CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Bifidobacterium breve UCC2003 employs multiple transcriptional regulators to control metabolism of particular human milk oligosaccharides
Author(s)	James, Kieran; O'Connell Motherway, Mary; Penno, Christophe; O'Brien, Rebecca L.; van Sinderen, Douwe
Publication date	2018-03-02
Original citation	James, K., O'Connell Motherway, M., Penno, C., O'Brien, R. L. and van Sinderen, D. (2018) 'Bifidobacterium breve UCC2003 employs multiple transcriptional regulators to control metabolism of particular human milk oligosaccharides', Applied and Environmental Microbiology. doi:10.1128/aem.02774-17
Type of publication	Article (peer-reviewed)
Link to publisher's version	http://dx.doi.org/10.1128/aem.02774-17 Access to the full text of the published version may require a subscription.
Rights	© 2018, American Society for Microbiology. All Rights Reserved.
Item downloaded from	http://hdl.handle.net/10468/5673

University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

1   *Bifidobacterium breve* **UCC2003 employs multiple transcriptional regulators to control**

2   **metabolism of particular human milk oligosaccharides.**

3

4   **Authors: Kieran James[1,2], Mary O'Connell Motherway[2], Christophe Penno[2], Rebecca**

5   **Louise O'Brien[1,2], Douwe van Sinderen[1,2*]**

6

7   **Affiliation(s): School of Microbiology[1] & APC Microbiome Institute[2], University**

8   **College Cork, Cork, Ireland**

9

10   *** Corresponding author:** **d.vansinderen@ucc.ie**

11

12

13   **Running Title:** Control of bifidobacterial HMO metabolism

14

15   **Key words:** Bifidobacteria, probiotic, prebiotic, transcriptional regulation, HMO,

16   carbohydrate metabolism.

Accepted Manuscript Posted Online

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

17 **Abstract**

18 Bifidobacterial carbohydrate metabolism has been studied in considerable detail for a variety

19 of both plant and human-derived glycans, particularly involving the bifidobacterial prototype

20 *Bifidobacterium breve* UCC2003. We recently elucidated the metabolic pathways by which

21 the human milk oligosaccharide (HMO) constituents lacto-N-tetraose (LNT), lacto-N-

22 neotetraose (LNnT) and lacto-N-biose (LNB) are utilized by *B. breve* UCC2003. However, to

23 date no work has been carried out on the regulatory mechanisms that control expression of

24 the genetic loci involved in these HMO metabolic pathways. In the current study, we describe

25 the characterization of three transcriptional regulators and corresponding operator and

26 associated (inducible) promoter sequences, the latter governing transcription of the genetic

27 elements involved in LN(n)T/LNB metabolism. The activity of these regulators is dependent

28 on the release of specific monosaccharides, which are believed to act as allosteric effectors,

29 and which are derived from the corresponding HMOs targeted by the particular locus.

30

31 **Importance**

32 Human milk oligosaccharides (HMOs) are a key factor in the development of the breastfed

33 infant microbiota. They function as prebiotics, selecting for a specific range of microbes,

34 including a number of infant-associated species of bifidobacteria, which are thought to

35 provide a range of health benefits to the infant host. While much research has been carried

36 out on elucidating the mechanisms of HMO metabolism in infant-associated bifidobacteria,

37 there is to date very little understanding of the transcriptional regulation of these pathways.

38 The current study reveals a multi-component transcriptional regulation system that controls

39 the recently-identified pathways of HMO metabolism in the infant-associated

40 *Bifidobacterium breve* prototype strain UCC2003. This not only provides insight into the

41 regulatory mechanisms present in other infant-associated bifidobacteria, but also provides an

42 example of a network of sequential steps regulating microbial carbohydrate metabolism.

43

2

44 **Introduction**

45 Bifidobacteria represent high-G+C, Gram-positive, anaerobic members of the phylum
46 Actinobacteria, and are common commensals of the mammalian, avian and insect gut. In
47 humans, they are particularly abundant and prevalent among the gut microbiota of healthy,
48 vaginally-delivered, breastfed infants (2), and are thought to confer a multitude of benefits to
49 the neonatal host (3-5). For this reason, as well as because of their purported health-
50 promoting activities in adults, bifidobacteria are used as functional ingredients in a variety of
51 foods and therapeutic products. The use of prebiotics is also becoming commonplace for the
52 improvement of both adult and infant (gut) health. A prebiotic has been defined as 'a non-
53 digestible food ingredient that beneficially affects the host by selectively stimulating the
54 growth and/or activity of one or a limited number of bacteria in the colon, and thus improves
55 host health' (6).

56 Fascinatingly, the archetypal prebiotic would appear to be human breastmilk, and in
57 particular its bifidogenic constituents known as H̲uman M̲ilk O̲ligosaccharides (HMOs).
58 HMOs represent specific glycans present in human breastmilk that are thought to shape, at
59 least partly, the compositional structure of the neonatal gut microbiota (7, 8). HMOs
60 represent, after lactose, the second-largest carbohydrate component of breastmilk (7, 9), and
61 constitute a heterogeneous mix of at least 200 distinct glycan structures (10). The majority of
62 complex HMO structures can be classified into one of two types, depending on their
63 backbone composition. The more abundant Type I HMOs contain the core tetrasaccharide
64 lacto-N-tetraose (LNT) within their structure (Galβ1-3GlcNAcβ1-3Galβ1-4Glc). Type II
65 HMOs contain lacto-N-neotetraose (LNnT), a stereoisomer of LNT, within their backbone
66 (Galβ1-4GlcNAcβ1-3Galβ1-4Glc). Lacto-N-biose (LNB; Galβ1-3GlcNAc) is a subunit of
67 LNT and other Type I HMO structures, and can be released by the degradation of these
68 sugars (11).

69 The effects of the specific human breastmilk components on the prevalence, abundance and
70 activity of members of the infant gut microbiota are currently enjoying a great deal of
71 scientific and commercial attention, due to the beneficial roles they are believed to play in
72 infant health and development (12, 13). Understanding the pathways by which specific
73 HMOs are metabolized by particular microbial species that inhabit the infant gut is important,
74 although our knowledge regarding these processes is still in its infancy, particularly with
75 regards to the manner in which they affect microbiota development.

