
539 grown to late exponential phase (0.6-0.8 OD_{600} as measured on an Ultraspec 2100 Pro,
540 Amersham) and protein expression induced with 1 mM final IPTG (Goldbio) for 4 hr at

541 37 ºC. Cells were harvested by centrifugation and pellets resuspended in 30 ml of
542 extraction buffer (1X PBS, 0.5M NaCl, pH 7.2) before being lysed in a French press.
543 Supernatant was collected after centrifugation at 16,000 × g for 10 min at 4ºC. The
544 supernatant was then fractionated using a HiTrap Ni^{2+} column (GE Healthcare)
545 following the manufacturer's instructions. Fractions containing RbpB were dialyzed
546 overnight in TGED buffer (10 mM Tris-HCl pH 8, 5% glycerol, 0.1mM EDTA, and DTT
547 0.015 mg/mL) and loaded onto a HiTrap-Q column (GE Healthcare). The column was
548 washed with TGED buffer, and protein was eluted with a linear gradient of NaCl (0.1 M
549 to 1 M) in TGED buffer. The fractions containing the protein were pooled, dialyzed, and
550 concentrated using Centricon 10 concentrators (Millipore-Sigma), mixed with an equal
551 volume of 100% glycerol and stored at -20ºC.

552 **Radiolabeled pentaprobe synthesis**. Twelve pentaprobes containing all the
553 possible 5-nt combinations were prepared based on a published protocol with some
554 modifications (52). All oligonucleotides used in pentaprobe synthesis are listed in Table
555 S1. Single-stranded oligonucleotides for PP1-PP6 were Q5 PCR amplified with a 5' T7
556 promoter for either the Watson strand or the Crick strand, generating twelve dsDNA
557 templates with a single T7 site, two each for PP1-PP6. The Watson strand of PP1
558 dsDNA is identical to the coding strand of the PP1 pentaprobe, and the Crick strand is
559 identical to the coding strand of the PP7 pentaprobe. These dsDNA fragments were
560 then used to produce twelve different ssRNA pentaprobes by *in vitro* transcription from
561 the T7 promoter with a MEGAscript T7 kit (Ambion). Transcribed RNA fragments were
562 5´end-labeled with ATP [γ^{32}P] (PerkinElmer) using the KinaseMax kit (Ambion) following
563 the manufacturer's protocol.

564 **Electrophoretic mobility shift assay and motif prediction.** RNA-protein gel
565 electrophoretic mobility shift assays were performed using 0.01 pmol of ^{32}P-labeled
566 pentaprobe RNA and the indicated amounts of RbpB in binding buffer (10mM Tris-HCl
567 pH 8.0, 0.5 mM DTT, 0.5 mM $MgCl_2$, 10mM KCl, 5mM Na_2HPO_4–NaH_2PO_4 pH 8.0). The
568 mixture was incubated at 37 ºC for 30 minutes, and non-denaturing loading buffer (50%
569 glycerol and 0.1% bromophenol blue) was added. The samples were resolved on a
570 4.6% native polyacrylamide gel for 1.5 hours at 10 mA. The fraction of RbpB bound was
571 determined using a Fluorescent Image Analyzer FLA-3000 (FUJIFILM) to quantify the
572 band intensities. K_D values were calculated using Sigmaplot software based on a
573 published method (90). The MEME program (55) was used to predict conserved motifs
574 for the positive pentaprobe sequences with the following parameters: maximum number
575 of motifs, 10; minimum motif width, 4; and maximum motif width, 8. K_D was calculated
576 for RbpB binding to PP3 and motif 1 using three technical replicates as done previously
577 (56).

578 **RNA sequencing sample prep and processing**

579 For rich media RNA-seq, strains were cultured in 5 ml of TYG in biological
580 triplicate to stationary phase overnight. Each culture was then sub-cultured 1:100 into
581 two 5 ml TYG cultures. One tube was cultured to mid-log (0.35-0.6 OD_{600}) and the
582 second tube was cultured to early stationary phase (1.2-1.4 OD_{600}) as measured in a
583 Thermo Spectronic 200 (Thermo Fisher Scientific brand, referred to as ThermoSpec
584 below). Then 500 µl of cells were then spun down at 7,500 x g for 3 min at room
585 temperature, supernatant removed, and pellets re-suspended in 600 µl of TriReagent

586 (Sigma). RNA was then isolated from the re-suspensions using the Zymo Direct-Zol
587 RNA Mini-Prep kit (Zymo) which includes on-column DNaseI treatment. RNA quality
588 was evaluated using a Qubit 2.0 fluorometer (Invitrogen) and Agilent 2100 Bioanalyzer
589 (UIUC Biotechnology Center). Total RNA was then submitted to the W. M. Keck Center
590 for Comparative and Functional Genomics at UIUC for rRNA depletion, library
591 construction, and sequencing. Briefly, ribosomal RNAs were removed from total RNA
592 with an Illumina Ribo-Zero rRNA Removal Kit for Bacteria. RNA-seq libraries were
593 produced with a ScriptSeq v2 kit (Illumina) and cleaned with AMPure beads (Beckman
594 Coulter) to remove any fragments <80 nt. Libraries were sequenced on an Illumina
595 HiSeq2500 using HiSeq SBS sequencing kit v4 to give 160 nt single-end reads. Results
596 were de-multiplexed with bcl2fastq v2.17.1.14 (Illumina). Reads were quality filtered and
597 trimmed using Bioconductor package ShortRead (91) to first remove reads with >1 N or
598 if ≥75% of a read is a single nucleotide and then the first 2 nucleotides were removed
599 from each sequence read. Sequencing adapters were removed with fastx_clipper
600 (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Residual rRNAs were removed
601 using bowtie2 (92, 93) and final reads mapped to the genome and analyzed using
602 Rockhopper v2.03. All raw Rockhopper calculated expression values were increased by
603 1 and fold changes (FC) calculated as $\log_2$(mutant expression value + 1/wild-type
604 expression value +1). RNA-seq processing statistics are summarized in Dataset S2C.

605 For RNA-seq of cultures grown in minimal media, a single colony/strain was
606 smeared onto half a 100 mm BHI-10%HB agar plate with a cotton swab and cultured for
607 24 hours. Lawns were then re-suspended in 5 ml of minimal medium + glucose (MMG)
608 and spun down at 4000g x 5 min. Cell pellets were washed three times with 1 ml of

609 MMG and then diluted to 0.07 OD_{630} in 200 µl of MMG, as measured on a Biotek
610 Synergy HT plate reader (referred to as Biotek below). Cells were then diluted 1:1000 in
611 25 ml of MMG and cultured overnight to 0.35-0.50 OD_{600} (ThermoSpec). Cells were
612 then spun down and re-suspended in 1 ml of minimal medium without a carbon source
613 per every 5 ml of culture. For each biological replicate, these suspensions were then
614 diluted to 0.1 OD_{600} (ThermoSpec) in 5 ml of MMG, minimal medium + melibiose
615 (MMM), or minimal medium + 0.25% w/v glucose + 0.25% w/v galactose (MMGG).
616 Cultures were then grown to 0.45-0.65 OD_{600} (ThermoSpec), and then stabilized in
617 Qiagen RNA protect. Briefly, 4 ml of culture was combined with 8 ml of RNA protect,
618 vortexed, and then incubated at room temperature for 5 min. Suspensions were then
619 spun at 4,000 x g for 10 minutes at 4ºC. Supernatants were decanted and pellets stored
620 at -80 ºC. Cell pellets were thawed, and RNA prepped using a Qiagen RNeasy Mini Kit.
621 RNA was sequenced and analyzed as done above with the RNA-seq for cells grown in
622 TYG, with the exception that libraries were sequenced on an Illumina HiSeq4000 to
623 produce 150 nt single-end reads.

