# Human Milk Oligosaccharide Utilization in Intestinal Bifidobacteria is Governed by a Global Transcriptional Regulator NagR

**Authors:** Aleksandr A. Arzamasov[a]#, Aruto Nakajima[b], Mikiyasu Sakanaka[b], Miriam N. Ojima[b], Takane Katayama[b], Dmitry A. Rodionov[a], Andrei L. Osterman[a]#

**Affiliations:**

[a]Infectious and Inflammatory Diseases Center, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California, USA

[b]Graduate School of Biostudies, Kyoto University, Kyoto, Japan

Running Head: Regulation of HMO utilization in B. infantis

#Adress correspondence to Aleksandr A. Arzamasov, arzamasov.alexander@gmail.com and Andrei L. Osterman, osterman@sbpdiscovery.org.

Word count for the abstract:

Word count for the text:

**Keywords**

bifidobacteria, HMO, regulon, comparative genomics, evolution, prebiotics, transcription factor, carbohydrate metabolism

**Abstract**

*Bifidobacterium longum* subsp. *infantis* (*B. infantis*) is a prevalent beneficial bacterium that colonizes the human neonatal gut and is uniquely adapted to efficiently use human milk oligosaccharides (HMOs) as a carbon and energy source. Multiple studies have focused on characterizing the elements of HMO utilization machinery in *B. infantis*; however, the regulatory mechanisms governing the expression of these catabolic pathways remain poorly understood. A bioinformatic regulon reconstruction approach used in this study implicated NagR, a transcription factor from the ROK family, as a negative global regulator of genomic loci encoding lacto-*N*-biose/galacto-*N*-biose (LNB/GNB), lacto-*N*-tetraose (LNT), and lacto-*N*-neotetraose (LNnT) utilization pathways in *B. infantis.* This conjecture was corroborated by transcriptome profiling upon *nagR* genetic inactivation and experimental assessment of binding of recombinant NagR to predicted DNA operators. The latter approach also implicated *N*-acetylglucosamine (GlcNAc), a universal intermediate of LNT and LNnT catabolism, and its phosphorylated derivatives as plausible NagR transcriptional effectors. Reconstruction of NagR regulons in various *Bifidobacterium* lineages revealed multiple regulon expansion events, suggesting evolution from a local regulator of GlcNAc catabolism in ancestral bifidobacteria to a global regulator controlling foraging of mixtures of GlcNAc-containing host-derived glycans in mammalian gut-colonizing *B. infantis* and *Bifidobacterium bifidum.*

**Importance**

The predominance of bifidobacteria in the gut of breastfed infants is attributed to the ability of these bacteria to utilize human milk oligosaccharides (HMOs). Thus, individual HMOs such as lacto-*N*-tetraose (LNT) and lacto-*N*-neotetraose (LNnT) are considered promising prebiotics that would stimulate the growth of bifidobacteria and confer multiple health benefits to preterm and malnourished children suffering from impaired (stunted) gut microbiota development. However, the rational selection of HMO-based prebiotics is hampered by the incomplete knowledge of regulatory mechanisms governing HMO utilization in target bifidobacteria. This study describes NagR-mediated transcriptional regulation of LNT and LNnT utilization in *Bifidobacterium longum* subsp. *infantis*. The elucidated regulatory network appears optimally adapted to simultaneous utilization of multiple HMOs, providing a rationale to add HMO mixtures (rather than individual components) into infant formulas. The study also provides insights into the

53 evolutionary trajectories of complex regulatory networks controlling carbohydrate metabolism in
54 bifidobacteria.
55
56 **Introduction**
57 Bifidobacteria are Gram-positive, anaerobic, saccharolytic microorganisms that colonize
58 the digestive tracts of humans and various animals (1). Certain *Bifidobacterium* species, namely
59 *Bifidobacterium longum* subsp. *infantis* (*B. infantis*), *Bifidobacterium bifidum*, and
60 *Bifidobacterium breve*, often predominate the human neonatal gut microbiota (GM) during
61 breastfeeding (2–5), and their predominance is directly linked with the healthy development of an
62 infant (6, 7). Among health-promoting effects attributed to bifidobacteria inhabiting the neonatal
63 gut are protection from pathogen colonization (8–10) and modulation of the immune system
64 towards the tolerance to commensal microorganisms (5, 11).
65 Decreased *Bifidobacterium* abundance is characteristic of stunted GMs observed in
66 preterm infants (14) and children suffering from severe acute malnutrition (73). Therapeutic
67 approaches aimed at restoring the bifidobacterial population in these affected groups include
68 administering exogenous *Bifidobacterium* species (e.g., *B. infantis*) as probiotics or/and food
69 formulas containing prebiotics that would selectively stimulate the growth of endogenous
70 bifidobacteria in the gut and thus confer beneficial properties to the infant (14–16, 73). Since the
71 prevalence of bifidobacteria in the neonatal gut is often attributed to their ability to selectively
72 utilize dietary human milk oligosaccharides (HMOs) (2, 17–21), these milk glycans are considered
73 "natural" prebiotics and added to infant formulas.
74 HMOs are the third most abundant (5 to 20 g/L) component of human milk after lactose
75 and lipids and are not assimilated by the infant (22–24). HMO building blocks include glucose,
76 galactose, *N*-acetylglucosamine (GlcNAc), L-fucose, and *N*-acetylneuraminic (sialic) acid; these
77 units form more than a hundred linear or branched oligosaccharide structures (22, 25, 26). Most
78 HMOs are composed of a lactose core (Galβ1-4Glc) conjugated to one or multiple lacto-*N*-biose
79 (LNB, Galβ1-3GlcNAc) or *N*-acetyllactosamine units (Galβ1-4GlcNAc) (22, 23, 26). The
80 resulting HMO structures are denoted as type I and type II chains, respectively, with lacto-*N*-
81 tetraose (LNT) and lacto-*N*-neotetraose (LNnT) as archetypes.
82 Previous studies have revealed substantial variation in HMO utilization strategies and
83 capabilities within the *Bifidobacterium* genus (19, 27). For example, *B. bifidum* uses a set of

84 membrane-attached glycoside hydrolases (GHs) that extracellularly degrade HMOs to di- and
85 monosaccharides (28–32) and then imports and catabolizes liberated LNB and lactose (19). In
86 contrast, *B. infantis* and *B. breve* import HMOs using ATP-binding cassette (ABC) transporters
87 and then degrade uptaken oligosaccharides to monosaccharides intracellularly using a repertoire
88 of exo-acting GHs (17, 20, 27).

89 *B. infantis*, which is widely used in probiotic and synbiotic formulations (8, 15, 16, 73),
90 possesses several unique genomic clusters (e.g., HMO cluster I or H1) that encode the most
91 elaborate HMO uptake (33, 34) and intracellular degradation machinery (35–38) among
92 bifidobacteria (**Fig. 1A-B**). While the molecular mechanisms of HMO utilization in *B. infantis*
93 have received considerable attention, the regulatory mechanisms governing this catabolic process
94 remain poorly understood. Previous studies demonstrated that LNT and LNnT induce a profound
95 and surprisingly similar transcriptomic response in *B. infantis* ATCC 15697, with multiple
96 genomic clusters (*nag*, *lnp*, H1; **Fig. 1B**) being upregulated, pointing to a possible global
97 regulatory mechanism(s) (39, 40). Uncovering this yet unknown mechanism in addition to
98 fundamental importance has a potential translational value for the rational selection of individual
99 HMOs for prebiotic and synbiotic formulations.

100 Comparative genomic analysis across multiple related genomes is a powerful approach for
101 reconstructing transcriptional regulatory networks (regulons) controlling carbohydrate utilization
102 (41–43). Our earlier bioinformatic analysis conducted on a limited set of *Bifidobacterium* genomes
103 implicated NagR, a transcription factor (TF) from the ROK family, as a regulator of *nag* and *lnp*
104 clusters (**Fig. 1B**) encoding GlcNAc and lacto-*N*-biose/galacto-*N*-biose (LNB/GNB) catabolic
105 pathways, respectively, in *B. bifidum*, *B. breve*, *B. infantis*, and *Bifidobacterium longum* subsp.
106 *longum* (*B. longum*) (42). James et al. further experimentally confirmed this prediction in *B. breve*
107 UCC2003 (44). However, it was unclear whether and how this knowledge translated to the global
108 regulation of the extensive HMO utilization machinery in *B. infantis.*

109 Here we reconstructed NagR regulons in a substantially larger collection of
110 *Bifidobacteriaceae* genomes focusing on HMO-utilizing species. This analysis revealed multiple
111 putative NagR-binding sites (operators) in the *B. infantis* ATCC 15697 genome, suggesting the
112 role of NagR as a global negative regulator of LNB/GNB, LNT, and LNnT utilization in this
113 bacterium. This conjecture was confirmed by transcriptome profiling of the *nagR* knockout mutant
114 and by direct assessment of the binding of recombinant NagR to its predicted operators. The

115 inferred NagR regulon structure suggests that at the transcriptional level, *B. infantis* is adapted to
116 simultaneous utilization of multiple HMOs, providing guidelines for using rationally formulated
117 mixtures rather than individual oligosaccharides as prebiotics. The reconstructed NagR regulons
118 also provided insights into the evolution of complex regulatory networks controlling carbohydrate
119 metabolism in bifidobacteria shaped by the ecological niche of these commensal microorganisms.
120

121 **Results**

122 **Genomic reconstruction reveals the complexity of the NagR regulon in *B. infantis***

123 We used a Position Weight Matrix (PWM)-based approach to reconstruct the NagR
124 regulon in *B. infantis* ATCC 15697 and identified 11 potential NagR-binding sites (operators) in
125 promoter regions of genes/operons encoding components of HMO utilization pathways including
126 six previously unknown NagR operators within in the HMO cluster I (H1) (**Fig. 1A-B and Table**
127 **S3**). Among the new putative NagR regulon members were genes encoding (i) LNnT (type II
128 HMOs) ABC transporters (HmoABC, HmoA2B2C2) (33), (ii) substrate-binding components of
129 ABC transporters possibly involved in HMO uptake (HmoA3, HmoA4, HmoA5, HmoA6) (33),
130 (iii) *N*-acetylneuraminate lyase NanA2 and exo-α-sialidase NanH2 (GH33) (35), (iv) lactose
131 permeases (LacS, LacS2) (**Table 1** and **Fig. 1A**). Overall, the reconstructed regulon contained 29
132 genes. The inferred NagR-binding motif had a palindrome structure (**Fig. 1C**), a common feature
133 of transcriptional regulators from the ROK family (45).

134 ROK-family TFs can function as transcriptional activators or repressors (45). To infer the
135 possible mode of action of NagR, we analyzed the position of predicted NagR operators relative
136 to −35 and −10 promoter elements recognized by bacterial RNA polymerase holoenzyme. Ten out
137 of 11 predicted NagR operators overlapped either with −10 or −35 sequences (**Fig. S1**), suggesting
138 that NagR was a potential transcriptional repressor. Therefore, based on the genomic
139 reconstruction, we hypothesized that NagR was a global negative regulator of genomic loci
140 involved in LNB/GNB, LNT, and LNnT utilization in *B. infantis* ATCC 15697.
141

142 **Engineered *nagR* insertional mutant displays comparable yet distinct physiological**
143 **properties compared to the parental wild-type strain**

144 To experimentally study the proposed regulatory role of NagR, we generated a *nagR*
145 knockout mutant of *B. infantis* ATCC 15697 (*nagR*-KO) by insertional mutagenesis and verified

146 the insertion position using genomic PCR (**Fig. S2A-B**). The *nagR*-KO mutant displayed a growth
147 rate comparable to the wild-type (WT) strain in the medium supplemented with lactose (MRS-CS-
148 Lac) but grew significantly slower (linear regression; $P_{adj} < 0.05$) in the medium supplemented
149 with a mixture of HMOs purified from pooled breastmilk (MRS-CS-HMO) (**Fig. 2**).
150 Unexpectedly, *nagR*-KO exhibited decreased growth in MRS-CS containing sucrose but not
151 fructooligosaccharides (FOS) (**Fig. 2**).
152 HMO consumption profiling of WT and *nagR*-KO strains grown in MRS-CS-HMO
153 revealed that both strains completely salvaged lactose, LNT, LNnT, 2'- and 3-fucosyllactose (FL),
154 and difucosyllactose (DFL) present in the HMO mixture (**Fig. S2C**). Notably, the *nagR*-KO mutant
155 consumed LNT faster compared to WT and displayed significantly delayed consumption (linear
156 regression; $P_{adj} < 0.05$) of large fucosylated HMOs, namely lacto-*N*-fucopentaoses (LNFP) I/II/III,
157 and lacto-*N*-difucohexaoses (LNDFH) I/II (**Fig. S2C**). The mutant also expelled less fucose into
158 the medium (**Fig. S2C**). Organic acid production profiling of supernatants of WT and *nagR*-KO
159 grown in MRS-CS-HMO showed that both strains released similar quantities of acetic acid;
160 however, the *nagR*-KO mutant produced significantly more formic and less lactic acid (linear
161 regression; $P_{adj} < 0.05$) (**Fig. S2D**). These data demonstrated comparable yet distinct physiological
162 properties of WT and *nagR*-KO strains, prompting a follow-up transcriptome profiling to infer
163 genes differentially expressed in the mutant.
164
165 **Comparative transcriptomics corroborates the role of NagR as a global negative**
166 **regulator of LNB/GNB, LNT, and LNnT utilization in *B. infantis***
167 We used RNA-seq to compare transcriptomes of *nagR*-KO and WT strains of *B. infantis*
168 ATCC 15697 grown in MRS-CS supplemented with lactose or LNnT, yielding four experimental
169 conditions (**Fig. 3A)**. The choice of carbon sources was based on a previous study, where LNnT
170 had induced many HMO utilization genes, and lactose had been used as a comparator (39).
171 Principle component analysis (PCA) of TMM-normalized counts revealed that each experimental
172 condition formed a distinct cluster (**Fig. 3B**). Specifically, PCA separated samples by carbon
173 source (PC1, 38.8% of the total variance) and strain (PC2, 21.1% of the total variance).
174 Linear modeling implemented in the *limma* framework (46) revealed significant
175 upregulation (Fold change (FC) > 2 and $P_{adj} < 0.01$) of multiple *nag*, *lnp*, and H1 cluster genes in
176 the *nagR*-KO strain grown in MRS-CS-Lac compared to WT grown in MRS-CS-Lac (**Table S2A**