76   It is hardly surprising that the dominant *Bifidobacterium* species found among the neonatal

77   gut microbiota can utilise various HMO components as their sole carbohydrate source (10).

78   These species chiefly include strains of *Bifidobacterium bifidum*, *Bifidobacterium longum*

79   subsp. *infantis* and *Bifidobacterium breve*. HMO utilization by *B. bifidum* and *B. longum*

80   subsp. *infantis* is relatively well characterised. *B. bifidum* extracellularly hydrolyses complex

81   HMO structures, including LNT and LNnT, employing secreted glycosyl hydrolases,

82   followed by the internalisation and intracellular degradation/metabolism of (most of) the

83   resulting mono- and di-saccharides, such as LNB (11, 14-20). *B. longum* subsp. *infantis*

84   internalises intact LNT, LNnT and LNB, and uses a series of sequential

85   hydrolytic/phosphorolytic reactions acting from the non-reducing end of the carbohydrate

86   structures to degrade them into their monosaccharide components for further metabolic

87   processing (10, 11, 21-24). However, *B. infantis* has also been demonstrated to take up and

88   utilise fucosyl- and sialyl-lactose (25-27).

89   The metabolic pathways of LNT, LNnT and LNB have recently been elucidated in the

90   prototype strain *B. breve* UCC2003 (28). In the latter study, converging pathways of LNT

91   and LNnT catabolism were identified, where monosaccharide moieties are sequentially

92   released from the non-reducing end of either sugar by hydrolytic reactions. The genetic units

93   responsible for the uptake and breakdown of these structures are the *lnt* locus (corresponding

94   to locus tags Bbr_0526-0530) and the *nah* locus (locus tags Bbr_1554-1560) (Table 3 and

95   Fig. 1). The *lnt* locus encodes proteins that are responsible for the internalisation of LNT and

96   intracellular hydrolysis of both LNT and LNnT, releasing a galactose (Gal) moiety from their

97   non-reducing end, and at the same time liberating the trisaccharide lacto-N-triose

98   (GlcNAcβ1-3Galβ1-4Glc). The *nah* locus specifies an LNT/LNnT uptake system, while it

99   furthermore encodes a glycosyl hydrolase which liberates GlcNAc from the non-reducing end

100   of lacto-N-triose, leaving lactose, which itself is further degraded by lactose-specific glycosyl

101   hydrolases. Additionally, the gene products of the *lnp/glt* locus (corresponding to locus tags

102   Bbr_1585-1590; Table 3 and Fig. 1) are responsible for the internalisation and subsequent

103   phosphorolysis of free LNB, releasing its constituent monosaccharides Gal-6-phosphate and

104   GlcNAc (28, 29). We also identified the transcriptional upregulation of genes in the *nag* locus

105   (locus tags Bbr_1247-1252; Table 3 and Fig. 1) during growth on LNT, LNnT and LNB,

106   indicating their role in the utilisation of these sugars, specifically in the multi-step

107   metabolism of GlcNAc. The *nag* locus has previously been implicated in the metabolism of

108   sialic acid and mucin-derived N-glycans, both of which contain GlcNAc as well (30, 31).

4

109   While the degradation routes of these key HMO structures have thus been identified, the
110   regulatory mechanisms that control expression of these pathways have remained unexplored,
111   both for *B. breve* and HMO-utilising *Bifidobacterium* species as a whole.

112   In the current study, we identified and characterized the genes encoding transcriptional
113   regulators responsible for control of gene expression in four key HMO-associated loci in *B.*
114   *breve* UCC2003 during growth on LNT, LNnT or LNB.

115

116   **Results**

117   **Identification of putative transcriptional regulator-encoding genes in the vicinity of**
118   **HMO-utilisation loci.** In a previous study, we had observed that genes within four
119   chromosomal loci exhibit transcriptional induction during growth of *B. breve* UCC2003 on
120   LNT, LNnT or LNB as the sole carbohydrate source (28). This indicates that these genes are
121   subject to transcriptional regulation, which was presumed to be either directly or indirectly
122   controlled by the presence of these HMO substrates. The four loci concerned are: the *lnt*
123   locus (Bbr_0526-530), the *nah* locus (Bbr_1554-1560), the *nag* locus (Bbr_1247-1250) and
124   the *lnp/glt* locus (Bbr_1585-1590) (Fig. 1; see Table 3 for a description of [predicted]
125   functions). Detailed scrutiny of these four loci and neighbouring regions showed that the *lnt*
126   and *nah* loci are flanked by or contain a predicted regulator-encoding gene, respectively: *lntR*
127   (Bbr_0526), encoding a LacI-type repressor, and *nahR* (Bbr_1555), encoding a NagC/XylR-
128   type repressor (Fig. 1A and 1B). The *nag* locus is associated with two genes, *nagR1*
129   (Bbr_1249) and *nagR2* (Bbr_1251), both predicted to encode ROK/NagC family-type
130   repressors, while no regulator-encoding gene was observed in close vicinity of the *lnp/glt*
131   *locus* (Fig. 1C and 1D). NagC/XylR-type and ROK/NagC-type repressors are both members
132   of the large family of ROK-type transcriptional regulators (32). The four identified putative
133   regulator-encoding genes were thus selected as candidates for mutagenesis in order to
134   ascertain their role, if any, in the transcriptional regulation of the *lnt*, *nah*, *nag* and *lnp/glt*
135   loci.

136

137   **Generation and transcriptomic analysis of insertional mutants in putative HMO-**
138   **associated regulator-encoding genes.** Individual insertional mutants were constructed in
139   *lntR*, *nahR*, *nagR1* and *nagR2*, resulting in *B. breve* strains UCC2003-lntR, UCC2003-nahR,
140   UCC2003-nagR1 and UCC2003-nagR2, respectively (see Materials and Methods). In order

5

141   to identify promoters/genes that are subject to transcriptional control of these predicted

142   regulators, global gene transcription data was obtained from microarray-based analyses

143   performed on the *B. breve* UCC2003-lntR, UCC2003-nahR, UCC2003-nagR1 and

144   UCC2003-nagR2 insertion mutants, grown in mMRS supplemented with ribose, as compared

145   to the transcriptome of the UCC2003 wild type strain grown under the same conditions.