624 **RNA-seq protein functional and pathway analyses**

625 RNA-seq data from each mutant strain (Δ*rbpA*, Δ*rbpB*, Δ*rbpA*Δ*rbpB*) was
626 compared to wild-type in rich medium at mid-log and stationary phase, yielding six
627 comparisons. All genes with a $\log_2FC \geq +1$ or ≤ -1 with a q-value <0.06 were assigned
628 to functional groups using the following framework. GSEA was used on each of the six
629 differentially expressed gene sets individually to identify enriched functional groups. *B.*
630 *thetaiotaomicron*-specific functional gene sets used included: KEGG pathways, CPS
631 loci, PULs, and corrinoid transport (22). GSEA identified enriched gene sets

632 corresponding to PULs, CPS loci, TCA cycle (KEGG pathway bth00020), corrinoid
633 transport, and microbial metabolism in diverse environments (bth01120). Since these
634 were identified as enriched categories, if one of the significantly differentially regulated
635 genes in Dataset S2A was part of one of these gene sets, it was assigned that as a
636 category identifier with the exception of "microbial metabolism in diverse environments."
637 GSEA-identified enriched genes in this KEGG category were further split into sub-
638 categories including TCA cycle and glyoxylate and dicarboxylate metabolism
639 (bth00630). If genes were associated with both TCA cycle and glyoxylate and
640 dicarboxylate metabolism, they were assigned TCA cycle since this category was
641 identified as enriched by GSEA directly, but glyoxylate and dicarboxylate metabolism
642 was not. Since corrinoid transport was enriched in our differentially expressed genes,
643 any gene associated with B_{12} metabolism that was in our differentially regulated genes
644 was assigned a "B-vitamin metabolism" category. Remaining gene functions were
645 assigned using gene ontology (GO) from QuickGO. Briefly, gene names were used to
646 extract UniprotKB identifiers that were then used to pull GO biological process
647 annotations from QuickGO when available. Only the first reported GO assignment was
648 used for each gene. GO terms were further grouped into custom functional categories to
649 make a more tractably sized list of functional categories for visualization. Genes without
650 any of the above functional assignments were labeled as either "hypothetical protein" if
651 they were annotated as such or "miscellaneous" if the gene had a putative functional
652 annotation that was not captured by the other functional categories. All resulting
653 functional groups are listed in Dataset S2A.

654 **Biolog carbon utilization assays**

655 Biolog carbon utilization assays were conducted according to manufacturer
656 recommendations as follows. Biolog carbon source PM1 and PM2A MicroPlates were
657 brought to room temperature to avoid condensation prior to opening the seals. Plates
658 were then cycled into the anaerobic chamber and maintained in an anaerobic desiccant
659 box for 24 hours. Single colonies of each strain were swabbed onto BHI-10% horse
660 blood plates and cultured overnight to produce a lawn of cells. Cells were aerobically
661 suspended into 5 ml of reduced minimal medium without a carbon source to 40%
662 turbidity (OD_{590}, ThermoSpec) using a cotton swab. Suspensions were cycled into the
663 chamber and 1.5 ml combined with 22 ml of anoxic, reduced minimal medium without a
664 carbon source. Each carbon source plate was then inoculated with 100 µl of diluted
665 cells and statically incubated for 30 minutes at room temperature to facilitate compound
666 dissolution before measuring time point zero. Plates were statically incubated at 37 ºC
667 with manual OD_{630} readings taken every hour in the plate reader for the first 11 hours of
668 growth. Plates were then left in the chamber overnight and optical density readings
669 were resumed after 24 hours of growth. Time points were then taken every 3 hours to a
670 final time point of 36 hours of growth. Linear regression and prediction curves were
671 calculated using Prism. Negative control wells and Xylitol (PM2A) were removed from
672 linear regression calculations. Xylitol was removed due to an unknown occlusion
673 (potentially condensation or precipitation) causing transiently high OD_{630} readings. In
674 the absence of these transient values, *B. thetaiotaomicron* could not grow on Xylitol as
675 a sole carbon source in these experiments.

676 **Minimal media growth assays**

677 Strains were cultured from a colony in 5 ml of TYG for 24 hours and then sub-
678 cultured 1:1000 into 5 ml of MMG for 24 hours. 1 ml of stationary phase MMG cultures
679 were spun down 4000 g x 10 min at room temperature. Supernatants were removed
680 and pellets resuspended in 1 ml of minimal medium without a carbon source. 2 µl of
681 cells were then sub-cultured into 198 µl of minimal media containing carbon sources to
682 appropriate final concentrations in flat bottom, 96 well Corning Costar tissue culture-
683 treated plates (Sigma). Plates were sealed with a Breathe-Easy gas permeable
684 membrane (Sigma) and statically cultured in the Biotek plate reader for 48 hours with
685 optical density recorded every 30 minutes.

686 **qRT-PCR of *rbpB* and *BT1871***

687 Strains were cultured in MMG to stationary phase overnight, sub-cultured 1:100
688 into 5 ml of MMG, and then cultured to mid-log (0.38-0.52 OD_{600}, ThermoSpec). All
689 cultures for strains containing pNBU2_*ermGb* vectors contained erythromycin. Four ml
690 of cells were pelleted at 4,000 g x 10 min, supernatant decanted, and then RNA isolated
691 with a Qiagen RNeasy mini kit. Residual DNA was degraded on-column using Qiagen
692 RNase-Free DNase Set and the RNA cleaned with a Qiagen RNeasy Mini kit. First-
693 strand cDNA synthesis was done with a SuperScript II RT kit (Invitrogen) and random
694 hexamers (Invitrogen). Post reverse transcription, the SuperScript reaction was
695 incubated with 27 µl of 1N NaOH at 65 ºC for 30 min, neutralized with 27 µl of 1N HCl,
696 and cleaned up with a Qiagen MinElute PCR purification kit. cDNA was diluted and
697 *rbpB*, *BT1871*, and 16s rRNA copies amplified using 2x QX200 ddPCR EvaGreen
698 Supermix (Biorad) and quantified using the QX200 Droplet Digital PCR system (Biorad)

699 according to manufacturer instructions. All ddPCR consumables were supplied by
700 Biorad and Rainin (pipette tips only). Relative ratios were calculated by dividing the
701 *rbpB* or *BT1871* counts by the 16s rRNA counts.

702 **Determination of operon structure of *rbpB* in *B. thetaiotaomicron***

703 Strains were cultured in MMG to stationary phase overnight and then sub-
704 cultured 1:1,000 into 4 ml of MMM and 1:10,000 into 4 ml MMG. MMG and MMM
705 cultures were grown to mid-log (OD_{630} 0.25-0.35, Biotek), pelleted at 4,000g x 10 min at
706 4ºC, and supernatant removed. RNA and cDNA were prepped as done for qPCR with
707 the exception that residual DNA was degraded on beads using an Ambion nuclease-
708 free DNAse kit. Overlap end-point PCR was done with KAPA HiFi (KAPA Biosystems).
709 gDNA was prepped using a Qiagen DNeasy Blood and tissue kit.

710 **Data Availability**

711 All RNA-seq datasets corresponding to the samples listed in Dataset S2C are publicly
712 available on NCBI under BioProject accession number PRJNA723047. Primary data
713 can be obtained at https://cluster.hpcc.ucr.edu/~patrickd/RBP_Datasets/. See file
714 "RBP_RNAseq_file_index" for file names corresponding to specific SRA Accession
715 numbers.

716

717

Acknowledgements

We thank Alvaro Hernandez, Chris Wright, and staff of the Roy J. Carver Biotechnology Center for assistance with RNA-seq. We also thank Danielle Campbell for guidance on comparative genomics and Auroni Gupta and Saika Hossain for assistance with EMSAs. We are grateful to Sandy Pernitzch (Scigraphix) for assistance with model graphics. This work was funded by UIUC and the UIUC Department of Microbiology, a Roy J. Carver Charitable Trust award (15-4501) to P.H.D. and initial complement funding to P.H.D. from UCR. A.N.D.A. was funded by an Alice Helm Graduate Research Excellence Fellowship and the UIUC Department of Microbiology.