177 and **Fig. 3C**). Overall, 19 out of 29 genes constituting the predicted NagR regulon were
178 upregulated. These results demonstrated that NagR, in line with the bioinformatic prediction,
179 functioned as a transcriptional repressor of the *nag*, *lnp*, and H1 loci. In addition, we observed: (i)
180 significant upregulation of *malEFG* encoding an ABC transport system for maltose (33) and (ii)
181 significant downregulation (FC < −2 and $P_{adj}$ < 0.01) of *cscA, cscB,* and *cscR* in the *nagR*-KO
182 mutant (**Table S2A** and **Fig. 3C**). The latter genes are involved in sucrose uptake and catabolism
183 (47) and are predicted to be controlled by a local regulator from the LacI family, CscR (42).
184 Notably, downregulation of the *csc* genes was consistent with decreased growth of the *nagR*-KO
185 mutant in the medium supplemented with sucrose (**Fig. 2**). No potential NagR operators were
186 identified in the promoter regions of *malEFG* and *csc* cluster genes, suggesting that the effect of
187 the *nagR* knockout on the expression of these genes was indirect.
188 RNA-seq of the WT strain revealed that 25 out of 29 genes constituting the predicted NagR
189 regulon were upregulated during growth in MRS-CS-LNnT compared to MRS-CS-Lac (**Table**
190 **S2B** and **Fig. 3D**). Interestingly, the expression of *nag*, *lnp*, and H1 cluster genes was higher in the
191 WT strain cultured in the presence of LNnT than in the *nagR*-KO mutant cultured in MRS-CS-
192 Lac (**Fig. S3**). These observations point to the presence of additional transcriptional activation
193 mechanism(s) that may contribute to the upregulation of these genomic clusters during the growth
194 of *B. infantis* in the medium containing LNnT. We also observed a significant upregulation of
195 *malK* (**Table S2B** and **Fig. S3**) encoding a shared ATPase component that energizes most
196 carbohydrate-specific ABC transporters in *B. infantis* (42). The upregulation of *malK* was likely
197 tied with the enhanced energetic needs associated with growth in the presence of LNnT,
198 specifically upregulation of genes encoding multiple ABC transport systems (GltFGH, GltABC,
199 HmoABC, HmoA2B2C2, HmoA3-A5).
200 Overall, the obtained transcriptomic data were consistent with our bioinformatic prediction
201 and demonstrated that NagR regulated LNB/GNB, LNT, and LNnT utilization in *B. infantis* by
202 repressing genomic loci encoding (i) transporters of respective glycans, (ii) GlcNAc and
203 LNB/GNB catabolic pathways.
204
205 **Interaction of NagR with predicted operators is dependent on GlcNAc and its**
206 **phosphorylated derivatives**

207 To test the interaction between NagR and its predicted operator sequences, we cloned the
208 *nagR* gene from *B. infantis* ATCC 15697 and expressed the recombinant protein as a fusion with
209 an N-terminal His-tag in *E. coli* BL21/DE3. Electrophoretic Mobility Shift Assay (EMSA)
210 demonstrated that NagR specifically bound DNA fragments (probes) containing predicted NagR
211 operators located upstream of *nagK*, *nagB*, *gltA*, *hmoA*, *hmoA2*, and *hmoA3* genes (**Fig. 4A-B, Fig.**
212 **5**, and **Table S1)**. Titration with increasing concentrations of NagR revealed that probes *nagK*,
213 *nagB_I*, *hmoA*, *hmoA2* had a high affinity for NagR ($EC_{50}$: 13-23 nM), whereas probes *gltA*,
214 *hmoA3* had a moderate one ($EC_{50}$: 140-180 nM) (**Fig. 5**). Reactions with probes *nagB_II* and
215 *hmoA6* did not manifest in robust shifts even at high (> 500 nM) protein concentrations (**Fig. S4A**).
216 These results showed that NagR exhibited different affinities to its various operator sequences.
217 Notably, we also observed a significant negative correlation (Pearson $R = -0.84$, $p < 0.05$) between
218 calculated $EC_{50}$ values for probes and expression FCs for respective genes in the RNA-seq
219 experiment (*nagR*-KO vs. WT grown in MRS-CS-Lac) (**Fig. 3C and 4C**), indicating the more
220 tightly NagR bound to an operator, the stronger was the observed upregulation of the respective
221 gene in *nagR*-KO.
222 To identify potential NagR effector molecules, we used a probe with a high affinity to
223 NagR (*hmoA*) and added various GlcNAc metabolism intermediates (GlcNAc, GlcNAc-6P, GlcN-
224 6P, GlcNAc-1P) to the binding reaction. GlcNAc, GlcNAc-6P, and GlcNAc-1P, but not GlcN-6P
225 disrupted the NagR-DNA complex at saturating (10 mM) concentration (**Fig. 4D**). Titration of the
226 effectors revealed that GlcNAc-6P lost its complex disrupting effect at 1 mM, whereas GlcNAc
227 and GlcNAc-1P lost their effect at 0.1 mM (**Fig. S4B**); the calculated effector $EC_{50}$ values were
228 $0.33 \pm 0.05$ mM for GlcNAc and $0.49 \pm 0.06$ mM for GlcNAc-1P (**Fig. 4E**).
229 Overall, the EMSA results demonstrated that (i) NagR bound its predicted operator
230 sequences with various affinities and (ii) multiple GlcNAc metabolism intermediates disrupted
231 NagR-DNA interactions *in vitro* and thus served as plausible NagR transcriptional effectors in *B.*
232 *infantis*.
233
234 **NagR regulon was a subject of evolutionary expansion in *Bifidobacteriaceae***
235 We used the PWM-based approach to identify NagR binding motifs and reconstruct
236 regulons in 25 genomes representing various phylogenetic lineages within the *Bifidobacteriaceae*
237 family to trace the potential evolutionary history of this gene regulatory network. Overall, the size

238 and composition of the reconstructed NagR regulons markedly varied among the studied genomes
239 (**Fig. 6A and Table S3**). For instance, early diverged *Bifidobacterium* species isolated from insects
240 and dairy products possessed concise NagR regulons comprised of a single genomic locus (*nag*)
241 encoding GlcNAc and predicted *N*,*N*'-diacetylchitobiose catabolic pathways (see **Supplementary**
242 **Results**). This observation suggests that in ancestral bifidobacteria, the role of NagR was likely
243 confined to local regulation of the GlcNAc catabolic pathway.
244 We observed gradual NagR regulon expansion in multiple *Bifidobacterium* genomes
245 isolated from mammalian hosts (**Fig. 6A, Table S3**, and **Supplementary Results**). The largest
246 and most complex NagR regulons were identified in strains isolated from the human neonatal gut,
247 namely *B. infantis* ATCC 15697 and *B. bifidum* PRL2010. Notably, while in *B. infantis*, the NagR
248 regulon expanded to include multiple HMO transporters encoded within the H1 cluster (**Fig. 1B**),
249 in *B. bifidum,* the reconstructed regulon contained genes encoding multiple GHs involved in
250 extracellular degradation of HMOs and mucin *O*-glycans (see **Supplementary Results**).
251 Additionally, the NagR regulon expansion was accompanied by minor variations in the NagR-
252 binding motif (**Fig. 6B** and **Supplementary Results**). Taken together, comparative reconstruction
253 of the NagR regulons suggests an evolutionary expansion from a local regulation of GlcNAc
254 catabolism in ancestral bifidobacteria toward a global regulation of various host glycan (e.g.,
255 HMOs) utilization in bifidobacterial species isolated from the mammalian neonatal gut.
256
257 **Discussion**
258 **Regulation of HMO utilization in *B. infantis***
259 The predominance of bifidobacteria in the infant gut is linked to their ability to consume
260 and use HMOs as a carbon source (17–21, 34). *B. infantis* possesses a unique genomic locus (H1)
261 that encodes multiple catabolic enzymes and components of ABC transporters that endow this
262 species with the ability to utilize a multitude of HMOs (17, 19, 27). Previous studies demonstrated
263 that a pooled HMO mixture, as well as individual HMOs (LNT and LNnT), induce expression of
264 H1 and *nag* cluster genes in *B. infantis* ATCC 15697, suggesting that the H1 cluster acts as an
265 HMO-inducible unit and is co-regulated with the GlcNAc catabolic pathway (39, 40). However,
266 the regulatory mechanisms underlying this phenomenon were not elucidated.
267 In this study, we have established NagR-mediated repression of H1, *lnp*, and *nag* cluster
268 genes in *B. infantis* ATCC 15697 by combining PWM-based regulon reconstruction with

269 transcriptome profiling of the *nagR*-KO mutant. The composition of the NagR regulon suggests
270 that this global TF regulates utilization of LNB/GNB, LNT, LNnT, and potentially other decorated
271 (e.g., sialylated) type I and II HMOs in *B. infantis*. We have also demonstrated concentration-
272 dependent binding of recombinant NagR to its predicted operators *in vitro*. The $EC_{50}$ values
273 inferred from EMSA experiments negatively correlated with fold-change values for upregulated
274 genes in the *nagR*-KO mutant. Thus, the degree of gene repression by NagR is strongly dependent
275 on the affinity of this TF to its cognate operator sequences in the promoter regions of corresponding
276 genes.

277 We identified GlcNAc and its phosphorylated derivatives, GlcNAc-6P and GlcNAc-1P, as
278 potential NagR transcriptional effectors in *B. infantis*. This result is somewhat unexpected since
279 James et al. previously described GlcNAc-6P, but not GlcNAc, as the NagR transcriptional
280 effector in *B. breve* UCC2003 (44). This discrepancy may reflect a metabolic adaptation to a more
281 global nature of NagR regulon in *B. infantis*, although, alternatively, it may be due to the
282 differences in the experimental approach, which in our case was based on the use of purified
283 recombinant NagR for EMSA experiments versus crude cell lysate as in (44). EMSA data indicated
284 that the acetyl group of these intermediary metabolites played a crucial role in NagR-effector
285 interactions, whereas phosphorylation of GlcNAc appeared to be dispensable. While GlcNAc and
286 GlcNAc-6P have been described as transcriptional effectors of the ROK family TFs (44, 48, 49),
287 the potential effector role of GlcNAc-1P is novel and unexpected. Although *N*-acetylhexosamine
288 1-kinase (LnpB) can phosphorylate GlcNAc to GlcNAc-1P (50), no enzymes that convert the latter
289 to GlcNAc-6P and thus shunt it to the GlcNAc catabolic pathway have been described in
290 prokaryotes (51). In contrast, GlcNAc-1P can be converted to UDP-GlcNAc by GlcNAc-1P
291 uridyltransferase (GlmU) and enter the peptidoglycan biosynthesis pathway (52). Therefore,
292 additional studies are required to assess the biological significance of GlcNAc-1P functioning as
293 a potential NagR transcriptional effector.

294 Based on the obtained data, we propose a model (**Fig. 7**) where the release of GlcNAc
295 during degradation of LNT and LNnT by intracellular GHs results in derepression of *nag*, *lnp*, and
296 H1 clusters, including genes encoding LNT and LNnT transporters. This model explains the
297 similarity of transcriptomic responses of *B. infantis* ATCC 15697 during growth in media
298 supplemented with LNT or LNnT (39, 40) and suggests that utilization of any GlcNAc-containing
299 glycan by this bacterium will result in upregulation of NagR-controlled genes. Among these

300 glycans might be particular fucosylated HMOs (e.g., LNFP I) and milk *N*-glycans imported by
301 different ABC transport systems (34, 73). Consistent with this notion, a previous study
302 demonstrated that *B. infantis* upregulates *nag* and H1 cluster genes (53) when utilizing *N*-
303 glycosylated human lactoferrin. Our model, however, does not all explain all transcriptional
304 responses observed during HMO utilization by *B. infantis*. The transcriptomic profile of the WT
305 strain grown in MRS-CS-LNnT suggests that additional mechanisms may co-activate the
306 expression of NagR regulon genes (particularly within the *lnp* cluster) in physiologically inducing
307 conditions.

308 The structure of the NagR-mediated transcriptional network in *B. infantis* likely reflects
309 the evolutionary adaptation of this bacterium to simultaneous foraging of multiple distinct HMOs
310 and other milk glycans. This notion suggests that using a mixture of LNT and LNnT (and
311 potentially other HMOs) rather than individual oligosaccharides as a prebiotic may be a more
312 efficient solution for selective stimulation of *B. infantis* growth in the neonatal gut since it
313 considers the nuanced regulatory mechanisms and physiology of the target organism.

314

315 **Evolution of the NagR regulon in bifidobacteria**

316 Evolution of *B. infantis* was shaped by its ecological niche, specifically the need to forage
317 dietary milk glycans (e.g., HMOs) abundantly present in the gut of breastfed infants (17). This
318 adaptation led to the emergence of several unique genomic clusters, such as H1, that, as we
319 demonstrated, are transcriptionally controlled by NagR. However, other *Bifidobacterium* species
320 inhabiting the neonatal gut, such as *B. breve* and *B. longum*, do not harbor the H1 cluster (42) and
321 have less complex NagR regulons (42, 44). Thus, to elucidate the evolutionary history of the NagR
322 regulon expansion, we reconstructed it in 25 representative genomes spanning 18 *Bifidobacterium*
323 and one *Alloscardovia* species isolated from various hosts and environments.