146   Transcriptome analysis of the *lntR* mutant revealed the upregulation of the adjacent *lntP1*,

147   *lntP2*, *lntA* and *lntS* genes of the *lnt* locus (Table 3 and Fig. 1), when this mutant was grown

148   on ribose (as compared to wild type UCC2003), all of which were also previously found to be

149   upregulated in expression during growth of wild type UCC2003 on LNT or LNnT (28). This

150   corroborates the notion that LntR is a LacI-type repressor and that this protein negatively

151   regulates the LNT/LNnT-dependent transcription of genes within the *lnt* cluster. Conversely,

152   the array data obtained for the *nahR* mutant only revealed transcriptional upregulation

153   (compared to the UCC2003 control) of the *nahS* gene (Table 3), when grown on ribose. This

154   is consistent with previously observed expression patterns in UCC2003, with the exception of

155   *nahA*, which may have been expected to exhibit transcriptional upregulation in the *nahR*

156   mutant, as its expression was increased during growth on LNT and LNnT in *B. breve*

157   UCC2003 (28). These results suggest that NahR, a NagC/XylR-type repressor, is responsible

158   for the transcriptional regulation of at least one gene of the *nah* cluster. For the *nagR1*

159   mutant, upregulation of *nagA2*, *nagB3* and *nagK* (but not *nagR2* or *nagK2*), as well as all of

160   the genes of the *lnp/glt* locus (Table 3) was observed when grown on ribose (as compared to

161   the UCC2003 control). These results suggest that NagR1, a ROK/NagC family-type

162   repressor, is responsible for the transcriptional regulation of (part of) the *nag* and *lnb* clusters.

163   This is consistent with transcriptomic data previously obtained for wild type UCC2003

164   during growth on LN(n)T and LNB, which demonstrated the transcriptional upregulation of

165   genes in both of these loci (28). When the transcriptome of UCC2003-nagR2 was compared

166   to that of UCC2003 when grown on ribose, the *nagR2* mutant exhibited increased

167   transcription of genes in the *mal* locus (locus tags Bbr_0118-0123), which is known to be

168   involved in maltooligosaccharide metabolism (33, 34), and Bbr_1719-1721 (predicted to

169   function in fatty acid metabolism) (35) (data not shown), none of which are predicted to

170   function in HMO metabolism, nor were shown to be upregulated in our previous wild type

171   arrays on LNT, LNnT or LNB (28). These results thus show that NagR2 is not involved in

172   the transcriptional control of the loci responsible for LNT, LNnT or LNB metabolism, and no

173   further investigation of this regulator was carried out. The *lntR*, *nahR* and *nagR1* genes,

174 however, were selected for further study, as described below, in order to further elucidate
175 their regulatory activity and specificity.

176

177 **Promoter mapping through identification of the transcription start sites.** Based on the
178 transcriptome findings, we presumed that LntR, NahR and NagR1 act as transcriptional
179 regulators of (certain genes of the) *lnt*, *nah* and *nag/lnp/glt* loci, respectively. Gene
180 expression patterns observed for the regulator gene mutants, and examination of the genetic
181 layout and transcriptome profiles of these loci allowed us to assign putative promoter-
182 containing regions within each locus. In order to verify these predicted promoter regions, the
183 associated transcription start sites (TSS) were experimentally determined by primer extension
184 analyses.

185 The *lnt* locus was deduced to contain at least two promoters: one just upstream of *lntP1* (Fig.
186 2A) and one in front of *lntS* (Fig. 2B). The *lntP1* and *lntS* genes on the *B. breve* UCC2003
187 genome encode a permease and a solute-binding protein of an ABC-transporter system,
188 respectively, and exhibit an increase in transcription upon growth on LNT, LNnT, LNB,
189 lactosamine or lactose (28). The transcription start sites (TSS) of the presumed *lntP1* and *lntS*
190 promoters were determined by primer extension analysis using RNA extracted from *B. breve*
191 UCC2003 grown in mMRS supplemented with 1 % LNnT. An extension product was
192 identified 41 nucleotides 5' of the predicted translational start site of the *lntP1* gene
193 (Supplemental Fig. S1A), while the TSS for the *lntS* gene was identified 154 nucleotides 5'
194 of the predicted translational start site (Supplemental Fig. S1B). In both cases, the TSS was
195 preceded by -10 and -35 hexamers that resemble (bifidobacterial) consensus vegetative
196 promoter recognition sequences (36, 37).

197 The *nah* locus was deduced to contain at least two promoters: one just upstream of *nahS* (Fig.
198 2C) and one in front of *nahA* (Fig. 2D). The *nahS* and *nahA* genes on the *B. breve* UCC2003
199 genome encode a solute-binding protein of an ABC-transporter system, and a GH20 *N*-
200 acetylhexosaminidase, respectively. While an increase in transcription was only observed for
201 *nahS* in the *nahR* mutant-based array, both this gene and *nahA* were found to be subject to
202 transcriptional induction when wild-type UCC2003 is grown on LNT, LNnT or lactosamine
203 (28). The TSSs of the presumed *nahS* and *nahA* promoters were determined by primer
204 extension analysis using RNA extracted from *B. breve* UCC2003 grown in mMRS
205 supplemented with 1 % LNnT. An extension product was identified 59 nucleotides 5' of the

206 predicted translational start site for the *nahS* gene (Supplemental Fig. S1C), while the TSS

207 upstream of *nahA* was identified 74 nucleotides 5' of the predicted *nahA* translational start

208 site (Supplemental Fig. S1D). The *nahS* upstream region contained a -10 and a -35 hexamer

209 just upstream of the TSS resembling bifidobacterial promoter sequences (36, 37), while in the

210 case of the *nahA* promoter region the TSS is preceded by a sequence that resembles a

211 canonical -10 promoter sequence, although no associated -35 hexamer could be identified.