A.N.D.A, M.S.A, Z.A.C., P.H.D, and C.K.V. designed research and analyzed data. A.N.D.A., M.S.A., Z.A.C., X.M., and P.H.D. performed research. A.N.D.A., M.S.A., P.H.D., and C.K.V. wrote the paper.

732 **References**

733

734 1. Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L,
735 Henrissat B, Knight R, Gordon JI. 2011. Diet drives convergence in gut microbiome
736 functions across mammalian phylogeny and within humans. Science 332:970–974.

737 2. Carmody RN, Gerber GK, Luevano JM, Gatti DM, Somes L, Svenson KL, Turnbaugh
738 PJ. 2015. Diet dominates host genotype in shaping the murine gut microbiota. Cell Host
739 Microbe 17:72–84.

740 3. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M,
741 Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L,
742 Li H, Bushman FD, Lewis JD. 2011. Linking long-term dietary patterns with gut microbial
743 enterotypes. Science 334:105–108.

744 4. Kaoutari AE, Armougom F, Gordon JI, Raoult D, Henrissat B. 2013. The abundance
745 and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev
746 Microbiol 11:497–504.

747 5. Martens EC, Chiang HC, Gordon JI. 2008. Mucosal glycan foraging enhances fitness
748 and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe
749 4:447–457.

750 6. Rogers TE, Pudlo NA, Koropatkin NM, Bell JSK, Balasch MM, Jasker K, Martens EC.
751 2013. Dynamic responses of *Bacteroides thetaiotaomicron* during growth on glycan
752 mixtures. Mol Microbiol 88:876–890.

753 7. Terrapon N, Lombard V, Drula É, Lapébie P, Al-Masaudi S, Gilbert HJ, Henrissat B.
754 2017. PULDB: the expanded database of Polysaccharide Utilization Loci. Nucleic Acids
755 Res 46:D677–D683.

756 8. Pereira GV, Abdel-Hamid AM, Dutta S, D'Alessandro-Gabazza CN, Wefers D, Farris
757 JA, Bajaj S, Wawrzak Z, Atomi H, Mackie RI, Gabazza EC, Shukla D, Koropatkin NM,
758 Cann I. 2021. Degradation of complex arabinoxylans by human colonic Bacteroidetes.
759 Nat Commun 12:459.

760 9. Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW,
761 Henrissat B, Gilbert HJ, Bolam DN, Gordon JI. 2011. Recognition and degradation of
762 plant cell wall polysaccharides by two human gut symbionts. PLoS Biol 9:e1001221.

763 10. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV,
764 Gordon JI. 2003. A genomic view of the human-*Bacteroides thetaiotaomicron*
765 symbiosis. Science 299:2074–2076.

766 11. Xu J, Chiang HC, Bjursell MK, Gordon JI. 2004. Message from a human gut
767 symbiont: sensitivity is a prerequisite for sharing. Trends Microbiol 12:21–28.

768 12. Cho KH, Cho D, Wang GR, Salyers AA. 2001. New regulatory gene that contributes
769 to control of *Bacteroides thetaiotaomicron* starch utilization genes. J Bacteriol
770 183:7198–7205.

771 13. Schwalm ND, Townsend GE, Groisman EA. 2016. Multiple signals govern utilization
772 of a polysaccharide in the gut bacterium *Bacteroides thetaiotaomicron*. mBio 7:e01342-
773 16.

774 14. Chang C, Tesar C, Li X, Kim Y, Rodionov DA, Joachimiak A. 2015. A novel
775 transcriptional regulator of L-arabinose utilization in human gut bacteria. Nucleic Acids
776 Res 43:10546–10559.

777 15. Ravcheev DA, Godzik A, Osterman AL, Rodionov DA. 2013. Polysaccharides
778 utilization in human gut bacterium *Bacteroides thetaiotaomicron*: comparative genomics
779 reconstruction of metabolic and regulatory networks. BMC Genomics 14:873.

780 16. Sonnenburg ED, Zheng H, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN,
781 Sonnenburg JL. 2010. Specificity of polysaccharide use in intestinal *Bacteroides*
782 species determines diet-induced microbiota alterations. Cell 141:1241–1252.

783 17. Martens EC, Roth R, Heuser JE, Gordon JI. 2009. Coordinate regulation of glycan
784 degradation and polysaccharide capsule biosynthesis by a prominent human gut
785 symbiont. J Biol Chem 284:18445–18457.

786 18. Pudlo NA, Urs K, Kumar SS, German JB, Mills DA, Martens EC. 2015. Symbiotic
787 human gut bacteria with variable metabolic priorities for host mucosal glycans. mBio
788 6:e01282-15.

789 19. Cao Y, Förstner KU, Vogel J, Smith CJ. 2016. *cis*-encoded small RNAs, a
790 conserved mechanism for repression of polysaccharide utilization in *Bacteroides*. J
791 Bacteriol 198:2410–2418.

792 20. Townsend GE, Han W, Schwalm ND, Hong X, Bencivenga-Barry NA, Goodman AL,
793 Groisman EA. 2020. A master regulator of *Bacteroides thetaiotaomicron* gut
794 colonization controls carbohydrate utilization and an alternative protein synthesis factor.
795 mBio 11:03221–19.

796 21. Ryan D, Jenniches L, Reichardt S, Barquist L, Westermann AJ. 2020. A high-
797 resolution transcriptome map identifies small RNA regulation of metabolism in the gut
798 microbe *Bacteroides thetaiotaomicron*. Nat Commun 11:3557.

799 22. Degnan PH, Barry NA, Mok KC, Taga ME, Goodman AL. 2014. Human gut
800 microbes use multiple transporters to distinguish vitamin B_{12} analogs and compete in
801 the gut. Cell Host Microbe 15:47–57.

802 23. Costliow ZA, Degnan PH. 2017. Thiamine acquisition strategies impact metabolism
803 and competition in the gut microbe *Bacteroides thetaiotaomicron*. mSystems 2:e00116-
804 17.

805 24. Costliow ZA, Degnan PH, Vanderpool CK. 2019. Thiamine pyrophosphate
806 riboswitches in *Bacteroides* species regulate transcription or translation of thiamine
807 transport and biosynthesis genes. bioRxiv 867226.

808 25. Papenfort K, Vogel J. 2014. Small RNA functions in carbon metabolism and
809 virulence of enteric pathogens. Front Cell Infect Microbiol 4:1–12.

810 26. Smirnov A, Schneider C, Hör J, Vogel J. 2017. Discovery of new RNA classes and
811 global RNA-binding proteins. Curr Opin Microbiol 39:1–9.

812 27. Hör J, Gorski SA, Vogel J. 2018. Bacterial RNA biology on a genome scale. Mol Cell
813 70:785–799.

814 28. Vogel J, Luisi BF. 2011. Hfq and its constellation of RNA. Nat Rev Microbiol 9:578–
815 589.

816 29. Santiago-Frangos A, Woodson SA. 2018. Hfq chaperone brings speed dating to
817 bacterial sRNA. Wiley Interdiscip Rev RNA 9:e1475.

818 30. Iosub IA, Nues RW van, McKellar SW, Nieken KJ, Marchioretto M, Sy B, Tree JJ,
819 Viero G, Granneman S. 2020. Hfq CLASH uncovers sRNA-target interaction networks
820 linked to nutrient availability adaptation. eLife 9:e54655.

821 31. Melamed S, Peer A, Faigenbaum-Romm R, Gatt YE, Reiss N, Bar A, Altuvia Y,
822 Argaman L, Margalit H. 2016. Global mapping of small RNA-target interactions in
823 bacteria. Molecular Cell 63:884–897.