324 The regulon structure in early diverged species (e.g., *B. asteroides* and *B. aquikefiri*)
325 suggests that NagR functioned as a local regulator of a single genomic locus involved GlcNAc and
326 possibly *N*,*N'*-diacetylchitobiose catabolism (**Table 2**). Not surprisingly, the activity of this TF
327 would be modulated by GlcNAc or/and GlcNAc-6P. During colonization of mammalian hosts,
328 bifidobacteria were exposed to various GlcNAc-containing glycans, which resulted in the
329 emergence of genomic loci involved in the metabolism of these carbohydrates. For example, most
330 bifidobacteria isolated from mammals (including humans) harbor the *lnp* cluster encoding the

331 LNB/GNB utilization pathway. This observation marks the next evolutionary step of NagR-
332 mediated regulation, namely transitioning from a local to a multi-locus (*nag* and *lnp*) controlling
333 TF. An independent alternative regulon expansion scenario presumably occurred in *B. catenulatum*
334 and *B. pseudocatenulatum* lacking *lnp* genes. In these strains, NagR expanded to control the
335 potential LNB/LNT utilization machinery (*gltABC-nagK-hex1-nagB-nagA* and *bga42A* genes)
336 (**Table 2**). Such regulon structure is consistent with a recently published transcriptomic dataset
337 where LNFP I induced the *gltABC-nagK-hex1-nagB-nagA* cluster in *B. pseudocatenulatum* (54).
338 Although LNFP I is imported into the cell by a different ABC transport system (21, 54), the
339 GlcNAc released during the degradation of this oligosaccharide was likely responsible for the
340 derepression of NagR-controlled genes in this species.

341 Multiple further NagR regulon expansion events occurred in prevalent infant-associated
342 species with the highest HMO utilization potential: *B. infantis* and *B. bifidum*. In *B. infantis*, the
343 regulon included multiple H1 cluster genes predominantly encoding HMO transporters, whereas
344 in *B. bifidum*, it expanded to control extracellular GHs involved in HMO and mucin *O*-glycan
345 degradation. These observations provide a fascinating example of how catabolic machinery
346 corresponding to two distinct (convergent) strategies of HMO utilization (intracellular vs.
347 extracellular) became controlled by the same regulatory system. The structure of the NagR regulon
348 partially explains the upregulation of *nag*, *lnp*, and genes encoding extracellular GHs in *B. bifidum*
349 PRL2010 during growth in a mucin-based medium (18). The predicted NagR operators also
350 coincide with a palindromic motif in promoter regions of these upregulated genes (18).

351 Regulon expansion has been previously described in various bacterial lineages (55, 56),
352 including bifidobacteria (57, 58); however, the underlying evolutionary rationale of such events
353 was not always clear. Here we propose that the general theme of the NagR regulon evolution in
354 the *Bifidobacterium* genus was an expansion to include genes involved in catabolism of GlcNAc-
355 containing host glycans, given the ability of NagR to sense GlcNAc and/or its phosphorylated
356 derivatives. Interestingly, the evolution of gene regulatory networks governing utilization of
357 GlcNAc-containing glycans in bifidobacteria was not limited to the NagR regulon expansion.
358 Another scenario involved duplication(s) of the *nagR* gene after speciation followed by the
359 functional divergence of emerged paralogs. For example, in *B. breve* UCC2003, while NagR
360 controls LNB/GNB and LNT utilization pathways, its paralogs, NahR and AtsR, repress the gene
361 encoding an LNnT transporter and a genomic locus involved in utilization of GlcNAc-6S (a mucin

362 *O*-glycan-constituting saccharide released by 6-sulfo-β-*N*-acetylglucosaminidase BbhII),
363 respectively (44, 59, 60) (**Table 2**). Interestingly, different GlcNAc derivatives serve as effector
364 molecules of these TFs: GlcNAc for NahR and GlcNAc-6S for AtsR (44, 59). Another NagR
365 paralog in *B. infantis* 2D9, NglR, potentially controls a genomic locus involved in utilizing
366 complex *N*-glycans (73) (**Table 2**). Overall, these observations illustrate how bifidobacteria
367 adapted to regulate the foraging of host glycans during colonization of mammalian hosts and shed
368 more light on how complex regulatory networks emerge and evolve.
369

370 **Methods**

371 **Reagents and bacterial strains**

372 Reagents were purchased from Alfa-Aesar (Tewksbury, MA, USA), Ambion (Austin, TX,
373 USA), Combi-Blocks (San Diego, CA, USA), Sigma-Aldrich (St. Louis, MO, USA), and
374 Invitrogen (Carlsbad, CA, USA) unless indicated otherwise. Synthetic LNnT (>95% purity) was
375 generously donated by DSM (Heerlen, Netherlands). The HMO mixture was prepared from pooled
376 human milk (21). The experimental protocol was reviewed and approved by the Ethics Committee
377 of Kyoto University (R0046); the study was performed per the Declaration of Helsinki, and
378 informed consent was obtained from all mothers (all subjects). Oligonucleotides were synthesized
379 by Integrated Genomic Technologies (Coralville, IA, USA). Phusion High-Fidelity DNA
380 Polymerase, restriction enzymes, and Quick Ligase were purchased from New England BioLabs
381 (Ipswich, MA, USA).

382 The type strain of *B. infantis* (ATCC 15697 = JCM 1222) was obtained from Japan
383 Collection of Microorganisms (RIKEN BioResource Research Center, Tsukuba, Japan).
384 *Escherichia coli* DH5α and One Shot TOP10 (Invitrogen) were used for genetic manipulations. *E.*
385 *coli* BL21/DE3 (New England BioLabs) was used for recombinant NagR overexpression.

386 **Bioinformatic analysis**

387 A previously established comparative genomic approach was used to identify putative
388 NagR operator sequences and reconstruct regulons in *B. infantis* ATCC 15697 and other selected
389 strains (42, 57). For the initial NagR regulon reconstruction, we built a PWM based on data
390 available in the RegPrecise database (42, 61) using SignalX (62). To improve identification of
391 NagR operators, we built additional PWMs representing two different NagR-binding motifs
392 specific for distant bifidobacterial lineages. Constructed PWMs were used to search for new

393 potential NagR operators using GenomeExplorer (62) with the following parameters: (i) -500 to
394 +50 bp position relative to the first codon of a gene, (iii) site score threshold of 4.3. Identified sites
395 were screened using the consistency check and phylogenetic footprinting approaches to filter out
396 false positives (63). NagR-binding motifs were visualized via WebLogo (64). Positions of the −10
397 and −35 promoter elements were determined via similar PMW-based searches based on data
398 available for *B. breve* (65) and *B. longum* (66). Promoter regions were aligned using Pro-Coffee
399 (67). Details on additional genomic analysis of *Bifidobacterium* strains and phylogenetic inference
400 are available in Supplementary Methods.

401 **Targeted *nagR* gene disruption in *B. infantis* ATCC 15697**

402 A single-crossover recombination event was used to inactivate the *nagR* gene (Blon_0880)
403 in *B. infantis* ATCC 15697. Briefly, a BamHI-digested, 2.0 kb fragment of pBS423 (68) that carries
404 the pUC ori and a spectinomycin resistance gene was self-ligated to generate pTK2051, a plasmid
405 incapable of replicating in bifidobacteria. The internal region of *nagR* was then amplified by PCR
406 using primers NagR_I/NagR_II (**Table S1**) and genomic DNA as a template. The amplified 0.5
407 kbp fragment was inserted into the BamHI site of pTK2051 using the In-Fusion Snap Assembly
408 kit (Takara Bio USA, Mountain View, CA, USA). The resulting suicide plasmid was introduced
409 into *B. infantis* by electroporation (34). Transformed cells were screened for spectinomycin
410 resistance, and the clones were subsequently subjected to a genomic PCR analysis at the *nagR*
411 locus using primers NagR_III/NagR_IV (**Table S1** and **Fig. S2A**). The amplicon was directly
412 sequenced to ensure that the suicide plasmid was integrated into the intended site.

413 **Culture conditions**

414 *B. infantis* was routinely grown in Gifu anaerobic medium (Nissui Pharmaceutical, Tokyo,
415 Japan) or Lactobacilli MRS Broth w/o Dextrose (MRS-CS; Alpha Biosciences, Baltimore, MD,
416 USA) with 0.34% (w/v) sodium ascorbate, 0.029% (w/v) L-cysteine-HCl monohydrate
417 (BeanTown Chemical, Hudson, NH, USA). The MRS-CS medium was supplemented with lactose,
418 LNnT, or a mixture of neutral HMOs at a final concentration of 1% (w/v). Cultures were incubated
419 at 37 °C in AnaeroPack system (Mitsubishi Gas Chemical Company, Tokyo, Japan) or an
420 anaerobic chamber maintained with a gas mix of 10% $H_2$, 10% $CO_2$, and 80% $N_2$ (Coy Laboratory
421 Products, Grass Lake, MI, USA). Growth was monitored either by measuring culture turbidity in
422 McFarland units using Densitometer DEN-1B (Grant Instruments, Shepreth, UK) or optical
423 density at 600 nm ($OD_{600}$) using DU800 Spectrophotometer (Beckman Coulter, Brea, CA, USA).

424 *E. coli* strains were cultured in Luria-Bertani broth at 37°C with vigorous agitation. Where
425 appropriate, growth media were supplemented with spectinomycin (15 μg/ml for *B. infantis* ATCC
426 15697 *nagR*-KO, 75 μg/ml for *E. coli* DH5α) or kanamycin (50-60 μg/ml for all other *E. coli*
427 strains). Details on measuring HMO consumption and organic acid production are described in
428 Supplementary Methods.

429 **Transcriptome analysis (RNA-seq)**

430 For RNA-seq experiments, *B. infantis* ATCC 15697 WT and *nagR*-KO strains were grown
431 as described above in MRS-CS (without spectinomycin) supplemented with either 1% lactose or
432 1% LNnT. Samples (2 ml; biological triplicates) were collected at the early-mid exponential phase
433 ($OD_{600}$=0.35) and immediately pelleted in a prechilled centrifuge at 4,800 × g for 5 min. Cell
434 pellets were snap-frozen in liquid nitrogen and stored at −80 °C until further use. RNA was
435 extracted as described previously (69) with minor modifications; the detailed protocol can be found
436 in Supplementary Methods. Ribosomal RNA was depleted with NEBNext rRNA Depletion Kit
437 for Bacteria (New England Biolabs). Barcoded libraries were made with NEBNext Ultra II
438 Directional RNA Library Prep Kit for Illumina (New England Biolabs). Libraries were pooled and
439 sequenced (single-end 75 bp reads) on Illumina NextSeq 500 using High Output V2 Kit (Illumina,
440 San Diego, CA, USA). Sequencing data were analyzed as described previously (70) with certain
441 modifications; the detailed pipeline is described in Supplementary Methods.

442 **Cloning, expression, and purification of recombinant NagR**

443 Codon-optimized nucleotide sequence of *nagR* (Blon_0880) was synthesized by GeneArt
444 Gene Synthesis (Thermo Fisher Scientific, Waltham, MA, USA), PCR-amplified using primers
445 NagR_HisN_F and NagR_HisN_R (**Table S1**), digested by BamHI and SalI, and ligated into a
446 pre-digested in-house pET-49b(+) vector conferring resistance to kanamycin. The ligation mixture
447 was introduced into *E. coli* One Shot TOP10 by chemical transformation, and transformants were
448 then selected based on kanamycin resistance. The recombinant NagR was expressed as fusion with
449 N-terminal His-tag under control of a T7 promoter in *E. coli* BL21/DE3. Cells were grown in LB
450 medium (50 mL) at 37 °C to an $OD_{600}$ of ~0.6 and then transferred to 16 °C. Protein expression
451 was induced by adding 0.2 mM isopropyl-β-D-thiogalactopyranoside. Cells were grown at 16° C
452 overnight and collected by centrifuging at 4,800 × g for 15 min. Harvested cells were resuspended
453 in a lysis buffer containing 10 mM HEPES buffer (pH 7.0), 100 mM NaCl, 0.15% Brij-35, and 5
454 mM β-mercaptoethanol. Cells were lysed by a freeze-thaw cycle, followed by sonication using

455 Misonix Sonicator 3000 (Misonix Inc, Farmingdale, NY, USA). The cell debris was removed; the
456 soluble fraction was loaded onto a Ni-NTA agarose mini-column (0.2 mL) (Qiagen, Hilden,
457 Germany). The column was washed with 10 column volumes of At buffer (50 mM Tris-HCl (pH
458 8.0), 500 mM NaCl, 20 mM imidazole, 0.3% Brij-35, 5 mM β-mercaptoethanol) and 10 column
459 volumes of At buffer with 1 M NaCl. Captured proteins were eluted with 0.6 mL of At buffer with
460 300 mM imidazole. The eluted protein fraction was concentrated and buffer-exchanged into 10
461 mM Tris-HCl (pH 8.0) with 50 mM NaCl using 30 kDa Amicon Ultra 0.5 mL Centrifugal Filters
462 (MilliporeSigma, Burlington, MA, USA). Protein concentration was determined by Qubit Protein
463 Assay Kit (Invitrogen).