212 The *nag* and *lnp/glt* loci were each deduced to contain at least two promoters, just upstream

213 of the genes: *nagB3* (Fig. 2E) and *nagK* (Fig. 2F), and *lnpB* (Fig. 2G) and *gltA* (Fig. 2H),

214 respectively, based on the associated genetic lay-out coupled to transcription patters of the

215 *nagR1* mutant or when UCC2003 was grown on LNB (28). The transcription start sites (TSS)

216 of the presumed *nagB3, nagK*, *lnpB* and *gltA* promoters were determined by primer extension

217 analysis using RNA extracted from *B. breve* UCC2003 grown in mMRS supplemented with 1

218 % LNB. An extension product was identified 155 nucleotides 5' of the predicted translational

219 start site for the *nagB3* gene (Supplemental Fig. S1E), while the transcriptional start site of

220 *nagK* was identified 35 nucleotides 5' of the predicted translational start site (Supplemental

221 Fig. S1F). An extension product was identified 43 nucleotides 5' of the predicted

222 translational start site for the *lnpB* gene (Supplemental Fig. S1G), while the transcription start

223 site for the *gltA* gene was identified 44 nucleotides 5' of the predicted translational start site

224 (Supplemental Fig. S1H). All four regions contained -10 and -35 hexamers just upstream of

225 the TSS that resembled bifidobacterial vegetative promoter recognition sequences.

226

227 **Identification of regulator-operator interactions using electromobility shift assays and**

228 ***in silico* analysis.** In order to establish if the LntR, NahR and NagR1 proteins directly and

229 specifically interact with operator sequences within the identified promoter regions of the *lnt*,

230 *nah*, and *nag*/*lnp*/*glt* gene clusters, respectively, electrophoretic mobility shift assays

231 (EMSAs) were performed. For the purpose of performing EMSAs, the *lntR*, *nahR* and *nagR1*

232 genes were first individually cloned into the nisin-inducible vector pNZ8150 with an N-

233 terminal His tag-encoding sequence to facilitate protein expression and purification in *L.*

234 *lactis* NZ9000 (see Materials and Methods). As had been noted previously for other

235 regulators from bifidobacteria (38-40), LntR, NahR and NagR1 could be obtained as purified

236 proteins, but had lost their DNA binding activity during some stage of the purification

237 process. Thus, instead of purified protein, crude cell extracts of (nisin-induced) *L. lactis*

238 NZ9000 pNZ-*lntR*_{His,} *L. lactis* NZ9000 pNZ-*nahR*_{His} and *L. lactis* NZ9000 pNZ-*nagR1*_{His}

239   were used to carry out the EMSAs. Crude cell extract obtained from nisin-induced *L. lactis*

240   NZ9000 pNZ8150 (empty vector) incubated with the respective DNA fragments was used as

241   a negative control. The DNA fragments used were various short amplicons representing

242   different segments of the putative promoter regions (Fig. 2, Supplemental Table S1).

243   LntR-containing crude extract was shown to specifically bind to the IRD700-labelled DNA

244   fragments lntP1a and lntP1b, but not with lntP1c (Fig. 2A, Supplemental Table S1). A double

245   mobility-shift was observed for fragment lntP1a, indicative of two distinct LntR binding sites

246   being present on this fragment, while a single mobility shift was visible for fragment lntP1b.

247   Similarly, LntR was able to bind to IRD700-labelled DNA fragments lntSb and lntSa, in the

248   latter case being visible as a double mobility-shift (suggesting the presence of two distinct

249   LntR-binding sites), while no binding was observed with lntSc (Fig. 2B, Supplemental Table

250   S1). Inspection and comparison of the four fragments in which binding was observed

251   revealed the presence of at least one complete conserved sequence, representing an inverted

252   repeat, in all four fragments, while two such conserved sequences were observed in

253   fragments lntP1a and lntSa (being consistent with the observed double mobility-shift).

254   Comparative analysis of these inverted repeats identified a 14-nucleotide consensus sequence

255   (Fig. 3A), containing a conserved 'CG' at its centre, which is a well-documented conserved

256   feature of operator sequences bound by LacI-type regulators (41, 42). This consensus

257   sequence furthermore contains a conserved 5' 'TG' and 3' 'CA' at its flanking ends, a feature

258   previously documented in operator sequences identified for other LacI-type regulators

259   encoded by *B. breve* UCC2003 (39, 43, 44). In both promoter regions, one such presumed

260   operator sequence was found closely downstream of or partially overlapping the predicted -

261   10 element of the promoter region, while the second was found closely upstream of the

262   predicted -35 element of the promoter region (Supplemental Fig. S1A, S1B). The positions of

263   these identified operators are consistent with LntR acting as a repressor for the identified *lnt*

264   promoters (45, 46).

265   The results obtained with the *L. lactis* NZ9000 pNZ-*nahR*$_{His}$ crude extract demonstrated

266   specific binding to the IRD700-labelled DNA fragments nahSa and nahSb, but not with

267   nahSc (Fig. 2C, Supplemental Table S1). Furthermore, binding was observed for IRD700-

268   labelled DNA fragment nahAa, but not to fragment nahAb or nahAc (Fig. 2D, Supplemental

269   Table S1). Sequence inspection and comparison of the NahR-bound DNA fragments revealed

270   the presence of an inverted repeat sequence, which was common to these fragments, yet not

271   present in fragments to which NahR did not bind. These inverted repeat elements therefore

9

272   represent putative operator sequences required for the NahR protein. Further analysis of these

273   inverted repeats identified a 10-nucleotide consensus sequence (Fig. 3B). Conserved 5' C and

274   3' G nucleotides at the extreme flanks of this consensus sequence have previously been

275   observed for operator sequences of certain NagC/XylR-type regulators (47). The presumed

276   operator upstream of *nahS* overlaps with the downstream end of the predicted -10 promoter