824 32. Chao Y, Vogel J. 2010. The role of Hfq in bacterial pathogens. Curr Opin Microbiol
825 13:24–33.

826 33. Richards GR, Vanderpool CK. 2011. Molecular call and response: the physiology of
827 bacterial small RNAs. Biochim Biophys Acta 1809:525–531.

828 34. Bobrovskyy M, Vanderpool CK. 2014. The small RNA SgrS: roles in metabolism and
829 pathogenesis of enteric bacteria. Front Cell Infect Microbiol 4:1–8.

830 35. Waters JL, Salyers AA. 2012. The small RNA RteR inhibits transfer of the
831 *Bacteroides conjugative* transposon CTnDOT. J Bacteriol 194:5228–5236.

832 36. Townsend GE, Han W, Schwalm ND, Raghavan V, Barry NA, Goodman AL,
833 Groisman EA. 2019. Dietary sugar silences a colonization factor in a mammalian gut
834 symbiont. Proc Natl Acad Sci U S A 116:233–238.

835 37. Methé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, Gevers D,
836 Petrosino JF, Abubucker S, Badger JH, Chinwalla AT, Earl AM, FitzGerald MG, Fulton
837 RS, Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny
838 DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK,
839 Zeng Q, Aagaard KM, Abolude OO, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL,
840 Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T, Baker CC,
841 Begg L, Belachew T, Bhonagiri V, Bihan M, Blaser MJ, Bloom T, Bonazzi VR, Brooks P,
842 Buck GA, Buhay CJ, Busam DA, Campbell JL, Canon SR, Cantarel BL, Chain PS,
843 Chen I-MA, Chen L, Chhibba S, Chu K, Ciulla DM, Clemente JC, Clifton SW, Conlan S,
844 Crabtree J, Cutting MA, Davidovics NJ, Davis CC, DeSantis TZ, Deal C, Delehaunty

845 KD, Dewhirst FE, Deych E, Ding Y, Dooling DJ, Dugan SP, Dunne WM, Durkin AS,
846 Edgar RC, Erlich RL, Farmer CN, Farrell RM, Faust K, Feldgarden M, Felix VM, Fisher
847 S, Fodor AA, Forney L, Foster L, Francesco VD, Friedman J, Friedrich DC, Fronick CC,
848 Fulton LL, Gao H, Garcia N, Giannoukos G, Giblin C, Giovanni MY, Goldberg JM, Goll
849 J, Gonzalez A, Griggs A, Gujja S, Haas BJ, Hamilton HA, Harris EL, Hepburn TA,
850 Herter B, Hoffmann DE, Holder ME, Howarth C, Huang KH, Huse SM, Izard J, Jansson
851 JK, Jiang H, Jordan C, Joshi V, Katancik JA, Keitel WA, Kelley ST, Kells C, Kinder-
852 Haake S, King NB, Knight R, Knights D, Kong HH, Koren O, Koren S, Kota KC, Kovar
853 CL, Kyrpides NC, Rosa PSL, Lee SL, Lemon KP, Lennon N, Lewis CM, Lewis L, Ley
854 RE, Li K, Liolios K, Liu B, Liu Y, Lo C-C, Lozupone CA, Lunsford RD, Madden T,
855 Mahurkar AA, Mannon PJ, Mardis ER, Markowitz VM, Mavrommatis K, McCorrison JM,
856 McDonald D, McEwen J, McGuire AL, McInnes P, Mehta T, Mihindukulasuriya KA,
857 Miller JR, Minx PJ, Newsham I, Nusbaum C, O'Laughlin M, Orvis J, Pagani I,
858 Palaniappan K, Patel SM, Pearson M, Peterson J, Podar M, Pohl C, Pollard KS, Priest
859 ME, Proctor LM, Qin X, Raes J, Ravel J, Reid JG, Rho M, Rhodes R, Riehle KP, Rivera
860 MC, Rodriguez-Mueller B, Rogers Y-H, Ross MC, Russ C, Sanka RK, Sankar P,
861 Sathirapongsasuti JF, Schloss JA, Schloss PD, Schmidt TM, Scholz M, Schriml L,
862 Schubert AM, Segata N, Segre JA, Shannon WD, Sharp RR, Sharpton TJ, Shenoy N,
863 Sheth NU, Simone GA, Singh I, Smillie CS, Sobel JD, Sommer DD, Spicer P, Sutton
864 GG, Sykes SM, Tabbaa DG, Thiagarajan M, Tomlinson CM, Torralba M, Treangen TJ,
865 Truty RM, Vishnivetskaya TA, Walker J, Wang L, Wang Z, Ward DV, Warren W,
866 Watson MA, Wellington C, Wetterstrand KA, White JR, Wilczek-Boney K, Wu YQ, Wylie
867 KM, Wylie T, Yandava C, Ye L, Ye Y, Yooseph S, Youmans BP, Zhang L, Zhou Y, Zhu

868 Y, Zoloth L, Zucker JD, Birren BW, Gibbs RA, Highlander SK, Weinstock GM, Wilson
869 RK, White O. 2012. A framework for human microbiome research. Nature 486:215–221.

870 38. Maris C, Dominguez C, Allain FHT. 2005. The RNA recognition motif, a plastic RNA-
871 binding platform to regulate post-transcriptional gene expression. FEBS J 272:2118–
872 2131.

873 39. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M,
874 Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan
875 D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. Nucleic
876 Acids Res 47:D427–D432.

877 40. Herren CD, Rocha ER, Smith CJ. 2003. Genetic analysis of an important oxidative
878 stress locus in the anaerobe *Bacteroides fragilis*. Gene 316:167–175.

879 41. Voss JE, Luisi BF, Hardwick SW. 2014. Molecular recognition of RhlB and RNase D
880 in the *Caulobacter crescentus* RNA degradosome. Nucleic Acids Res 42:13294–13305.

881 42. Al-Husini N, Tomares DT, Bitar O, Childers WS, Schrader JM. 2018. α-
882 Proteobacterial RNA degradosomes assemble liquid-liquid phase-separated RNP
883 bodies. Mol Cell 71:1027–1039.

884 43. Santiago-Frangos A, Kavita K, Schu DJ, Gottesman S, Woodson SA. 2016. C-
885 terminal domain of the RNA chaperone Hfq drives sRNA competition and release of
886 target RNA. Proc Natl Acad Sci U S A 113:E6089–E6096.

887 44. Večerek B, Rajkowitsch L, Sonnleitner E, Schroeder R, Bläsi U. 2007. The C-
888 terminal domain of *Escherichia coli* Hfq is required for regulation. Nucleic Acids Res
889 36:133–143.

890 45. Santiago-Frangos A, Jeliazkov JR, Gray JJ, Woodson SA. 2017. Acidic C-terminal
891 domains autoregulate the RNA chaperone Hfq. eLife 6:e27049.

892 46. Sobrero P, Valverde C. 2012. The bacterial protein Hfq: much more than a mere
893 RNA-binding factor. Crit Rev Microbiol 38:276–299.

894 47. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, Pudlo
895 NA, Kitamoto S, Terrapon N, Muller A, Young VB, Henrissat B, Wilmes P, Stappenbeck
896 TS, Núñez G, Martens EC. 2016. A dietary fiber-deprived gut microbiota degrades the
897 colonic mucus barrier and enhances pathogen susceptibility. Cell 167:1339–1353.

898 48. Wu M, McNulty NP, Rodionov DA, Khoroshkin MS, Griffin NW, Cheng J, Latreille P,
899 Kerstetter RA, Terrapon N, Henrissat B, Osterman AL, Gordon JI. 2015. Genetic
900 determinants of *in vivo* fitness and diet responsiveness in multiple human gut
901 *Bacteroides*. Science 350:aac5992.