464 **Electrophoretic mobility shift assays**

465 Oligonucleotides containing predicted 21-bp NagR operators and surrounding genomic
466 regions (14 bp from each end) were synthesized by Integrated DNA Technologies. DNA fragment
467 sequences, sizes, and labels used for testing are given in **Table S1**. Double-stranded labeled DNA
468 probes were obtained by annealing IRD700-labelled oligonucleotides with unlabeled
469 complementary oligonucleotides (ratio 1:5) in 4 mM Tris-HCl (pH 8.0), 20 mM NaCl, 0.4 mM
470 EDTA in Mastercycler PRO Thermal Cycler (Eppendorf) overnight. Binding reactions were
471 carried out with a final volume of 20 µl in binding buffer containing 10 mM Tris–HCl (pH 7.5),
472 50 mM KCl, 5 mM $MgCl_2$, 2.25 mM DTT, 0.125% Tween 20, and 2.5% glycerol. DNA probes (1
473 nM) were incubated with increasing concentrations of the purified NagR (0-2000 nM) for 60 min
474 at room temperature. Reaction mixtures were loaded on a Novex 6% DNA Retardation gel
475 (Thermo Fisher Scientific) and run in 0.5x Tris-Borate-EDTA buffer (Thermo Fisher Scientific)
476 at 100 V and room temperature for 45 min in XCell SureLock Mini-Cell Electrophoresis System
477 (Thermo Fisher Scientific). Gels were visualized using Odyssey CLx (Li-COR Biosciences,
478 Lincoln, NE, USA). Bands were quantified in Image Studio v5.2 (Li-COR Biosciences). The
479 resulting data were imported into R and approximated by a 4-parameter logistic (4PL) equation in
480 the *drc* package (71) to calculate $EC_{50}$ values. In the 4PL model, the lower limit was fixed at 0 and
481 the upper limit at 1. To identify possible NagR effectors, binding reactions were carried out with
482 25 nM of NagR and the addition of 0.1-10 mM of GlcNAc or its phosphorylated derivatives
483 (GlcNAc-6P, GlcN-6P, GlcNAc-1P). $EC_{50}$ values for effectors were calculated using the 4PL
484 equation with the upper limit fixed at 1.

485 **Data availability**

The RNA-seq dataset is deposited on Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE196064). Raw EMSA gel quantification, growth, HMO consumption, and organic acid production data are available on GitHub (https://github.com/Arzamasov/NagR_manuscript). Code detailing the data analysis steps is available via GitHub and in Supplementary Code File 1.

**Abbreviations**

HMOs — human milk oligosaccharides
LNB — lacto-*N*-biose
GNB — galacto-*N*-biose
LNT — lacto-*N*-tetraose
LNnT— lacto-*N*-neotetraose
GlcNAc — *N*-acetylglucosamine
GM — gut microbiota
GH — glycoside hydrolase
ABC — ATP-binding cassette
TF — transcription factor
PWM — position weight matrix
WT — wild-type
PCA — principal component analysis
FC — fold change
EMSA — electrophoretic mobility shift assay
4PL — 4-parameter logistic

**Acknowledgments**

We thank Semen Leyn (SBP) for help with computational analyses; Kazi Ahsan (Washington University in St. Louis) for guidance in RNA isolation experiments; Kang Liu and Brian James (SBP Genomics Core) for library preparation and sequencing; Daniel Beiting (University of Pennsylvania) for the DIY.transcriptomics course; Irina Rodionova (University of California San Diego) and Oleg Kurnasov (SBP) for guidance in protein expression and purification experiments; Motomitsu Kitaoka (Niigata University), Junko Hirose (Kyoto Women's

517 University), Tadasu Urashima (Obihiro University of Agriculture and Veterinary Medicine) for
518 technical support; RIKEN BRC through the National BioResource Project of the MEXT/AMED,
519 Japan for providing pBS423, Glycom A/S and DSM for generously providing LNnT.
520 This work was supported by JSPS KAKENHI (grant 21H02116 to T.K) and NIH (grant
521 DK30292 to A.L.O).
522 Conceptualization: A.A.A; data curation: A.A.A; formal analysis: A.A.A, investigation:
523 A.A.A, A.N., M.S., M.N.O; methodology: A.A.A, A.N., M.S., M.N.O; supervision: T.K., D.A.R,
524 A.L.O; visualization: A.A.A; writing – original draft: A.A.A with input from A.N; writing – review
525 & editing: T.K., D.A.R, A.L.O.
526 A.O. and D.R. are co-founders of Phenobiome Inc., a company pursuing development of
527 computational tools for predictive phenotype profiling of microbial communities. Employment of
528 M.S and M.N.O at Kyoto University is supported by Morinaga Milk Industry Co., Ltd. All other
529 authors declare no conflict of interest.
530
531 **Figure legends**
532 **FIG 1.** Reconstructed NagR regulon in *B. infantis* ATCC 15697. (A) Schematic
533 representation of LNT, LNnT, and LNB/GNB utilization pathways in *B. infantis* ATCC 15697.
534 *Step 1*: HMOs and their constituents are transported into the cell by various transport systems. *Step*
535 *2*: once inside the cell, HMOs are degraded from the non-reducing end by a coordinated action of
536 exo-acting GHs. Breakdown of glycosidic bonds by specific GHs is indicated in light orange
537 circles. *Step 3*: released monosaccharides are converted to fructose-6P and enter the bifid shunt.
538 (B) Genomic loci constituting the reconstructed NagR regulon in *B. infantis* ATCC 15697.
539 Numbers represent locus tags in the Blon_XXXX format (GenBank accession no. CP001095.1).
540 (C) NagR-binding motif in *B. infantis* ATCC 15697 based on 11 predicted operator sequences.
541 **FIG 2.** Growth curves of *B. infantis* ATCC 15697 WT and *nagR*-KO strains in MRS-CS
542 supplemented with various carbon sources (1% w/v). Growth was monitored by measuring $OD_{600}$
543 at specific time points. Data points represent the mean of three biological replicates. Error bars
544 depict 95% confidence intervals for the mean obtained by bootstrapping. Timepoints where the
545 $OD_{600}$ values for WT and *nagR*-KO strains were significantly different (*, $P_{adj}<0.05$) were
546 identified using a linear regression. Bonferroni correction was used to adjust for multiple testing.

547 **FIG 3**. RNA-seq of WT and *nagR*-KO strains grown in MRS-CS supplemented with
548 lactose and LNnT. (A) Schematic representation of the experimental design. (B) PCA of TMM-
549 normalized count data. Each data point represents one sample. (C) and (D) Volcano plots depicting
550 the $\log_2$FC of gene expression versus the $-\log_{10}P_{adj}$. (C) compares *nagR*-KO and WT strains grown
551 in MRS-CS-Lac, whereas (D) compares the WT strain grown in MRS-CS-LNnT and MRS-CS-
552 Lac. Criteria for calling differentially expressed genes were the follows: $P_{adj} < 0.01$ and absolute
553 FC > 2. Genes constituting the predicted NagR and CscR regulons are labeled and colored in
554 orange and purple, respectively.

555 **FIG 4.** Interactions of recombinant NagR with predicted operators and screening of NagR
556 effectors. (A) Predicted NagR operator sequences identified in the promoter regions of listed
557 genes. Conserved nucleotides are in bold. The full sequences of probes used in EMSA experiments
558 are given in Table S1. (B) An EMSA gel showing interactions of NagR with DNA probes
559 containing predicted operators. NagR concentration was 25 nM, and probe concentrations were 1
560 nM. The *ybdO* probe was used as a control for non-specific binding. (C) Correlation between probe
561 $EC_{50}$ values determined via EMSA and expression fold-changes of cognate genes in the RNA-seq
562 experiment (*nagR*-KO vs. WT grown in MRS-CS-Lac). (D) Effect of various GlcNAc metabolism
563 intermediates (10 mM) on the interaction between NagR (25 nM) and the *hmoA* probe (1 nM). (E)
564 $EC_{50}$ values of selected NagR effector molecules. The y-axis depicts the ratio $\frac{\%\ shift\ with\ effector}{\%\ shift\ w/o\ effector}$.
565 Calculated $EC_{50}$ are shown as big circles. 25 nM of NagR and 1 nM of the *hmoA* probe were used.

566 **FIG 5**. EMSA gels depicting titration of DNA probes (1 nM) containing predicted NagR
567 operators with recombinant NagR. Gels were quantified, and the results were approximated by the
568 4PL equation. NagR concentration at which half of the probe is shifted ($EC_{50}$) is shown. Grey
569 shading depicts 95% confidence intervals. The x-axis is in the Log10 scale.

570 **FIG 6.** Evolution of the NagR regulon and binding motif within the *Bifidobacteriaceae*
571 family. (A) NagR regulon composition in 25 *Bifidobacteriaceae* strains mapped on a species tree
572 built based on the alignment of 247 core genes. Bootstrap values are shown as purple circles. Black
573 symbols indicate strain isolation sources. Regulon members are colored according to their
574 function: catabolic enzymes are green, GHs in orange, and transporters in blue. (B) NagR-binding
575 motifs mapped on a tree of NagR proteins from 25 *Bifidobacteriaceae* strains. Bootstrap values
576 are shown as purple circles. NagR paralogs in *B. bohemicum* and *B. bifidum* are denoted by
577 numbers (1 and 2).

578 **FIG 7**. Model of NagR-mediated regulation of HMO utilization in *B. infantis*. *Step 1*:
579 GlcNAc-containing milk glycans (e.g., LNT and LNnT) are uptaken into the cell by various ABC
580 transporters. *Step 2*: once inside the cell, the glycans are degraded by intracellular GHs, and
581 GlcNAc is released. *Step 3*: released GlcNAc interacts with NagR and disrupts the NagR-operator
582 complex, leading to derepression of NagR-controlled genes.

583 **FIG S1.** Multiple sequence alignments of promoter regions of NagR-regulated genes in
584 *Bifidobacterium* spp. Predicted NagR operators are highlighted in yellow. Predicted promoter
585 elements (-35 and -10 sequences) and in bold black. Experimentally described transcription start
586 sites (TSS) in *B. breve* UCC2003 (65) are in bold blue. Predicted Shine-Dalgarno sequences are
587 in bold red. Coding sequences are highlighted in grey; initiation codons are underlined. Conserved
588 nucleotides are marked with asterisks.

589 **FIG S2**. Construction and metabolic profiling of the *B. infantis* ATCC 15697 *nagR*-KO
590 mutant. (A) Schematic representation of insertional inactivation of *nagR*. The region used for a
591 single crossover recombination event is in orange. The primers for genomic PCR (green) were
592 designed to anneal outside the region used for recombination. (B) Results of genomic PCR. The
593 amplicon sizes were expected to be 679 bp and 3164 bp for WT and *nagR*-KO strains, respectively.
594 Four clones were analyzed. (C) HMO consumption and (D) organic acid production profiles of *B.*
595 *infantis* ATCC 15697 WT and *nagR*-KO strains grown in MRS-CS-HMO. Data points represent
596 the mean of three biological replicates. Error bars depict 95% confidence intervals for the mean.
597 Timepoints where the metabolite concentrations for WT and *nagR*-KO strains were significantly
598 different (*, $P_{adj}<0.05$) were identified using a linear regression. Bonferroni correction was used
599 to adjust for multiple testing.

600 **FIG S3**. Heatmap depicting the expression of genes from Table S2A-B across four
601 experimental conditions (WT and *nagR*-KO strains grown in MRS-CS supplemented either with
602 Lac or LNnT). Genes constituting the predicted NagR regulon are in bold. Genomic clusters from
603 Fig. 1B are marked by lines.

604 **FIG S4**. (A) EMSA gels depicting titration of DNA probes (1 nM) containing predicted
605 NagR operators with recombinant NagR. (B) EMSA gels depicting titration of the NagR-DNA
606 probe complex with various effector molecules. 25 nM of NagR and 1 nM of the *hmoA* probe were
607 used in all reactions.
608

609 **Table legends**

610 **TABLE S1**. Oligonucleotides used in this study. IRD oligonucleotides are 5'-labeled with
611 the IRD700 dye, whereas RC oligonucleotides are unlabeled. Predicted NagR operators are in
612 bold.

613 **TABLE S2**. List of differentially expressed genes ($P_{adj} < 0.01$; absolute $log_2FC > 1$)
614 between selected experimental conditions in the RNA-seq experiment. (A) *nagR*-KO grown in
615 MRS-CS-Lac vs. WT grown in MRS-CS-Lac. (B) WT grown in MRS-CS-LNnT vs. WT grown
616 in MRS-CS-Lac.

617 **TABLE S3**. Reconstructed NagR regulons and predicted NagR operators in 25
618 *Bifidobacteriaceae* genomes.