277   element (Supplemental Fig. S1C), while the *nahA*-associated operator was found to be

278   roughly 110 bp upstream of the predicted -10 element (Supplemental Fig. S1D). The position

279   of the identified *nahS* operator, and the consensus obtained between this and the putative

280   operator identified for *nahA*, confirm the function of NahR as a repressor for *nahS*. While

281   binding of NahR may occur at the *nahA* operator, its binding does not appear to directly

282   interfere with the *nahA* promoter. This agrees with the lack of upregulation in *nahA*

283   expression observed for the *nahR* mutant, though the transcriptional role of NahR in this case,

284   if any, is not clear. The mapped locations of the operator, -10 and -35 sequences are shown in

285   Supplemental Figure S1.

286   The results obtained with the crude extract obtained from nisin-induced *L. lactis* NZ900 pNZ-

287   *nagR1*$_{His}$ revealed specific binding to the IRD700-labelled DNA fragments nagB3a, nagB3b

288   and nagB3c (with a weak apparent double-shift observed for nagB3a), but not to fragment

289   nagB3d (Fig. 2E, Supplemental Table S1). Specific binding was identified for IRD700-

290   labelled DNA fragment nagKa, while no binding was detected when fragment nagKb was

291   used (Fig. 2F, Supplemental Table S1). Binding of the NagR protein was also demonstrated

292   for the IRD700-labelled DNA fragments lnpBa and lnpBc, but not with lnpBb (Fig. 2G,

293   Supplemental Table S1). Finally, NagR1 was shown to bind IRD700-labelled DNA

294   fragments gltAa, gltAb and gltAc (Fig. 2H, Supplemental Table S1). Inspection and

295   comparison of the *nagB3*, *nagK*, *lnpB*, and gltA-associated fragments in which binding was

296   observed revealed the presence of a common sequence, representing an inverted repeat (with

297   two repeats present in the fragment nagB3a, consistent with the observed double-shift), while

298   being absent within fragments for which no binding was observed. These sequence motifs are

299   presumed to act as operator sequences for the NagR1 protein. *In silico* analysis of these

300   inverted repeat sequences revealed a 23-nucleotide consensus (Fig. 3C). Interestingly, while

301   this obtained consensus motif bears little resemblance to many previously proposed binding

302   motifs for ROK/NagC family-type repressors from other bacteria (48), a substantial degree of

303   similarity can be observed to motifs identified previously for other ROK/NagC-type

304   regulators encoded by *B. breve* UCC2003 (31, 39). The putative *nagB3*, *nagK*, *lnpB* and *gltA*

305   operators were all found to be overlapping with or encompassing the predicted -10 or -35
306   elements (Supplemental Fig. S1E, S1F, S1G, S1H). The positions of these identified
307   operators corroborate the notion that NagR1 acts as a transcriptional repressor of its target
308   genes (i.e. *nagB3*, *nagK*, *lnpB* and *gltA*).

309

310   **Identification of transcriptional effectors.** In order to identify effectors that control the
311   binding activity of LntR, NahR and NagR1, we performed EMSAs with fragments containing
312   the binding motifs for each regulator, in the presence of a range of carbohydrates, including:
313   lactose, LNB, LNT, LNnT, galactose, galactose-6-phosphate (Gal-6-P), galactose-1-
314   phosphate (Gal-1-P), GlcNAc, GalNAc or glucose (at a standard concentration of 20 mM)
315   (Supplemental Fig. S2). These carbohydrates were chosen as they include both the complete
316   structures and various components of LNT, LNnT or LNB. Carbohydrates which did not
317   elicit any effect on fragment binding by the regulator (at a concentration of 20 mM) were
318   assumed not to represent transcriptional effectors for that particular regulator. If an inhibition
319   in binding was observed at 20 mM, the EMSA was repeated at a range of descending
320   concentrations (or in some cases higher concentrations were used for a related molecule [e.g.
321   Gal, Gal-1-P and Gal-6-P]). For LntR, galactose was found to reduce binding of this regulator
322   to its DNA targets at a concentration of 10 mM or less (Fig. 4A). Gal-6-P and Gal-1-P were
323   also found to reduce target DNA binding of LntR, but at considerably higher, and perhaps
324   biologically irrelevant concentrations of ≥20 mM (Supplemental Figure S3). For NahR, only
325   GlcNAc was found to reduce interaction between NahR and its DNA target at a minimum
326   concentration of 0.0625 mM (Fig. 4B), while in the case of NagR1, Gal-6-P was found to
327   prevent NagR1-binding activity at a minimum concentration of 1 mM (Fig. 4C, 4D).

328

329   **Discussion**

330   The dominance of (certain) bifidobacteria within the breastfed neonatal gut microbiota (2) is
331   substantially aided by the ability of these infant-associated species to utilise indigestible
332   HMO residues as a carbon source (10). Our previous work demonstrated that consumption
333   and utilisation of LNT, LNnT or LNB by *B. breve* UCC2003 is facilitated by interrelated
334   catabolic pathways (28). While pathways for HMO utilisation in other *Bifidobacterium*
335   species have been identified and elucidated (14, 22, 23), very little work has been carried out
336   with regard to their regulation. Our results reveal molecular details of the transcriptional

11

337   regulation of *B. breve* UCC2003 loci responsible for LN(n)T/LNB metabolism, and provide

338   insights into how metabolism of these HMOs is controlled in *B. breve* UCC2003.

339   In the current study, we identified four transcriptional regulators, three of which were shown

340   to be involved in regulating LN(n)T/LNB metabolism in UCC2003. Microarray analysis of

341   insertional mutants in *lntR*, *nahR*, *nagR1* and *nagR2* identified genes under the regulation of

342   each encoded regulator. LntR and NahR were shown to represent 'local' regulators, i.e.