902 49. Huang Y, Sheth RU, Kaufman A, Wang HH. 2019. Scalable and cost-effective
903 ribonuclease-based rRNA depletion for transcriptomics. Nucleic Acids Res 48:e20.

904 50. Despres J, Forano E, Lepercq P, Comtet-Marre S, Jubelin G, Chambon C, Yeoman
905 CJ, Miller MEB, Fields CJ, Martens E, Terrapon N, Henrissat B, White BA, Mosoni P.
906 2016. Xylan degradation by the human gut *Bacteroides xylanisolvens* XB1AT involves

907 two distinct gene clusters that are linked at the transcriptional level. BMC Genomics
908 17:1–14.

909 51. Schofield WB, Zimmermann-Kogadeeva M, Zimmermann M, Barry NA, Goodman
910 AL. 2018. The stringent response determines the ability of a commensal bacterium to
911 survive starvation and to persist in the gut. Cell Host Microbe 24:120–132.

912 52. Bendak K, Loughlin FE, Cheung V, O'Connell MR, Crossley M, Mackay JP. 2012. A
913 rapid method for assessing the RNA-binding potential of a protein. Nucleic Acids Res
914 40:e105.

915 53. Teplova M, Farazi TA, Tuschl T, Patel DJ. 2016. Structural basis underlying CAC
916 RNA recognition by the RRM domain of dimeric RNA-binding protein RBPMS. Q Rev
917 Biophys 49:e1.

918 54. Afroz T, Cienikova Z, Cléry A, Allain FHT. 2015. One, two, three, four! How multiple
919 RRMs read the genome sequence. Methods Enzymol 558:235–278.

920 55. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble
921 WS. 2009. MEME Suite: tools for motif discovery and searching. Nucleic Acids Res
922 37:W202–W208.

923 56. Azam MS, Vanderpool CK. 2017. Translational regulation by bacterial small RNAs
924 via an unusual Hfq-dependent mechanism. Nucleic Acids Res 46:2585–2599.

925 57. Stein EM, Kwiatkowska J, Basczok MM, Gravel CM, Berry KE, Olejniczak M. 2020.
926 Determinants of RNA recognition by the FinO domain of the *Escherichia coli* ProQ
927 protein. Nucleic Acids Res 48:7502–7519.

928 58. Nithin C, Mukherjee S, Bahadur RP. 2019. A structure-based model for the
929 prediction of protein-RNA binding affinity. RNA 25:1628–1645.

930 59. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillete MA,
931 Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set
932 enrichment analysis: a knowledge-based approach for interpreting genome-wide
933 expression profiles 102:15545–15550.

934 60. Krinos CM, Coyne MJ, Weinacht KG, Tzianabos AO, Kasper DL, Comstock LE.
935 2001. Extensive surface diversity of a commensal microorganism by multiple DNA
936 inversions. Nature 414:555–558.

937 61. Gloster TM, Turkenburg JP, Potts JR, Henrissat B, Davies GJ. 2008. Divergence of
938 catalytic mechanism within a glycosidase family provides insight into evolution of
939 carbohydrate metabolism by human gut flora. Chem Biol 15:1058–1067.

940 62. Liu H, Shiver AL, Price MN, Carlson HK, Trotter VV, Chen Y, Escalante V, Ray J,
941 Hern KE, Petzold CJ, Turnbaugh PJ, Huang KC, Arkin AP, Deutschbauer AM. 2021.
942 Functional genetics of human gut commensal *Bacteroides thetaiotaomicron* reveals
943 metabolic requirements for growth across environments. Cell Rep 34:108789.

944 63. Faith JJ, McNulty NP, Rey FE, Gordon JI. 2011. Predicting a human gut
945 microbiota's response to diet in gnotobiotic mice. Science 333:101–104.

946 64. Sonnenburg JL, Xu J, Leip DD, Chen CH, Westover BP, Weatherford J, Buhler JD,
947 Gordon JI. 2005. Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont.
948 Science 307:1955–1959.

949 65. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. 2009. The effect
950 of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic
951 mice. Sci Transl Med 1:6ra14.

952 66. Görke B, Stülke J. 2008. Carbon catabolite repression in bacteria: many ways to
953 make the most out of nutrients. Nat Rev Microbiol 6:613–624.

954 67. Hylemon PB, Phibbs PV. 1974. Evidence against the presence of cyclic AMP and
955 related enzymes in selected strains of *Bacteroides fragilis*. Biochem Biophys Res
956 Commun 60:88–95.

957 68. Barabote RD, Saier MH. 2005. Comparative genomic analyses of the bacterial
958 phosphotransferase system. Microbiol Mol Biol Rev 69:608–634.

959 69. Brigham CJ, Malamy MH. 2005. Characterization of the RokA and HexA broad-
960 substrate-specificity hexokinases from *Bacteroides fragilis* and their role in hexose and
961 *N*-acetylglucosamine utilization. J Bacteriol 187:890–901.

962 70. Holmqvist E, Berggren S, Rizvanovic A. 2020. RNA-binding activity and regulatory
963 functions of the emerging sRNA-binding protein ProQ. Biochim Biophys Acta Gene
964 Regul Mech 1863:194596.

965 71. Kavita K, Mets F de, Gottesman S. 2018. New aspects of RNA-based regulation by
966 Hfq and its partner sRNAs. Curr Opin Microbiol 42:53–61.

967 72. Olejniczak M, Storz G. 2017. ProQ/FinO-domain proteins: another ubiquitous family
968 of RNA matchmakers? Mol Microbiol 104:905–915.

969 73. Updegrove TB, Zhang A, Storz G. 2016. Hfq: the flexible RNA matchmaker. Curr
970 Opin Microbiol 30:133–138.

971 74. Coyne MJ, Reinap B, Lee MM, Comstock LE. 2005. Human symbionts use a host-
972 like pathway for surface fucosylation. Science 307:1778–1781.

973 75. Porter NT, Canales P, Peterson DA, Martens EC. 2017. A subset of polysaccharide
974 capsules in the human symbiont *Bacteroides thetaiotaomicron* promote increased
975 competitive fitness in the mouse gut. Cell Host Microbe 22:494–506.

976 76. Bjursell MK, Martens EC, Gordon JI. 2006. Functional genomic and metabolic
977 studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides*
978 *thetaiotaomicron*, to the suckling period. J Biol Chem 281:36269–36279.

979 77. Peterson DA, McNulty NP, Guruge JL, Gordon JI. 2007. IgA response to symbiotic
980 bacteria as a mediator of gut homeostasis. Cell Host Microbe 2:328–339.

981 78. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R,
982 Gordon JI. 2009. Identifying genetic determinants needed to establish a human gut
983 symbiont in its habitat. Cell Host Microbe 6:279–289.

984 79. Holdeman L, Cato E, Moore W. 1977. Anaerobe laboratory manual, 4th Edition.
985 Anaerobe Laboratory, Virginia Polytechnic Institute and State University, Blacksburg,
986 VA.

987 80. Koropatkin NM, Martens EC, Gordon JI, Smith TJ. 2008. Starch catabolism by a
988 prominent human gut symbiont Is directed by the recognition of amylose helices.
989 Structure 16:1105–1115.

990 81. García-Bayona L, Comstock LE. 2019. Streamlined genetic manipulation of diverse
991 *Bacteroides* and *Parabacteroides* isolates from the human gut microbiota. mBio
992 10:e01762-19.

993 82. Wang J, Shoemaker NB, Wang GR, Salyers AA. 2000. Characterization of a
994 *Bacteroides* mobilizable transposon, NBU2, which carries a functional lincomycin
995 resistance gene. J Bacteriol 182:3559–3571.

996 83. Brown JR, Douady CJ, Italia MJ, Marshal WE, Stanhope MJ. 2001. Universal trees
997 based on large combined protein sequence data sets. Nat Genet 28:281–285.