619

620 **TABLE 1. Composition of the reconstructed NagR regulon in *B. infantis* ATCC 15697**

| Locus tag | Gene | Annotation | Reference |
|---|---|---|---|
| Blon_0879 | *nagK* | Predicted *N*-acetyl-glucosamine kinase, ROK family (EC 2.7.1.59) | - |
| Blon_0880 | *nagR* | Transcriptional regulator of LNB/GNB, LNT, and LNnT utilization, ROK family | This study |
| Blon_0881 | *nagB* | Glucosamine-6-phosphate deaminase (EC 3.5.99.6) | - |
| Blon_0882 | *nagA* | *N*-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25) | - |
| Blon_0883 | *gltF* | LNB and GNB ABC transporter, substrate-binding protein | (33) |
| Blon_0884 | *gltG* | LNB and GNB ABC transporter, permease protein 1 | - |
| Blon_0885 | *gltH* | LNB and GNB ABC transporter, permease protein 2 | - |
| Blon_2171 | *lnpD* | UDP-hexose 4-epimerase (EC 5.1.3.2) | (50) |
| Blon_2172 | *lnpC* | UTP-hexose-1-phosphate uridylyltransferase (EC 2.7.7.10) | (50) |
| Blon_2173 | *lnpB* | *N*-acetylhexosamine 1-kinase (EC 2.7.1.162) | (50) |
| Blon_2174 | *lnpA* | 1,3-β-galactosyl-*N*-acetylhexosamine phosphorylase (EC 2.4.1.211) | (72) |
| Blon_2175 | *gltC* | LNT, LNB, and GNB ABC transporter, permease protein 2 | - |
| Blon_2176 | *gltB* | LNT, LNB, and GNB ABC transporter, permease protein 1 | - |
| Blon_2177 | *gltA* | LNT, LNB, and GNB ABC transporter, substrate-binding protein | (33) |
| Blon_2331 | *lacS2* | Lactose permease, GPH translocator family | - |
| Blon_2332 | *lacS* | Lactose permease, GPH translocator family | - |
| Blon_2341 | - | Hypothetical protein | - |
| Blon_2342 | *hmoC2* | Type II HMO ABC transporter, permease protein 2 | - |
| Blon_2343 | *hmoB2* | Type II HMO ABC transporter, permease protein 1 | - |
| Blon_2344 | *hmoA2* | Type II HMO ABC transporter II, substrate-binding protein | (33) |
| Blon_2345 | *hmoC* | Type II HMO ABC transporter, permease protein 2 | - |
| Blon_2346 | *hmoB* | Type II HMO ABC transporter, permease protein 1 | - |
| Blon_2347 | *hmoA* | Type II HMO ABC transporter I, substrate-binding protein | (33) |
| Blon_2348 | *nanH2* | H1 cluster exo-α-2,3/6-sialidase (GH33) | (35) |
| Blon_2349 | *nanA2* | *N*-acetylneuraminate lyase (EC 4.1.3.3) | - |
| Blon_2350 | *hmoA3* | Predicted HMO ABC transporter, substrate-binding protein | (33) |
| Blon_2351 | *hmoA4* | Predicted HMO ABC transporter, substrate-binding protein | (33) |
| Blon_2352 | *hmoA5* | Predicted HMO ABC transporter, substrate-binding protein | (33) |
| Blon_2354 | *hmoA6* | Predicted HMO ABC transporter, substrate-binding protein | (33) |

621

622 **TABLE 2. Characteristics of NagR orthologs and paralogs from various**
623 ***Bifidobacterium* species**

| TF | Species | % identity* | Effector | Regulated pathway | Ref. |
|---|---|---|---|---|---|
| NagR | *B. infantis* ATCC 15697 | 100 | GlcNAc, GlcNAc-1P, GlcNAc-6P | LNB/GNB, LNT, and LNnT utilization | This study |
| NagR | *B. breve* UCC2003 | 84 | GlcNAc-6P | LNB/GNB, LNT utilization | (44) |
| NagR | *B. bifidum* PRL2010 | 60/53** | Unknown | LNB/GNB, HMO, and mucin *O*-glycan utilization# | This study |
| NagR | *B. pseudocatenulatum* JCM 1200 | 40 | Unknown | LNB and LNT utilization# | This study |
| NagR | *B. asteroides* PRL2011 | 40 | Unknown | GlcNAc catabolism# | This study |
| NahR | *B. breve* UCC2003 | 45 | GlcNAc | LNnT utilization | (44) |
| AtsR | *B. breve* UCC2003 | 40 | GlcNAc-6S | GlcNAc-6S utilization | (59) |
| NglR | *B. infantis* Bg_2D9 | 40 | Unknown | Complex *N*-glycan utilization# | (73) |

624 * Sequence identity with NagR from *B. infantis* ATCC 15697
625 ** Percentages for two NagR copies in *B. bifidum* PRL2010
626 # Prediction based on regulon reconstruction
627

628 **References**


629 1. Alessandri G, van Sinderen D, Ventura M. 2021. The genus bifidobacterium: From genomics
630 to functionality of an important component of the mammalian gut microbiota running title:
631 Bifidobacterial adaptation to and interaction with the host. Comput Struct Biotechnol J
632 19:1472–1487.
633 2. Matsuki T, Yahagi K, Mori H, Matsumoto H, Hara T, Tajima S, Ogawa E, Kodama H,
634 Yamamoto K, Yamada T, Matsumoto S, Kurokawa K. 2016. A key genetic factor for
635 fucosyllactose utilization affects infant gut microbiota development. Nat Commun 7:11939.
636 3. Vatanen T, Franzosa EA, Schwager R, Tripathi S, Arthur TD, Vehik K, Lernmark Å,
637 Hagopian WA, Rewers MJ, She J-X, Toppari J, Ziegler A-G, Akolkar B, Krischer JP,
638 Stewart CJ, Ajami NJ, Petrosino JF, Gevers D, Lähdesmäki H, Vlamakis H, Huttenhower
639 C, Xavier RJ. 2018. The human gut microbiome in early-onset type 1 diabetes from the
640 TEDDY study. Nature 562:589–594.
641 4. Tsukuda N, Yahagi K, Hara T, Watanabe Y, Matsumoto H, Mori H, Higashi K, Tsuji H,
642 Matsumoto S, Kurokawa K, Matsuki T. 2021. Key bacterial taxa and metabolic pathways
643 affecting gut short-chain fatty acid profiles in early life. ISME J
644 https://doi.org/10.1038/s41396-021-00937-7.
645 5. Laursen MF, Sakanaka M, von Burg N, Mörbe U, Andersen D, Moll JM, Pekmez CT,
646 Rivollier A, Michaelsen KF, Mølgaard C, Lind MV, Dragsted LO, Katayama T, Frandsen
647 HL, Vinggaard AM, Bahl MI, Brix S, Agace W, Licht TR, Roager HM. 2021.
648 Bifidobacterium species associated with breastfeeding produce aromatic lactic acids in the
649 infant gut. Nat Microbiol 6:1367–1382.
650 6. Raman AS, Gehrig JL, Venkatesh S, Chang H-W, Hibberd MC, Subramanian S, Kang G,
651 Bessong PO, Lima AAM, Kosek MN, Petri WA, Rodionov DA, Arzamasov AA, Leyn SA,
652 Osterman AL, Huq S, Mostafa I, Islam M, Mahfuz M, Haque R, Ahmed T, Barratt MJ,
653 Gordon JI. 2019. A sparse covarying unit that describes healthy and impaired human gut
654 microbiota development. Science 365.
655 7. Gehrig JL, Venkatesh S, Chang H-W, Hibberd MC, Kung VL, Cheng J, Chen RY,
656 Subramanian S, Cowardin CA, Meier MF, O'Donnell D, Talcott M, Spears LD,
657 Semenkovich CF, Henrissat B, Giannone RJ, Hettich RL, Ilkayeva O, Muehlbauer M,
658 Newgard CB, Sawyer C, Head RD, Rodionov DA, Arzamasov AA, Leyn SA, Osterman
659 AL, Hossain MI, Islam M, Choudhury N, Sarker SA, Huq S, Mahmud I, Mostafa I, Mahfuz
660 M, Barratt MJ, Ahmed T, Gordon JI. 2019. Effects of microbiota-directed foods in
661 gnotobiotic animals and undernourished children. Science 365.
662 8. Frese SA, Hutton AA, Contreras LN, Shaw CA, Palumbo MC, Casaburi G, Xu G, Davis JCC,
663 Lebrilla CB, Henrick BM, Freeman SL, Barile D, German JB, Mills DA, Smilowitz JT,
664 Underwood MA. 2017. Persistence of Supplemented Bifidobacterium longum subsp.
665 infantis EVC001 in Breastfed Infants. mSphere 2:e00501-17.
666 9. Casaburi G, Duar RM, Vance DP, Mitchell R, Contreras L, Frese SA, Smilowitz JT,
667 Underwood MA. 2019. Early-life gut microbiome modulation reduces the abundance of
668 antibiotic-resistant bacteria. Antimicrob Resist Infect Control 8:131.
669 10. Hirano R, Sakanaka M, Yoshimi K, Sugimoto N, Eguchi S, Yamauchi Y, Nara M, Maeda S,
670 Ami Y, Gotoh A, Katayama T, Iida N, Kato T, Ohno H, Fukiya S, Yokota A, Nishimoto M,
671 Kitaoka M, Nakai H, Kurihara S. 2021. Next-generation prebiotic promotes selective
672 growth of bifidobacteria, suppressing Clostridioides difficile. Gut Microbes 13:1973835.

673 11. Henrick BM, Rodriguez L, Lakshmikanth T, Pou C, Henckel E, Arzoomand A, Olin A,
674 Wang J, Mikes J, Tan Z, Chen Y, Ehrlich AM, Bernhardsson AK, Mugabo CH, Ambrosiani
675 Y, Gustafsson A, Chew S, Brown HK, Prambs J, Bohlin K, Mitchell RD, Underwood MA,
676 Smilowitz JT, German JB, Frese SA, Brodin P. 2021. Bifidobacteria-mediated immune
677 system imprinting early in life. Cell 184:3884-3898.e11.
678 12. Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, Tobe T, Clarke JM,
679 Topping DL, Suzuki T, Taylor TD, Itoh K, Kikuchi J, Morita H, Hattori M, Ohno H. 2011.
680 Bifidobacteria can protect from enteropathogenic infection through production of acetate.
681 Nature 469:543–547.
682 13. Alessandri G, Ossiprandi MC, MacSharry J, van Sinderen D, Ventura M. 2019.
683 Bifidobacterial Dialogue With Its Human Host and Consequent Modulation of the Immune
684 System. Front Immunol 10:2348.
685 14. Alcon-Giner C, Dalby MJ, Caim S, Ketskemety J, Shaw A, Sim K, Lawson MAE, Kiu R,
686 Leclaire C, Chalklen L, Kujawska M, Mitra S, Fardus-Reid F, Belteki G, McColl K, Swann
687 JR, Kroll JS, Clarke P, Hall LJ. 2020. Microbiota Supplementation with Bifidobacterium
688 and Lactobacillus Modifies the Preterm Infant Gut Microbiota and Metabolome: An
689 Observational Study. Cell Rep Med 1:100077.
690 15. Nguyen M, Holdbrooks H, Mishra P, Abrantes MA, Eskew S, Garma M, Oca C-G,
691 McGuckin C, Hein CB, Mitchell RD, Kazi S, Chew S, Casaburi G, Brown HK, Frese SA,
692 Henrick BM. 2021. Impact of Probiotic B. infantis EVC001 Feeding in Premature Infants
693 on the Gut Microbiome, Nosocomially Acquired Antibiotic Resistance, and Enteric
694 Inflammation. Front Pediatr 9:618009.
695 16. Bajorek S, Duar RM, Corrigan M, Matrone C, Winn KA, Norman S, Mitchell RD, Cagney
696 O, Aksenov AA, Melnik AV, Kopylova E, Perez J. 2021. B. infantis EVC001 Is Well-
697 Tolerated and Improves Human Milk Oligosaccharide Utilization in Preterm Infants in the
698 Neonatal Intensive Care Unit. Front Pediatr 9:795970.
699 17. Sela DA, Chapman J, Adeuya A, Kim JH, Chen F, Whitehead TR, Lapidus A, Rokhsar DS,
700 Lebrilla CB, German JB, Price NP, Richardson PM, Mills DA. 2008. The genome sequence
701 of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within
702 the infant microbiome. Proc Natl Acad Sci USA 105:18964–18969.
703 18. Turroni F, Bottacini F, Foroni E, Mulder I, Kim J-H, Zomer A, Sánchez B, Bidossi A,
704 Ferrarini A, Giubellini V, Delledonne M, Henrissat B, Coutinho P, Oggioni M, Fitzgerald
705 GF, Mills D, Margolles A, Kelly D, van Sinderen D, Ventura M. 2010. Genome analysis of
706 Bifidobacterium bifidum PRL2010 reveals metabolic pathways for host-derived glycan
707 foraging. Proc Natl Acad Sci USA 107:19514–19519.
708 19. Asakuma S, Hatakeyama E, Urashima T, Yoshida E, Katayama T, Yamamoto K, Kumagai
709 H, Ashida H, Hirose J, Kitaoka M. 2011. Physiology of consumption of human milk
710 oligosaccharides by infant gut-associated bifidobacteria. J Biol Chem 286:34583–34592.
711 20. James K, Motherway MO, Bottacini F, van Sinderen D. 2016. Bifidobacterium breve
712 UCC2003 metabolises the human milk oligosaccharides lacto-N-tetraose and lacto-N-neo-
713 tetraose through overlapping, yet distinct pathways. Sci Rep 6.
714 21. Ojima MN, Asao Y, Nakajima A, Katoh T, Kitaoka M, Gotoh A, Hirose J, Urashima T,
715 Fukiya S, Yokota A, Abou Hachem M, Sakanaka M, Katayama T. 2021. Diversification of
716 a fucosyllactose transporter within the genus Bifidobacterium. Appl Environ Microbiol
717 AEM0143721.