343   controlling transcription of genes adjacent to *lntR* and *nahR*, respectively. In contrast, NagR1

344   regulates transcription of not only the 'local' *nag* locus, but also of the genetically unlinked

345   *lnp/glt* locus. We also investigated the transcriptome effect of a mutation in Bbr_1251

346   (*nagR2*), however, the affected genes are not believed to be involved in HMO metabolism,

347   but apparently in malto-oligosaccharide and fatty acid metabolism. While LacI-type,

348   NagC/XylR-type and ROK/NagC-type regulators have all previously been identified and

349   characterised in *B. breve* UCC2003 (31, 38, 39, 43), functional analysis of regulators in other

350   bifidobacteria is comparatively undocumented. However, a recent study identified

351   homologous transcription factors for those of LntR, NagR1and NagR2 in a range of different

352   *Bifidobacterium* species (49).

353   Details of promoter and operator sequences specific to the LntR, NahR and NagR1 regulators

354   were elucidated using a combination of electromobility shift and primer extension analyses.

355   These operator results, for the most part, agree with those predicted by Khoroshkin et al. (49).

356   The operator sequences predicted in their study concur with our experimentally determined

357   data, both in approximate location and number, for both LntR and NagR1, with the exception

358   of one additional predicted operator for LntR, and two for NagR1. An additional NagR1

359   operator sequence was predicted upstream of *gltA*, however, this did not appear to be

360   functional, based on the lack of a double mobility-shift in the EMSA's for that region. This

361   operator may indeed be a non-functional relic resulting from a duplication event. Khoroshkin

362   et al. (49) also predicted an operator sequence upstream of the gene Bbr_1884 for NagR1

363   binding, though we did not examine this. However, based on the predicted functions of this

364   gene in the Bifid Shunt, it may also be tied into the overall regulation of LNB and LacNAc

365   metabolism carried out by NagR1. An additional LntR operator was predicted upstream of

366   *lntR* itself, which may function in *lntR* transcriptional auto-regulation. The observed lack of

367   upregulation in *nahA* transcription for the *nahR* mutant appears discordant with the

368   transcriptional increase of this gene that was previously observed in wild type UCC2003

369   during growth on LN(n)T (28), as well as the presence of the functional *nahA* operator

370 sequence for NahR binding identified in this study. However, this may be explained if
371 transcriptional induction of *nahA* is mediated by both LntR and NahR. This possibility is
372 corroborated by the presence of an inverted repeat sequence resembling an LntR operator and
373 this intriguing possibility merits further experimental investigation.

374 Perhaps most interestingly of all was the identification of the effectors for each
375 transcriptional regulator. Binding of LntR to its targets is impeded by Gal, NahR-mediated
376 operator binding is prevented by the presence of GlcNAc, while NagR1-operator interaction
377 is prevented by the presence of Gal-6-P. In each case, the genes under transcriptional control
378 by their respective regulator encode the metabolic machinery responsible for the release of
379 the effector monosaccharide from the substrate at that metabolic step. For example, Gal is
380 released from the non-reducing end of LN(n)T through the hydrolytic activity of LntA, which
381 is encoded by the *lnt* locus (28). Transcriptional repression of this locus is thus believed to be
382 relieved by the presence of the released monosaccharide, which is presumed to interact with
383 the allosteric effector site typical of LacI-type repressors (39, 45, 50, 51). A similar scenario
384 applies to GlcNAc release which acts as the effector for the NahR regulator that controls
385 transcription of *nahS*, and to Gal-6-P which governs the activity of the NagR1, the presumed
386 transcriptional regulator of the *lnp/glt* and *nag* loci. The possible dual regulation of *nahA*
387 transcription, as mentioned above, would mean that both the presence of the *lnt* locus activity
388 product (and LntR effector) galactose, and the *nah* locus activity product (and NahR effector)
389 GlcNAc, are required for the induction of *nahA* expression. This provides an extra level of
390 transcriptional and thus metabolic control, ensuring the expression of *nahA* strictly during
391 LN(n)T metabolism, despite GlcNAc release during metabolism of other sugars, such as
392 LNB, sialic acid and sulphated GlcNAc (28, 30, 31). Interestingly, in the case of Gal-6-P and
393 NagR1, the *lnp/glt* locus is required for the degradation of LNB, though the *nag* locus is not
394 directly involved in this, yet plays a role in N-acetylglucosamine (GlcNAc) and sialic acid
395 metabolism (30). This may not be surprising, as sialic acid residues are commonly found in
396 HMO (7), and more importantly, GlcNAc is a breakdown product of LNB (as well as LNT
397 and LNnT).

398 Interestingly, previous work has shown that transcriptional induction takes place of the *lnt*
399 locus during growth of UCC2003 on galacto-oligosaccharides (GOS) (52). This would appear
400 to disagree with the high degree of specificity of transcriptional induction by effectors of
401 these HMO-associated loci. However, it is worth noting that GOS consist mainly of galactose

13

402 (53, 54), and that the intracellular release of galactose during GOS metabolism by UCC2003

403 would be sufficient to cause transcriptional induction of the *lnt* locus.

404 Thus, the presence and initial degradation of such a structure (i.e. LNT, LNnT or LNB)

405 indirectly induces further expression of the locus required for its degradation, until the sugar

406 is no longer available, at which point the absence of inducers will cause a return to

407 transcriptional repression. Initial internalisation and degradation is likely facilitated by a low

408 level of 'leaky' gene expression of the locus. In the case of LNT and LNnT degradation, this

409 regulation is a two-step process, at the level of LN(n)T degradation first (by the *lnt* locus),

410 and then at the level of (LN(n)T breakdown product) lacto-N-triose degradation (by the *nah*

411 locus). Regulation of LNB metabolism is managed in a single step, at the level of LNB

412 phosphorolysis (by the *lnp/glt* locus). We see that all three regulators in this transcriptional

413 control network belong to distinct families of regulator proteins, despite functioning in

414 similar roles as saccharide-controlled repressors. In conclusion, our results reveal a tightly

415 controlled system for transcriptional regulation of genes encoding the metabolic machinery

416 required for (certain) HMO metabolism in *B. breve* UCC2003. Such tight regulation is

417 necessary for infant-associated bifidobacteria such as *B. breve*, where switching metabolic

418 processing to and from milk-derived sugars such as HMO and lactose, and plant-derived

419 carbohydrate sources (55) is a regular occurrence during the weaning period. Moreover, this

420 suggests the evolution of specific catabolic responses to the presence of and for the utilisation

421 of specific HMO moieties by *B. breve*, and poses the question as to whether such regulatory

422 systems have similarly evolved in other infant-associated *Bifidobacterium* species.