998 84. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and
999 high throughput. Nucleic Acids Res 32:1792–1797.

1000 85. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-
1001 analysis of large phylogenies. Bioinformatics 30:1312–1313.

1002 86. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
1003 sequence data. Bioinformatics 30:2114–2120.

1004 87. McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco CA,
1005 Vanderpool CK, Tjaden B. 2013. Computational analysis of bacterial RNA-Seq data.
1006 Nucleic Acids Res 41:e140.

1007 88. Tjaden B. 2015. *De novo* assembly of bacterial transcriptomes from RNA-seq data.
1008 Genome Biol 16:1–10.

1009 89. 1989. JMP® Version 15. SAS Institute Inc., Cary, NC.

1010 90. Ryder SP, Recht MI, Williamson JR. 2008. Quantitative analysis of protein–RNA
1011 interactions by gel mobility shift. Methods Mol Biol 488:99–115.

1012 91. Morgan M, Anders S, Lawrence M, Aboyoun P, Pagès H, Gentleman R. 2009.
1013 ShortRead: a bioconductor package for input, quality assessment and exploration of
1014 high-throughput sequence data. Bioinformatics 25:2607–2608.

1015 92. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat
1016 Methods 9:357–359.

1017 93. Langmead B, Wilks C, Antonescu V, Charles R. 2018. Scaling read aligners to
1018 hundreds of threads on general-purpose processors. Bioinformatics 35:421–432.

1019

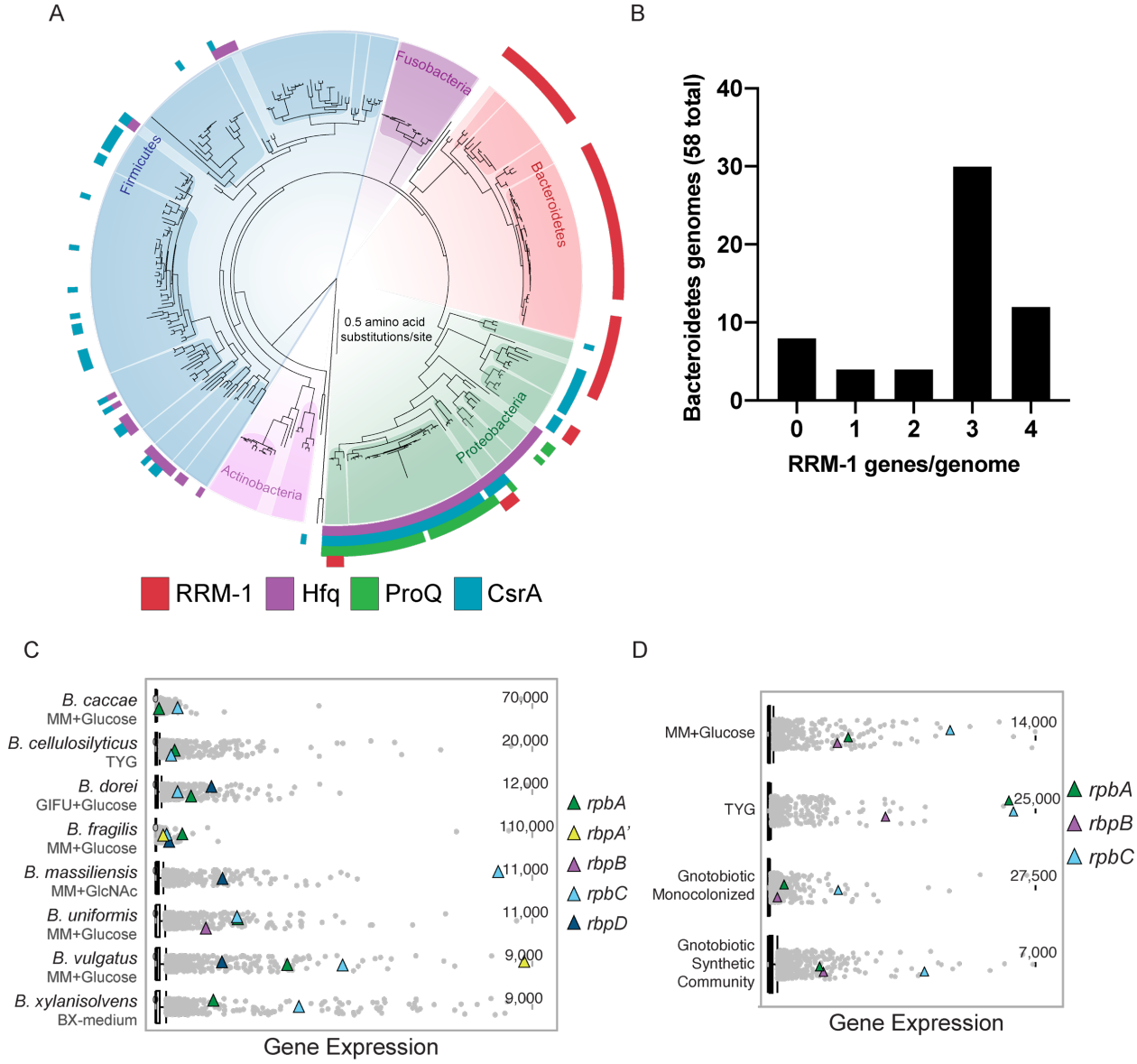

Figure 1: RRM-1 is a conserved, abundantly expressed RNA-binding domain in gut bacteria. (A) Maximum likelihood phylogenetic species tree of 313 human gut-associated microbial genomes. Colored bars indicate presence of at least one copy of the indicated RNA regulatory proteins in a given genome. (B) Histogram of total RRM-1 genes per genome in the 58 Bacteroidetes genomes represented in (A). (C-D) RNA-seq expression plots of all genes in publicly available transcriptomes for various *Bacteroides* species (C) or *B. thetaiotaomicron* only in various growth conditions (D). Grey dots represent a single gene and triangles *rbp* genes. The top 10% of expressed genes lie above the whiskers.

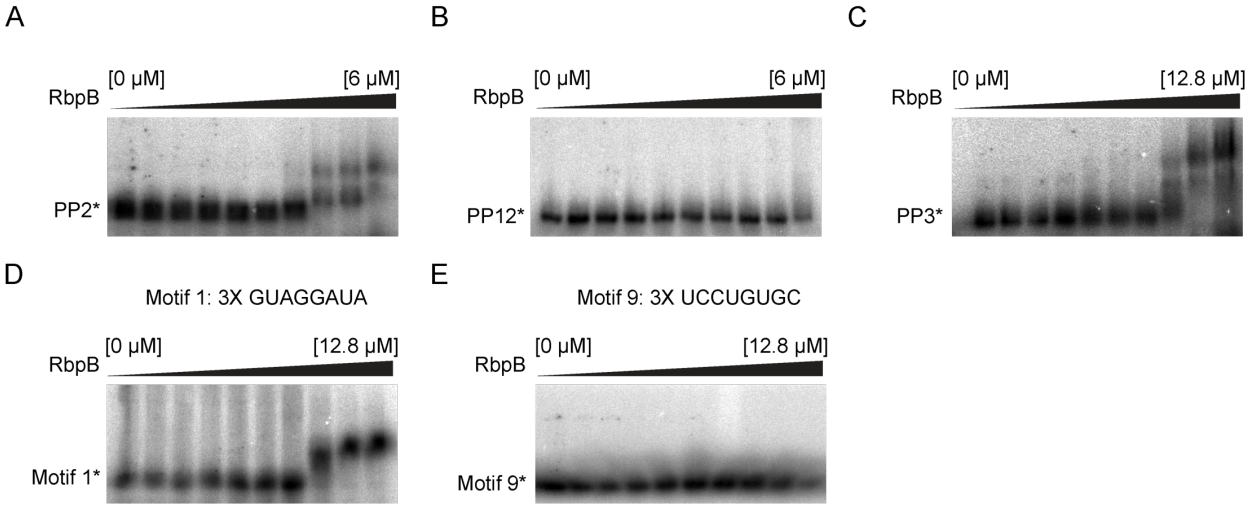

Figure 2: RbpB is a ssRNA-binding protein. (A-B) RbpB-pentaprobe EMSAs for pentaprobes 2 (A) and 12 (B) (PP2 and PP12, respectively). RbpB [μM] increases from left to right as follows: 0, 0.02, 0.05, 0.09, 0.19, 0.38, 0.75, 1.50, 3.00, 6.00. (C) Pentaprobe 3 repeat EMSA with RbpB [μM] increasing from left to right in the gel as follows: 0, 0.05, 0.10, 0.20, 0.40, 0.80, 1.60, 3.20, 6.40, 12.80. (D-E) EMSAs of a 3x repeat of MEME motif 1 (D) or motif 9 (E) with RbpB increasing from left to right as in (C). Asterisk * indicates the unbound radiolabeled pentaprobe.