718 22. Kunz C, Rudloff S, Baier W, Klein N, Strobel S. 2000. Oligosaccharides in human milk:
719 structural, functional, and metabolic aspects. Annu Rev Nutr 20:699–722.
720 23. Bode L. 2012. Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology
721 22:1147–1162.
722 24. Ballard O, Morrow AL. 2013. Human Milk Composition: Nutrients and Bioactive Factors.
723 Pediatr Clin North Am 60:49–74.
724 25. Wu S, Tao N, German JB, Grimm R, Lebrilla CB. 2010. Development of an annotated
725 library of neutral human milk oligosaccharides. J Proteome Res 9:4138–4151.
726 26. Chen X. 2015. Human Milk Oligosaccharides (HMOS): Structure, Function, and Enzyme-
727 Catalyzed Synthesis. Adv Carbohydr Chem Biochem 72:113–190.
728 27. Sakanaka M, Gotoh A, Yoshida K, Odamaki T, Koguchi H, Xiao J-Z, Kitaoka M, Katayama
729 T. 2019. Varied Pathways of Infant Gut-Associated Bifidobacterium to Assimilate Human
730 Milk Oligosaccharides: Prevalence of the Gene Set and Its Correlation with Bifidobacteria-
731 Rich Microbiota Formation. Nutrients 12.
732 28. Katayama T, Sakuma A, Kimura T, Makimura Y, Hiratake J, Sakata K, Yamanoi T,
733 Kumagai H, Yamamoto K. 2004. Molecular cloning and characterization of
734 Bifidobacterium bifidum 1,2-alpha-L-fucosidase (AfcA), a novel inverting glycosidase
735 (glycoside hydrolase family 95). J Bacteriol 186:4885–4893.
736 29. Wada J, Ando T, Kiyohara M, Ashida H, Kitaoka M, Yamaguchi M, Kumagai H, Katayama
737 T, Yamamoto K. 2008. Bifidobacterium bifidum lacto-N-biosidase, a critical enzyme for
738 the degradation of human milk oligosaccharides with a type 1 structure. Appl Environ
739 Microbiol 74:3996–4004.
740 30. Ashida H, Miyake A, Kiyohara M, Wada J, Yoshida E, Kumagai H, Katayama T, Yamamoto
741 K. 2009. Two distinct alpha-L-fucosidases from Bifidobacterium bifidum are essential for
742 the utilization of fucosylated milk oligosaccharides and glycoconjugates. Glycobiology
743 19:1010–1017.
744 31. Miwa M, Horimoto T, Kiyohara M, Katayama T, Kitaoka M, Ashida H, Yamamoto K. 2010.
745 Cooperation of β-galactosidase and β-N-acetylhexosaminidase from bifidobacteria in
746 assimilation of human milk oligosaccharides with type 2 structure. Glycobiology 20:1402–
747 1409.
748 32. Kiyohara M, Tanigawa K, Chaiwangsri T, Katayama T, Ashida H, Yamamoto K. 2011. An
749 exo-alpha-sialidase from bifidobacteria involved in the degradation of
750 sialyloligosaccharides in human milk and intestinal glycoconjugates. Glycobiology 21:437–
751 447.
752 33. Garrido D, Kim JH, German JB, Raybould HE, Mills DA. 2011. Oligosaccharide Binding
753 Proteins from Bifidobacterium longum subsp. infantis Reveal a Preference for Host
754 Glycans. PLoS One 6.
755 34. Sakanaka M, Hansen ME, Gotoh A, Katoh T, Yoshida K, Odamaki T, Yachi H, Sugiyama Y,
756 Kurihara S, Hirose J, Urashima T, Xiao J-Z, Kitaoka M, Fukiya S, Yokota A, Lo Leggio L,
757 Abou Hachem M, Katayama T. 2019. Evolutionary adaptation in fucosyllactose uptake
758 systems supports bifidobacteria-infant symbiosis. Sci Adv 5:eaaw7696.
759 35. Sela DA, Li Y, Lerno L, Wu S, Marcobal AM, German JB, Chen X, Lebrilla CB, Mills DA.
760 2011. An infant-associated bacterial commensal utilizes breast milk sialyloligosaccharides.
761 J Biol Chem 286:11909–11918.

762 36. Garrido D, Ruiz-Moyano S, Mills DA. 2012. Release and utilization of N-acetyl-d-
763 glucosamine from human milk oligosaccharides by Bifidobacterium longum subsp. infantis.
764 Anaerobe 18:430–435.
765 37. Sela DA, Garrido D, Lerno L, Wu S, Tan K, Eom H-J, Joachimiak A, Lebrilla CB, Mills
766 DA. 2012. Bifidobacterium longum subsp. infantis ATCC 15697 α-Fucosidases Are Active
767 on Fucosylated Human Milk Oligosaccharides. Appl Environ Microbiol 78:795–803.
768 38. Yoshida E, Sakurama H, Kiyohara M, Nakajima M, Kitaoka M, Ashida H, Hirose J,
769 Katayama T, Yamamoto K, Kumagai H. 2012. Bifidobacterium longum subsp. infantis uses
770 two different β-galactosidases for selectively degrading type-1 and type-2 human milk
771 oligosaccharides. Glycobiology 22:361–368.
772 39. Garrido D, Ruiz-Moyano S, Lemay DG, Sela DA, German JB, Mills DA. 2015. Comparative
773 transcriptomics reveals key differences in the response to milk oligosaccharides of infant
774 gut-associated bifidobacteria. Sci Rep 5.
775 40. Özcan E, Sela DA. 2018. Inefficient Metabolism of the Human Milk Oligosaccharides
776 Lacto-N-tetraose and Lacto-N-neotetraose Shifts Bifidobacterium longum subsp. infantis
777 Physiology. Front Nutr 5.
778 41. Ravcheev DA, Godzik A, Osterman AL, Rodionov DA. 2013. Polysaccharides utilization in
779 human gut bacterium Bacteroides thetaiotaomicron: comparative genomics reconstruction
780 of metabolic and regulatory networks. BMC Genomics 14:873.
781 42. Khoroshkin MS, Leyn SA, Van Sinderen D, Rodionov DA. 2016. Transcriptional Regulation
782 of Carbohydrate Utilization Pathways in the Bifidobacterium Genus. Front Microbiol 7.
783 43. Rodionov DA, Rodionova IA, Rodionov VA, Arzamasov AA, Zhang K, Rubinstein GM,
784 Tanwee TNN, Bing RG, Crosby JR, Nookaew I, Basen M, Brown SD, Wilson CM,
785 Klingeman DM, Poole FL, Zhang Y, Kelly RM, Adams MWW. 2021. Transcriptional
786 Regulation of Plant Biomass Degradation and Carbohydrate Utilization Genes in the
787 Extreme Thermophile Caldicellulosiruptor bescii. mSystems e0134520.
788 44. James K, O'Connell Motherway M, Penno C, O'Brien RL, van Sinderen D. 2018.
789 Bifidobacterium breve UCC2003 Employs Multiple Transcriptional Regulators To Control
790 Metabolism of Particular Human Milk Oligosaccharides. Appl Environ Microbiol 84.
791 45. Kazanov MD, Li X, Gelfand MS, Osterman AL, Rodionov DA. 2013. Functional
792 diversification of ROK-family transcriptional regulators of sugar catabolism in the
793 Thermotogae phylum. Nucleic Acids Res 41:790–803.
794 46. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers
795 differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids
796 Res 43:e47.
797 47. Warchol M, Perrin S, Grill J-P, Schneider F. 2002. Characterization of a purified beta-
798 fructofuranosidase from Bifidobacterium infantis ATCC 15697. Lett Appl Microbiol
799 35:462–467.
800 48. Plumbridge JA. 1991. Repression and induction of the nag regulon of Escherichia coli K-12:
801 the roles of nagC and nagA in maintenance of the uninduced state. Mol Microbiol 5:2053–
802 2062.
803 49. Plumbridge J, Kolb A. 1991. CAP and Nag repressor binding to the regulatory regions of the
804 nagE-B and manX genes of Escherichia coli. J Mol Biol 217:661–679.
805 50. Nishimoto M, Kitaoka M. 2007. Identification of N-Acetylhexosamine 1-Kinase in the
806 Complete Lacto-N-Biose I/Galacto-N-Biose Metabolic Pathway in Bifidobacterium
807 longum. Appl Environ Microbiol 73:6444–6449.

808 51. Stiers KM, Muenks AG, Beamer LJ. 2017. Biology, mechanism, and structure of enzymes in
809 the α-D-phosphohexomutase superfamily. Adv Protein Chem Struct Biol 109:265–304.
810 52. Barreteau H, Kovac A, Boniface A, Sova M, Gobec S, Blanot D. 2008. Cytoplasmic steps of
811 peptidoglycan biosynthesis. FEMS Microbiol Rev 32:168–207.
812 53. Garrido D, Nwosu C, Ruiz-Moyano S, Aldredge D, German JB, Lebrilla CB, Mills DA.
813 2012. Endo-β-N-acetylglucosaminidases from infant gut-associated bifidobacteria release
814 complex N-glycans from human milk glycoproteins. Mol Cell Proteomics 11:775–785.
815 54. Shani G, Hoeflinger JL, Heiss BE, Masarweh CF, Larke JA, Jensen NM, Wickramasinghe S,
816 Davis JC, Goonatilleke E, El-Hawiet A, Nguyen L, Klassen JS, Slupsky CM, Lebrilla CB,
817 Mills DA. 2021. Fucosylated human milk oligosaccharide foraging within the species
818 Bifidobacterium pseudocatenulatum is driven by glycosyl hydrolase content and specificity.
819 Appl Environ Microbiol AEM0170721.
820 55. Leyn SA, Li X, Zheng Q, Novichkov PS, Reed S, Romine MF, Fredrickson JK, Yang C,
821 Osterman AL, Rodionov DA. 2011. Control of Proteobacterial Central Carbon Metabolism
822 by the HexR Transcriptional Regulator. J Biol Chem 286:35782–35794.
823 56. Ravcheev DA, Khoroshkin MS, Laikova ON, Tsoy OV, Sernova NV, Petrova SA,
824 Rakhmaninova AB, Novichkov PS, Gelfand MS, Rodionov DA. 2014. Comparative
825 genomics and evolution of regulons of the LacI-family transcription factors. Front
826 Microbiol 5.
827 57. Arzamasov AA, van Sinderen D, Rodionov DA. 2018. Comparative Genomics Reveals the
828 Regulatory Complexity of Bifidobacterial Arabinose and Arabino-Oligosaccharide
829 Utilization. Front Microbiol 9:776.
830 58. Lanigan N, Kelly E, Arzamasov AA, Stanton C, Rodionov DA, van Sinderen D. 2019.
831 Transcriptional control of central carbon metabolic flux in Bifidobacteria by two
832 functionally similar, yet distinct LacI-type regulators. Sci Rep 9:17851.
833 59. Egan M, Jiang H, O'Connell Motherway M, Oscarson S, van Sinderen D. 2016.
834 Glycosulfatase-Encoding Gene Cluster in Bifidobacterium breve UCC2003. Appl Environ
835 Microbiol 82:6611–6623.
836 60. Katoh T, Maeshibu T, Kikkawa K-I, Gotoh A, Tomabechi Y, Nakamura M, Liao W-H,
837 Yamaguchi M, Ashida H, Yamamoto K, Katayama T. 2017. Identification and
838 characterization of a sulfoglycosidase from Bifidobacterium bifidum implicated in mucin
839 glycan utilization. Biosci Biotechnol Biochem 81:2018–2027.
840 61. Novichkov PS, Laikova ON, Novichkova ES, Gelfand MS, Arkin AP, Dubchak I, Rodionov
841 DA. 2010. RegPrecise: a database of curated genomic inferences of transcriptional
842 regulatory interactions in prokaryotes. Nucleic Acids Res 38:D111–D118.
843 62. Mironov AA, Vinokurova NP, Gelfand MS. 2000. Software for analysis of bacterial
844 genomes. Mol Biol 34:222–231.
845 63. Rodionov DA. 2007. Comparative Genomic Reconstruction of Transcriptional Regulatory
846 Networks in Bacteria. Chem Rev 107:3467–3497.
847 64. Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: A Sequence Logo
848 Generator. Genome Res 14:1188–1190.
849 65. Bottacini F, Zomer A, Milani C, Ferrario C, Lugli GA, Egan M, Ventura M, van Sinderen D.
850 2017. Global transcriptional landscape and promoter mapping of the gut commensal
851 Bifidobacterium breve UCC2003. BMC Genomics 18.
852 66. Kozakai T, Izumi A, Horigome A, Odamaki T, Xiao J, Nomura I, Suzuki T. 2020. Structure
853 of a Core Promoter in Bifidobacterium longum NCC2705. J Bacteriol 202.

854 67. Erb I, González-Vallinas JR, Bussotti G, Blanco E, Eyras E, Notredame C. 2012. Use of
855 ChIP-Seq data for the design of a multiple promoter-alignment method. Nucleic Acids Res
856 40:e52.
857 68. Hirayama Y, Sakanaka M, Fukuma H, Murayama H, Kano Y, Fukiya S, Yokota A. 2012.
858 Development of a Double-Crossover Markerless Gene Deletion System in Bifidobacterium
859 longum: Functional Analysis of the α-Galactosidase Gene for Raffinose Assimilation. Appl
860 Environ Microbiol 78:4984–4994.
861 69. Rey FE, Faith JJ, Bain J, Muehlbauer MJ, Stevens RD, Newgard CB, Gordon JI. 2010.
862 Dissecting the in vivo metabolic potential of two human gut acetogens. J Biol Chem
863 285:22082–22090.
864 70. Amorim CF, Novais FO, Nguyen BT, Misic AM, Carvalho LP, Carvalho EM, Beiting DP,
865 Scott P. 2019. Variable gene expression and parasite load predict treatment outcome in
866 cutaneous leishmaniasis. Sci Transl Med 11.
867 71. Ritz C, Baty F, Streibig JC, Gerhard D. 2015. Dose-Response Analysis Using R. PLoS One
868 10:e0146021.
869 72. Kitaoka M, Tian J, Nishimoto M. 2005. Novel Putative Galactose Operon Involving Lacto-
870 N-Biose Phosphorylase in Bifidobacterium longum. Appl Environ Microbiol 71:3158–
871 3162.
872 73. Barratt M, Nuzhat S, Ahsan K, Frese SA, Arzamasov AA, Sarker SA, Islam MM, Palit P,
873 Islam MR, Hibberd MC, Nakshatri S, Cowardin CA, Guruge JL, Byrne AE, Venkatesh S,
874 Sundaresan V, Henrick B, Duar RM, Mitchell RD, Casaburi G, Prambs J, Flannery R,
875 Mahfuz M, Rodionov DA, Osterman AL, Kyle D, Ahmed T, Gordon J. 2022.
876 Bifidobacterium longum subsp. infantis strains for treating severe acute malnutrition in
877 Bangladeshi infants. Sci Transl Med in press.
878