423

424 **Materials and Methods**

425 **Bacterial strains, plasmids, and culture conditions.** Bacterial strains and plasmids used in

426 this study are listed in Table 1. *B. breve* UCC2003 was routinely cultured in either de Man

427 Rogosa and Sharpe medium (MRS medium; Difco, BD, Le Pont de Claix, France)

428 supplemented with 0.05 % cysteine-HCl or reinforced clostridial medium (RCM; Oxoid Ltd.,

429 Basingstoke, England). Growth of bifidobacterial strains for transcriptional and primer

430 extension analyses was carried out in modified de Man Rogosa and Sharpe (mMRS) medium,

431 which was prepared from first principles (56), and which does not contain a fixed

432 carbohydrate source. Prior to inoculation, the mMRS medium was supplemented with

433 cysteine-HCl (0.05 %, wt/vol) and a particular carbohydrate source (1 %, wt/vol). It has

14

434   previously been shown that mMRS does not support growth of *B. breve* UCC2003 in the
435   absence of an added carbohydrate (57). Carbohydrates used were ribose (Sigma Aldrich,
436   Steinheim, Germany), LNB (Elicityl Oligotech, Crolles, France) and LNnT (Glycom,
437   Lyngby, Denmark). A 1 % wt/vol concentration of carbohydrate was considered sufficient to
438   encourage adequate growth for RNA harvesting. The addition of these carbohydrates did not
439   significantly alter the pH of the medium, and therefore subsequent pH adjustment was not
440   required.

441   *B. breve* cultures were incubated under anaerobic conditions in a modular atmosphere-
442   controlled system (Davidson and Hardy, Belfast, Ireland) at 37°C. *Lactococcus lactis* strains
443   were cultivated in M17 broth (Oxoid Ltd., Basingstoke, England) containing 0.5 % glucose
444   (58) at 30°C. *Escherichia coli* strains were cultured in Luria-Bertani (LB) broth (59) at 37°C
445   with agitation. Where appropriate, growth media contained tetracycline (Tet; 10 μg ml^{-1}),
446   chloramphenicol (Cm; 5 μg ml^{-1} for *L. lactis* and *E. coli*, 2.5 μg ml^{-1} for *B. breve*),
447   erythromycin (Em; 100 μg ml^{-1}) or kanamycin (Kan; 50 μg ml^{-1}). Recombinant *E. coli* EC101
448   cells containing (derivatives of) pORI19 were selected on LB agar containing Em and Kan,
449   and supplemented with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) (40 μg
450   ml^{-1}) and 1 mM IPTG (isopropyl-β-D-galactopyranoside).

451

452   **Nucleotide sequence analysis.** Sequence information was obtained from the Artemis-
453   mediated (60) genome annotations of *B. breve* UCC2003 (61). Database searches were
454   performed using non-redundant sequences accessible at the National Center for
455   Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov) using the basic local
456   alignment search tool (BLAST) (62, 63). Sequences were verified and analysed using the
457   SeqMan and SeqBuilder programs of the DNAStar software package (version 10.1.2;
458   DNAStar, Madison, WI, USA).

459

460   **DNA Manipulations.** Chromosomal DNA was isolated from *B. breve* UCC2003 as
461   previously described (64). Plasmid DNA was isolated from *Escherichia coli*, *Lactococcus*
462   *lactis* and *B. breve* using the Roche High Pure Plasmid Isolation kit (Roche Diagnostics,
463   Basel, Switzerland). An initial lysis step was performed using 30 mg ml^{-1} of lysozyme for 30
464   minutes at 37°C prior to plasmid isolation from *L. lactis* or *B. breve*. Procedures for DNA
465   manipulations were essentially performed as described previously (59). All restriction

15

466   enzymes and T4 DNA ligase were used according to the supplier's instructions (Roche

467   Diagnostics, Basel, Switzerland). Synthetic single stranded oligonucleotide primers used in

468   this study (Table 2) were synthesized by Eurofins (Ebersberg, Germany). Standard PCRs

469   were performed using Taq PCRmaster mix (Qiagen) or Extensor Hi-Fidelity PCR Master

470   Mix (Thermo Scientific, Waltham, United States) in a Biometra T3000 thermocycler

471   Biometra, Göttingen, Germany) or a Life Technologies Proflex PCR System (Thermo

472   Scientific, Waltham, United States). PCR products were visualized by ethidium bromide

473   (EtBr) staining following agarose gel electrophoresis (1 % agarose). *B. breve* colony PCR

474   reactions were performed as described previously (65). PCR fragments were purified using

475   the Roche high Pure PCR purification kit (Roche Diagnostics, Basel, Switzerland). Plasmid

476   DNA was isolated using the Roche High Pure Plasmid Isolation kit (Roche Diagnostics,

477   Basel, Switzerland). Plasmid DNA was introduced into *E. coli* by electroporation as

478   described previously (59). *B. breve* UCC2003 (66) and *L. lactis* (67) were transformed by

479   electroporation according to published protocols. Correct orientation of DNA inserts and

480   integrity of all plasmid constructs (see also below) were verified by DNA sequencing,

481   performed at Eurofins (Ebersberg, Germany).