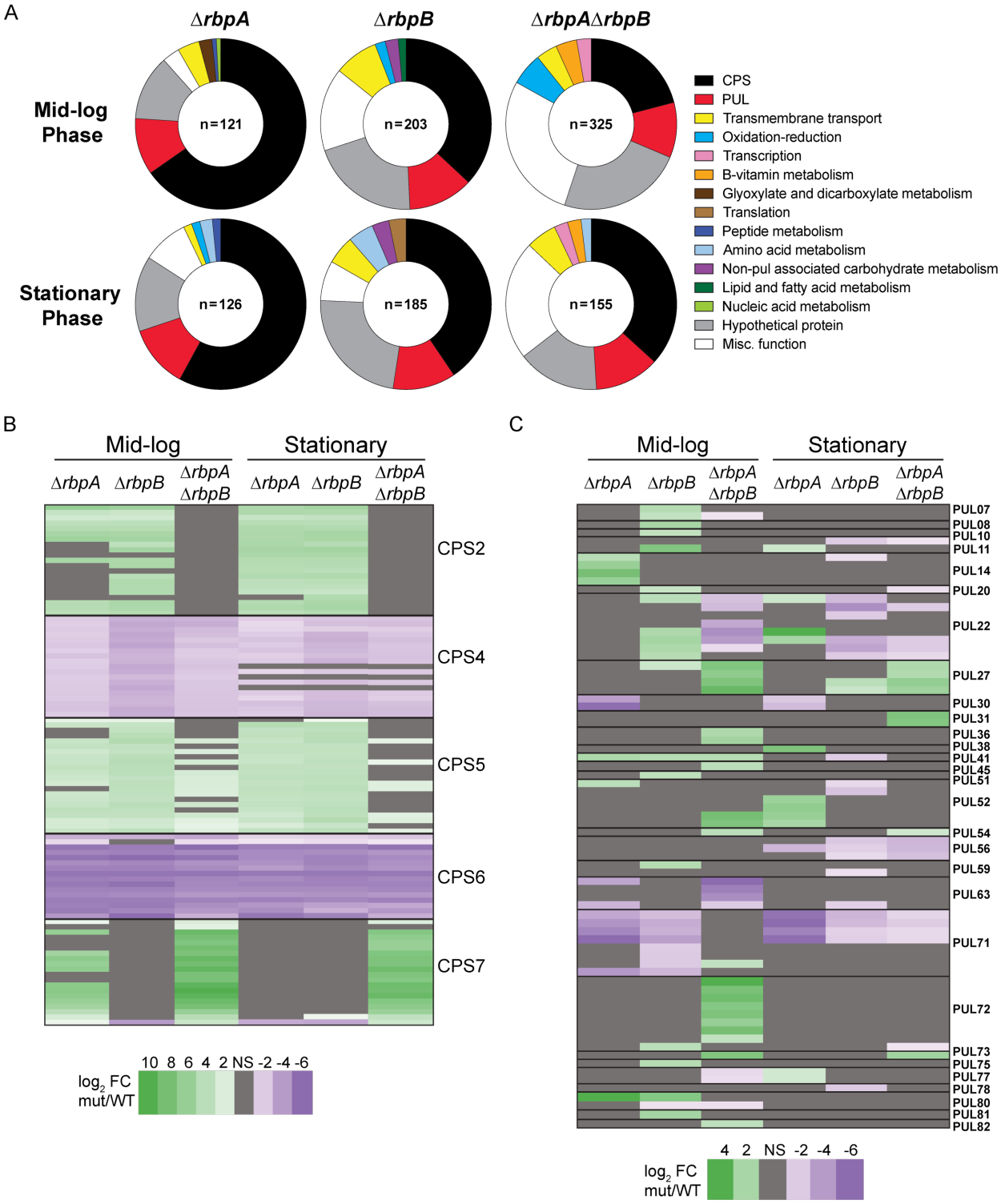

Figure 3: Loss of RBPs leads to altered expression of PULs and CPS loci. (A) Functional categories enriched in differentially regulated genes in rich media RNA-sequencing (genes with a $\log_2 FC \geq +1$ or ≤ -1 with a q-value <0.06). (B-C) Differentially regulated genes for PULs and CPS loci. Only genes that were significantly differentially regulated are shown. Gene names listed in Dataset S2A.