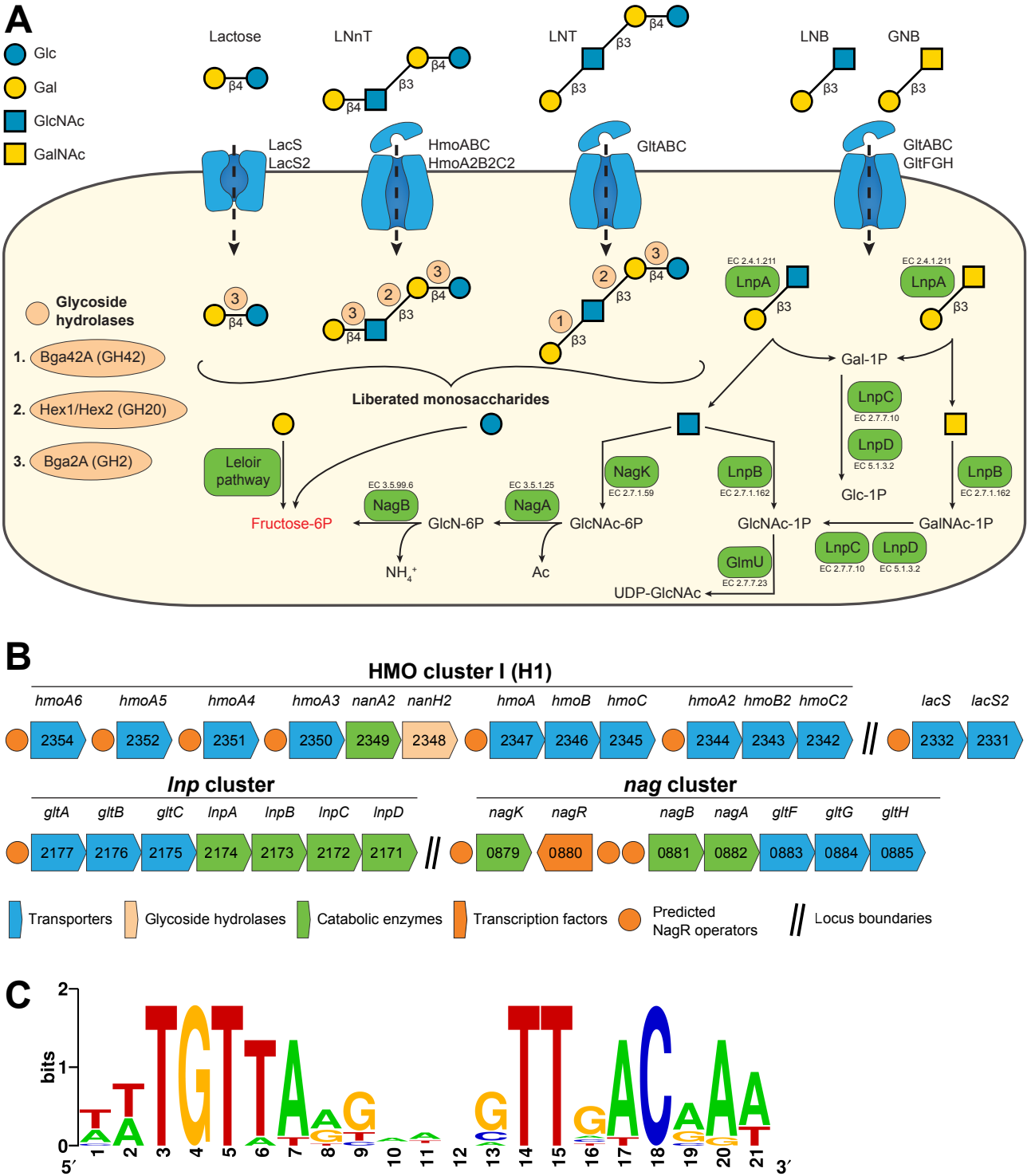



**FIG 1**. Reconstructed NagR regulon in *B. infantis* ATCC 15697. (A) Schematic representation of LNT, LNnT, and LNB/GNB utilization pathways in *B. infantis* ATCC 15697. *Step 1*: HMOs and their constituents are transported into the cell by various transport systems. *Step 2*: once inside the cell, HMOs are degraded from the non-reducing end by a coordinated action of exo-acting GHs. Breakdown of glycosidic bonds by specific GHs is indicated in light orange circles. *Step 3*: released monosaccharides are converted to fructose-6P and enter the bifid shunt. (B) Genomic loci constituting the reconstructed NagR regulon in *B. infantis* ATCC 15697. Numbers represent locus tags in the Blon_XXXX format (GenBank accession no. CP001095.1). (C) NagR-binding motif in *B. infantis* ATCC 15697 based on 11 predicted operator sequences.

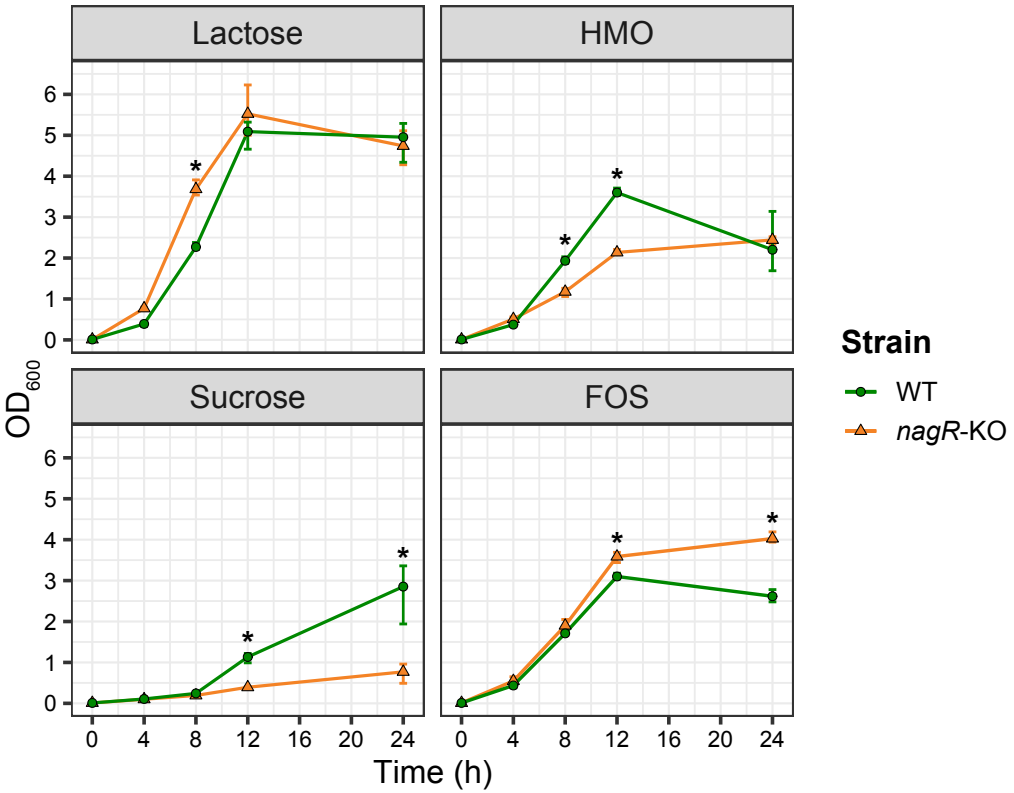



**FIG 2**. Growth curves of *B. infantis* ATCC 15697 WT and *nagR*-KO strains in MRS-CS supplemented with various carbon sources (1% w/v). Growth was monitored by measuring $OD_{600}$ at specific time points. Data points represent the mean of three biological replicates. Error bars depict 95% confidence intervals for the mean obtained by bootstrapping. Timepoints where the $OD_{600}$ values for WT and *nagR*-KO strains were significantly different (*, Padj<0.05) were identified using a linear regression. Bonferroni correction was used to adjust for multiple testing.

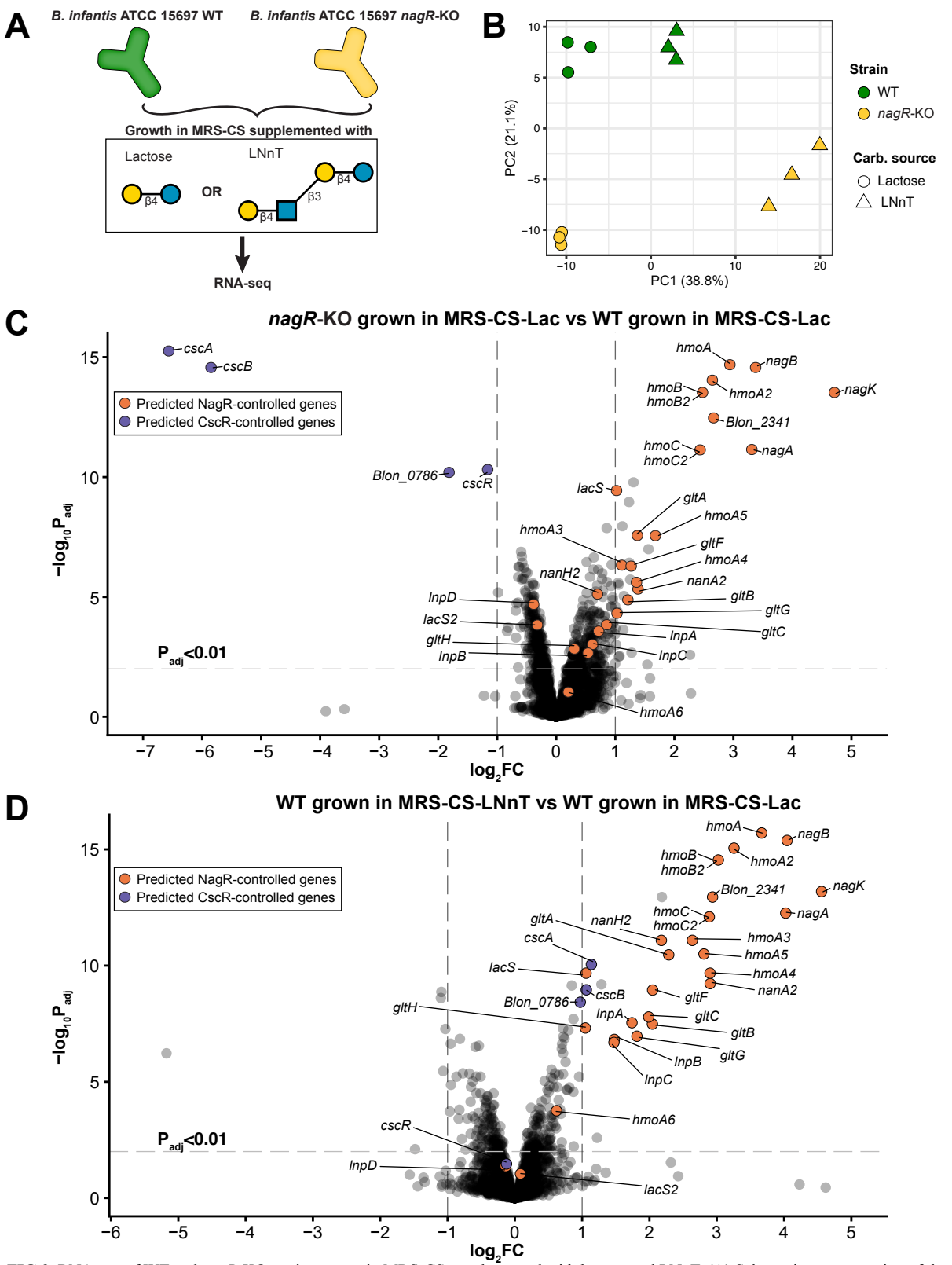



**FIG 3**. RNA-seq of WT and *nagR*-KO strains grown in MRS-CS supplemented with lactose and LNnT. (A) Schematic representation of the experimental design. (B) PCA of TMM-normalized count data. Each data point represents one sample. (C) and (D) Volcano plots depicting the $\log_2$FC of gene expression versus the $-\log_{10}$Padj. (C) compares *nagR*-KO and WT strains grown in MRS-CS-Lac, whereas (D) compares the WT strain grown in MRS-CS-LNnT and MRS-CS-Lac. Criteria for calling differentially expressed genes were the follows: Padj < 0.01 and absolute FC > 2. Genes constituting the predicted NagR and CscR regulons are labeled and colored in orange and purple, respectively.

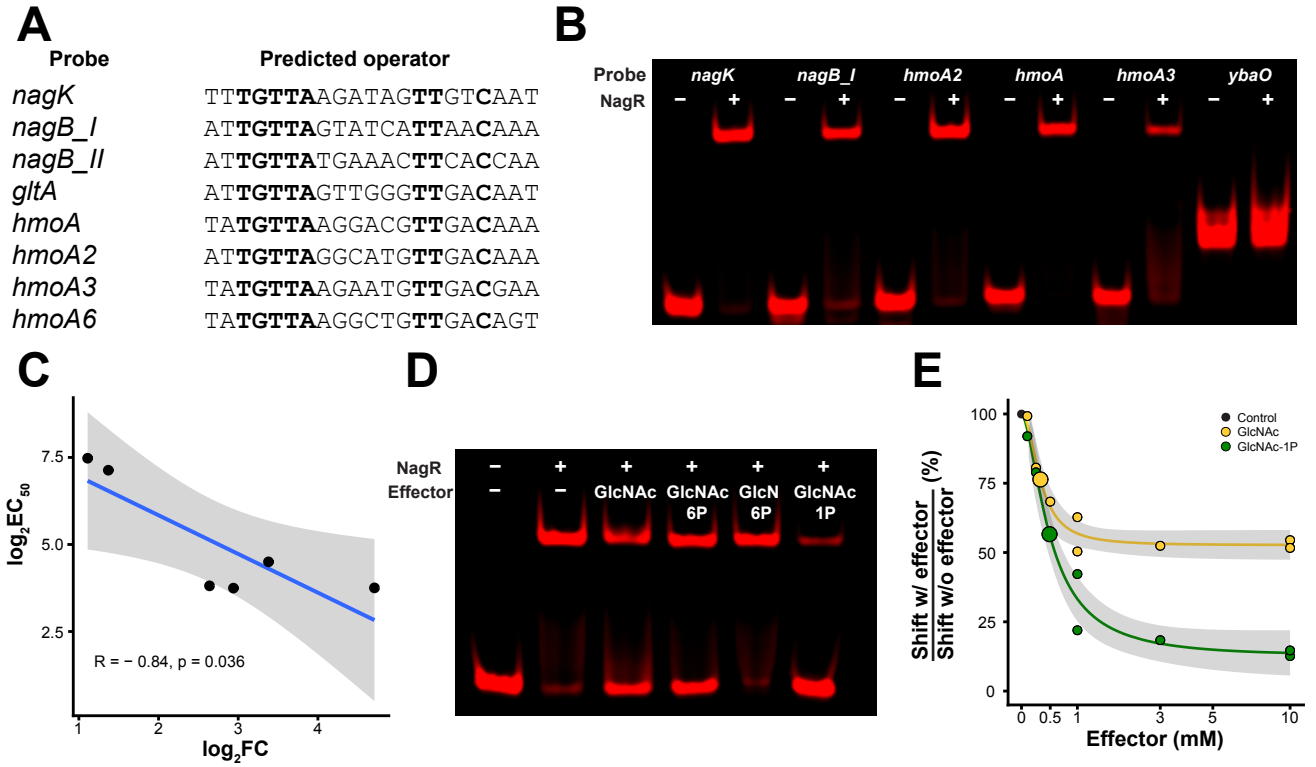
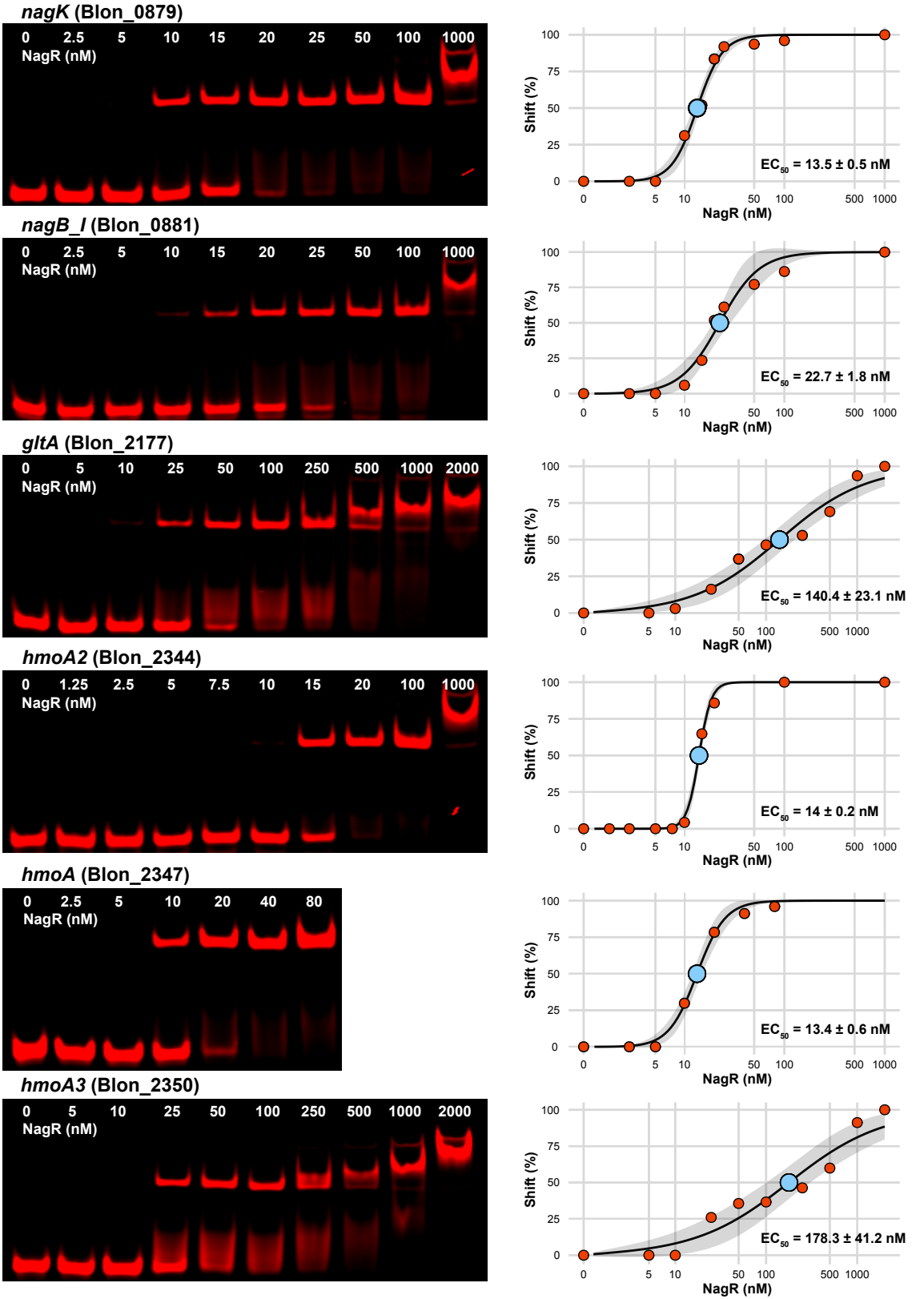



**FIG 4**. Interactions of recombinant NagR with predicted operators and screening of NagR effectors. (A) Predicted NagR operator sequences identified in the promoter regions of listed genes. Conserved nucleotides are in bold. The full sequences of probes used in EMSA experiments are given in Table S1. (B) An EMSA gel showing interactions of NagR with DNA probes containing predicted operators. NagR concentration was 25 nM, and probe concentrations were 1 nM. The *ybdO* probe was used as a control for non-specific binding. (C) Correlation between probe $EC_{50}$ values determined via EMSA and expression fold-changes of cognate genes in the RNA-seq experiment (*nagR*-KO vs. WT grown in MRS-CS-Lac). (D) Effect of various GlcNAc metabolism intermediates (10 mM) on the interaction between NagR (25 nM) and the *hmoA* probe (1 nM). (E) $EC_{50}$ values of selected NagR effector molecules. The y-axis depicts the ratio (% shift with effector )/(% shift w/o effector). Calculated $EC_{50}$ are shown as big circles. 25 nM of NagR and 1 nM of the *hmoA* probe were used.

**FIG 5**. EMSA gels depicteing titration of DNA probes (1 nM) containing predicted NagR operators with recombinant NagR. Gels were quantified, and the results were approximated by the 4PL equation. Protein concentration when half of the probe is shifted ($EC_{50}$) is shown. Grey shading depicts the 95% confidence interval. The x-axis is in the Log10 scale.



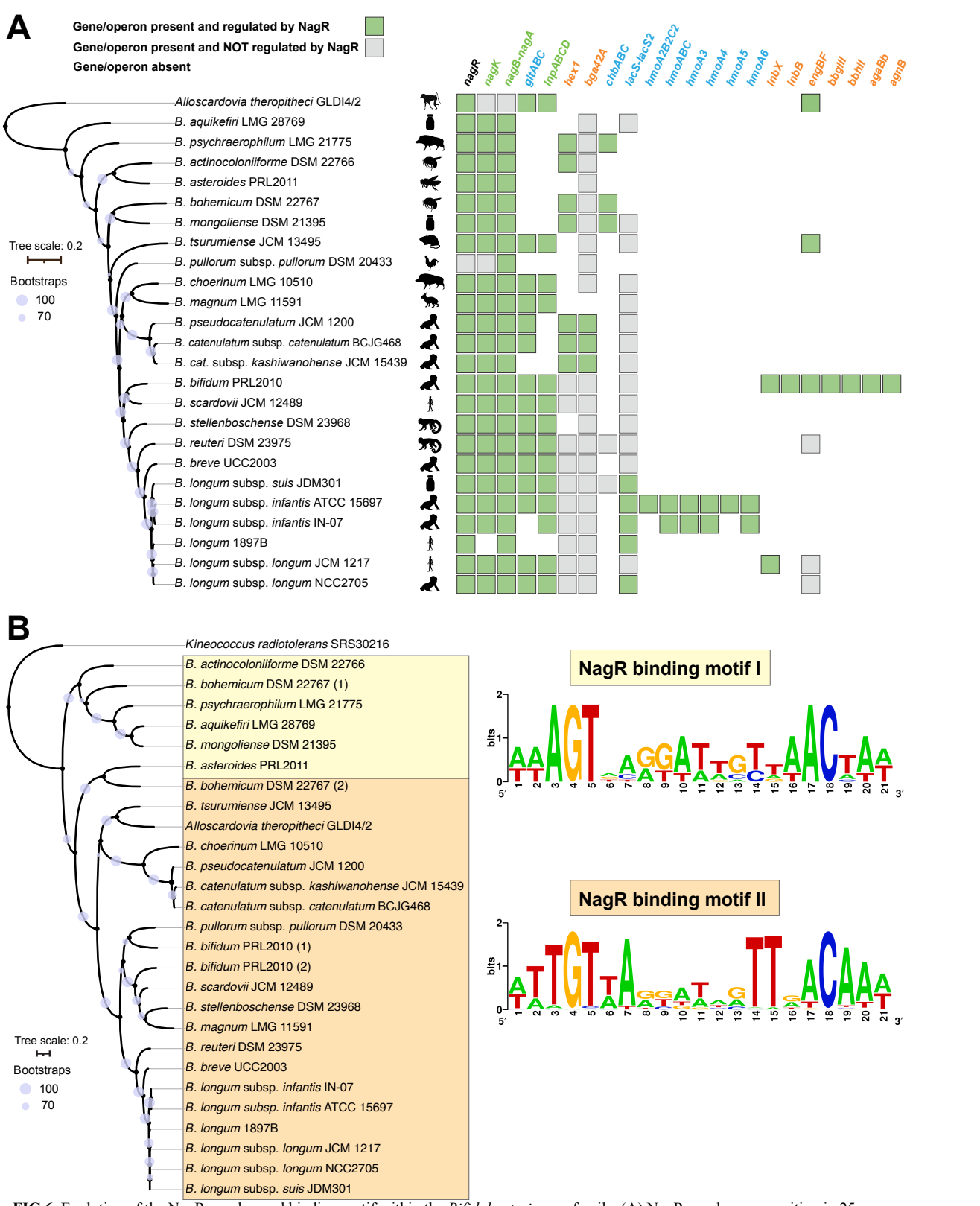



**FIG 6**. Evolution of the NagR regulon and binding motif within the *Bifidobacteriaceae* family. (A) NagR regulon composition in 25 *Bifidobacteriaceae* strains mapped on a species tree built based on the alignment of 247 core genes. Bootstrap values are shown as purple circles. Black symbols indicate strain isolation sources. Regulon members are colored according to their function: catabolic enzymes are green, GHs in orange, and transporters in blue. (B) NagR-binding motifs mapped on a tree of NagR proteins from 25 *Bifidobacteriaceae* strains. Bootstrap values are shown as purple circles. NagR paralogs in *B. bohemicum* and *B. bifidum* are denoted by numbers (1 and 2).

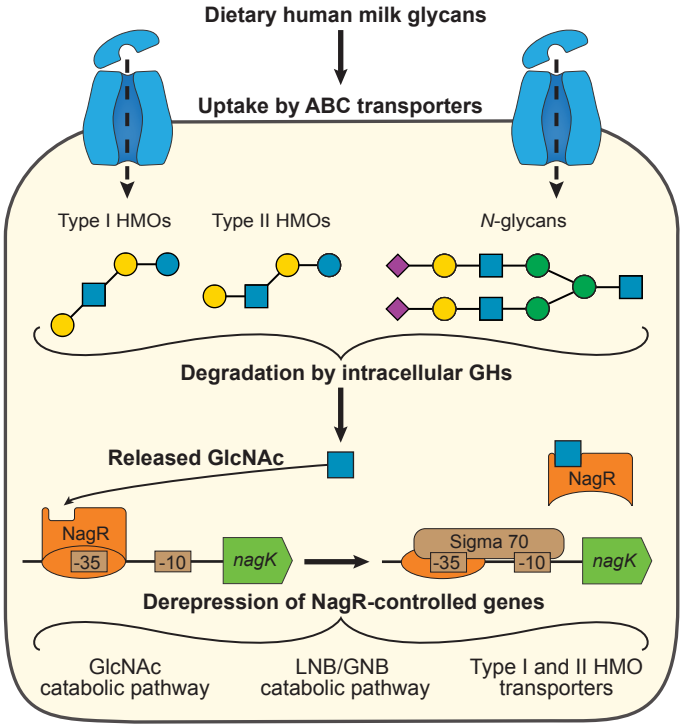



**FIG 7**. Model of NagR-mediated regulation of HMO utilization in *B. infantis*. *Step 1*: GlcNAc-containing milk glycans (e.g., LNT and LNnT) are uptaken into the cell by various ABC transporters. *Step 2*: once inside the cell, the glycans are degraded by intracellular GHs, and GlcNAc is released. *Step 3*: released GlcNAc interacts with NagR and disrupts the NagR-operator complex, leading to derepression of NagR-controlled genes.