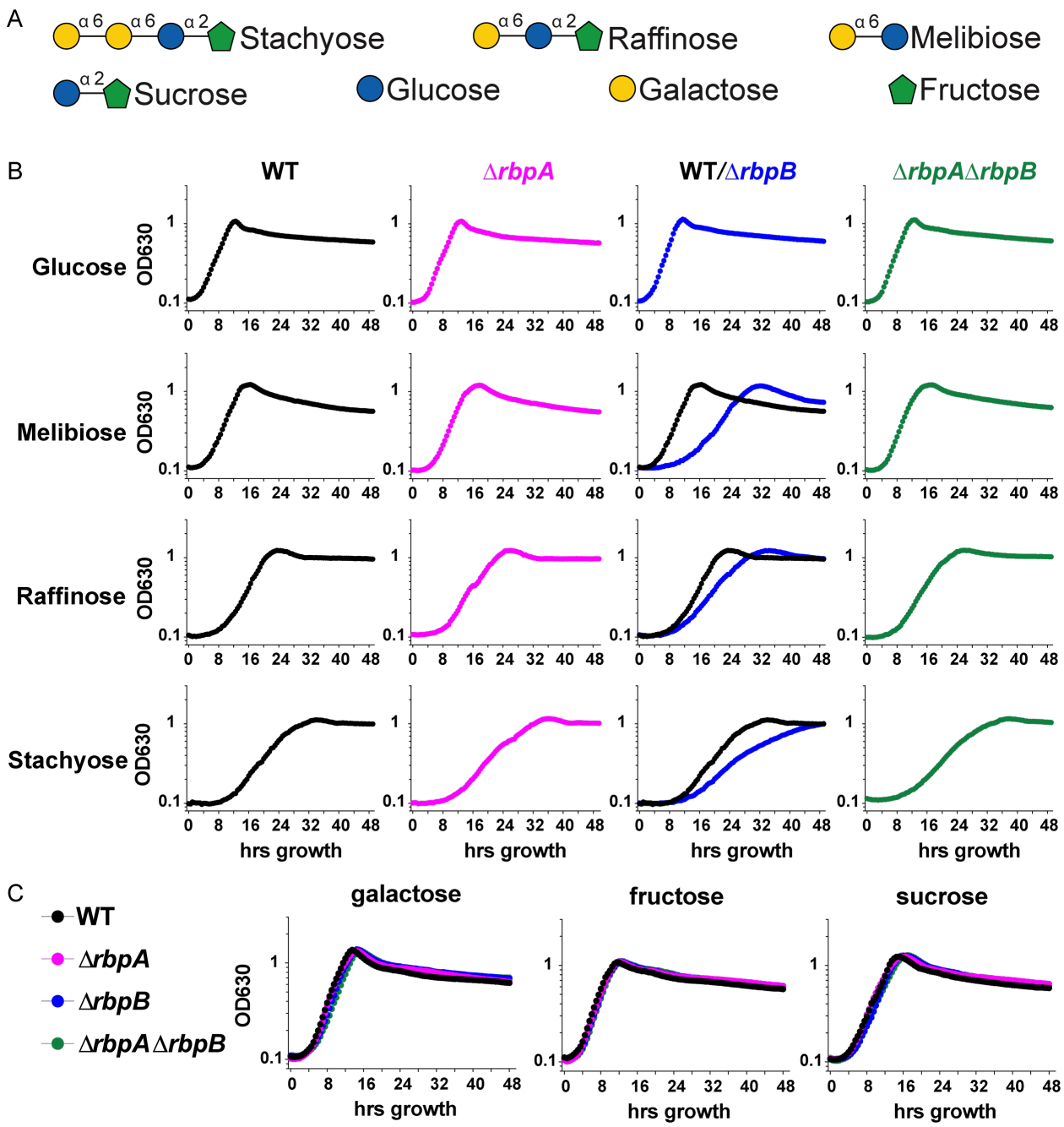

Figure 4: Δ*rbpB* is defective for growth on raffinose family oligosaccharides. (A) Raffinose family oligosaccharides and their subunits. (B-C) Representative normalized growth curves in minimal media from a single biological replicate (n=3) for wild-type (WT) and *rbp* mutants with optical densities (OD_{630}) recorded every 30 minutes.

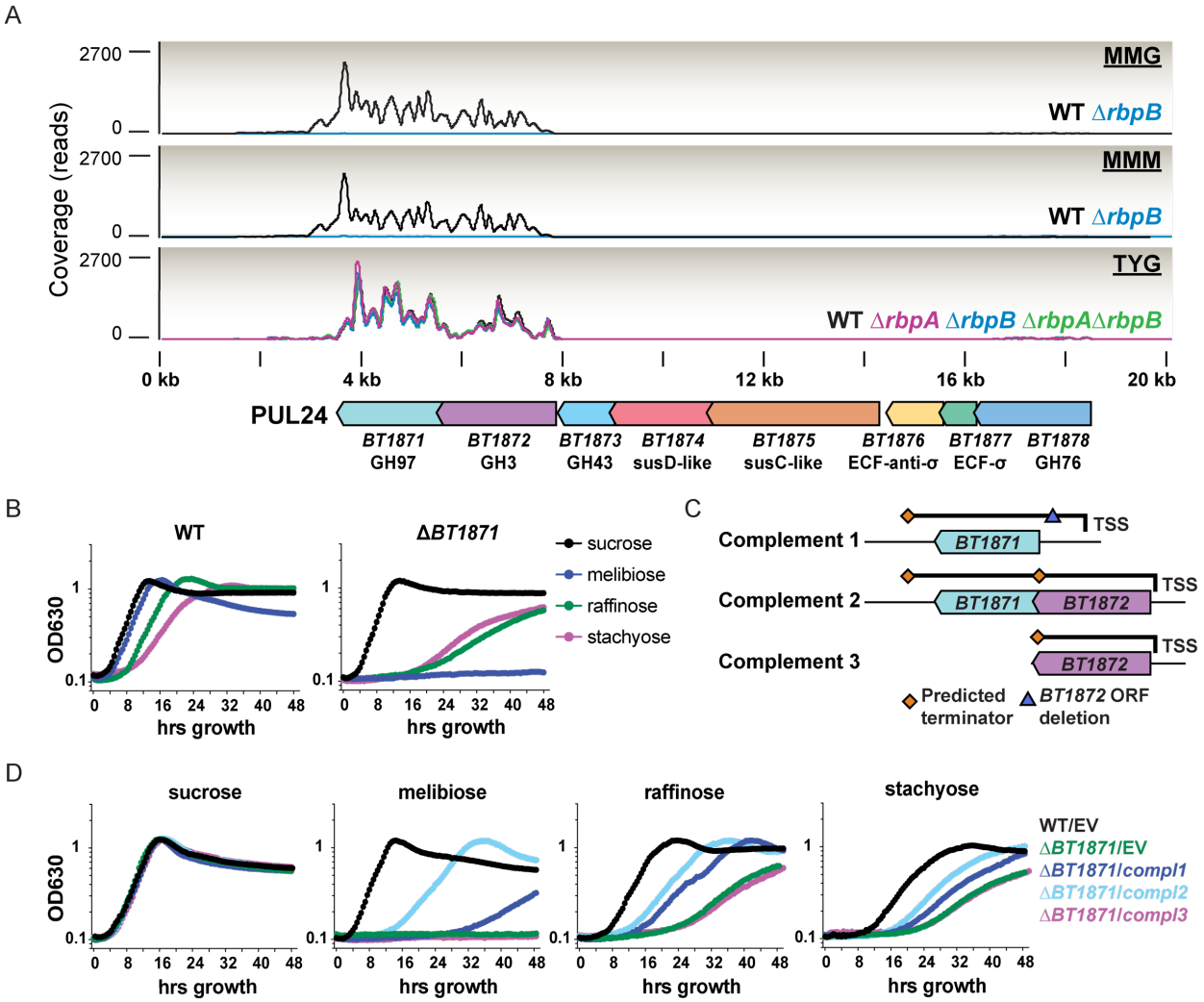

Figure 5: Loss of *rbpB* leads to loss of expression of an essential melibiase in PUL24. (A) Normalized expression coverage curves for mid-log phase cultures in PUL24 with putative gene annotations. (B and D) Representative growth curves in minimal media from a single biological replicate (n=3) with optical densities (OD_{630}) recorded every 30 minutes. (C) Genomic regions inserted into pNBU2 vectors for complementation of *BT1871-BT1872* shown in (D). (D) Empty vector (EV) controls contain integrated pNBU2 without an insert.

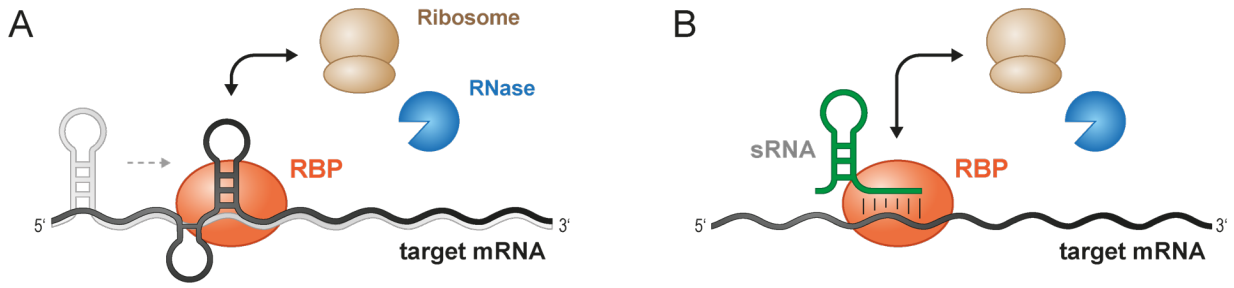

Figure 6: Model for possible mechanisms of RBP-mediated regulation
(A) RBP binding directly to mRNAs could promote structural changes (represented by the transition from light gray to dark gray conformation) that alters access of ribosomes or RNases to change translation or mRNA stability. (B) RBPs could facilitate sRNA binding to mRNA targets and alter access of ribosomes or RNases to change translation or mRNA stability.