482

483   **Construction of *B. breve* UCC2003 insertion mutants.** Internal fragments of Bbr_0526

484   (designated here as *lntR*) (367 base pairs [bp] representing codon numbers 40 through to 162

485   of the 320 codons of this gene), Bbr_1249 (designated here as *nagR1*) (502 bp representing

486   codon numbers 64 through to 231 of the 375 codons of this gene), Bbr_1251 (designated here

487   as *nagR2*) (507 bp representing codon numbers 62 through to 230 of the 405 codons of this

488   gene), and Bbr_1555 (designated here as *nahR*) (448 bp representing codon numbers 74

489   through to 223 of the 380 codons of this gene) were amplified by PCR using *B. breve*

490   UCC2003 chromosomal DNA as a template and primer pairs 526LacIInsFHindIII and

491   526LacIInsRXbaI, 1249LacIInsFHindIII and 1249LacIInsRXbaI, 1251LacIInsFHindIII and

492   1251LacIInsRXbaI, or 1555LacIInsFHindIII and 1555LacIInsRXbaI (Table 2), respectively.

493   The insertion mutants were constructed using a previously described approach (65),

494   generating mutant strains *B. breve* UCC2003-lntR, *B. breve* UCC2003-nagR1, *B. breve*

495   UCC2003-nagR2 and *B. breve* UCC2003-nahR, which carried disrupted *lntR*, *nagR1*, *nagR2*

496   and *nahR* genes, respectively (Table 1). Site-specific recombination of potential tet-resistant

497   mutant isolates was confirmed by colony PCR using primer combinations TetWF and TetWR

498   to verify *tetW* gene integration, and primers Bbr_526ConfirmP1 or Bbr_526ConfirmP2,

499   Bbr_1249ConfirmP1 or Bbr_1249ConfirmP2, Bbr_1251ConfirmP1 or Bbr_1251ConfirmP2,

500   and Bbr_1555ConfirmP1 or Bbr_1555ConfirmP2 (positioned upstream of the selected

501   internal fragments of Bbr_0526, Bbr_1249, Bbr_1251 and Bbr_1555, respectively) in

502   combination with primer TetWF to confirm integration at the correct chromosomal location

503   (Table 2).

504

505   **Analysis of global gene expression using *B. breve* DNA microarrays.** Global gene

506   expression was determined during log-phase growth of the insertional mutant strains *B. breve*

507   UCC2003-*lntR*, *B. breve* UCC2003-*nagR1*, *B. breve* UCC2003-*nagR2* and *B. breve*

508   UCC2003-*nahR* in mMRS supplemented with ribose. The generated transcriptome data sets

509   were compared to the transcriptome information obtained for log-phase wild-type *B. breve*

510   UCC2003 cells when grown in mMRS supplemented with ribose. Ribose was selected as a

511   suitable transcriptomic reference, as the metabolic pathway and gene expression profile for

512   growth of UCC2003 on ribose is known and has been employed previously as a reference

513   (44, 68). DNA microarrays containing oligonucleotide primers representing each of the 1864

514   identified open reading frames on the genome of *B. breve* UCC2003 were designed and

515   obtained from Agilent Technologies (Palo Alto, Ca., USA). Methods for cell disruption,

516   RNA isolation, RNA quality control, complementary DNA synthesis and labelling were

517   performed as, described previously (69). Two independent biological replicates were used for

518   each array using a Cy3/Cy5 dye-swap, as described previously (69). Labelled cDNA was

519   hybridized using the Agilent Gene Expression hybridization kit (part number 5188-5242) as

520   described in the Agilent Two-Colour Microarray-Based Gene Expression Analysis v4.0

521   manual (publication number G4140-90050). Following hybridization, microarrays were

522   washed in accordance with Agilent's standard procedures and scanned using an Agilent DNA

523   microarray scanner (model G2565A). Generated scans were converted to data files with

524   Agilent's Feature Extraction software (Version 9.5). DNA-microarray data were processed as

525   previously described (70-72). Differential expression tests were performed with the Cyber-T

526   implementation of a variant of the t-test (73).

527

528   **Construction of overexpression vectors, protein overproduction and purification.** For

529   the construction of plasmids pNZ-lntR, pNZ-nagR1 and pNZ-nahR, DNA fragments

530   encompassing *lntR*, *nagR1* and *nahR* were generated by PCR amplification from

17

531 chromosomal DNA of *B. breve* UCC2003 using Q5 High-Fidelity DNA polymerase and
532 primer combinations 526PurFSmaI and 526PurRXbaI, 1249PurFPvuII and 1249PurRXbaI,
533 or 1555PurFEcoRV and 1555PurXbaI, respectively (Table 2). An in-frame N-terminal His10-
534 encoding sequence was incorporated into the forward primers 526PurFSmaI, 1249PurFPvuII
535 and 1555PurFEcoRV to facilitate downstream protein purification. The generated amplicons
536 were digested with SmaI and XbaI, PvuII and XbaI, or EcoRV and XbaI, respectively, and
537 ligated into the ScaI and XbaI digested, nisin-inducible translational fusion plasmid pNZ8150
538 (74). The ligation mixtures were introduced into *L. lactis* NZ9000 by electrotransformation
539 and transformants were then selected based on chloramphenicol (Cm) resistance. The plasmid
540 content of a number of Cm-resistant transformants was screened by restriction analysis and
541 the integrity of positively identified clones was verified by sequencing.

542 Nisin-inducible gene expression and protein overproduction was performed as described
543 previously (38, 43, 75). In brief, 50 ml of M17 broth supplemented with 0.5 % (wt/vol)
544 glucose was inoculated with a 2 % inoculum of a particular *L. lactis* strain, followed by
545 incubation at 30°C until an OD600 of 0.5 was reached, at which point protein expression was
546 induced by addition of cell-free supernatant of a nisin-producing strain (76), followed by
547 continued incubation for a further 2 hours. Cells were harvested by centrifugation, and crude
548 cell extract was obtained as described previously (39). Although protein purification of LntR-
549 His, NahR-His and NagR1-His was achieved using His tag affinity chromatography, the
550 purification procedure appeared to render the proteins inactive in subsequent electrophoretic
551 mobility shift assays (EMSAs). For this reason, crude cell extracts, prepared in a 10 mM Tris-
552 HCl lysis buffer (pH 7.0), were adopted for the EMSAs (see below).

553

554 **Electrophoretic mobility shift assay (EMSA).** DNA fragments representing different
555 portions of the promoter regions upstream of *lntP1* (locus tag Bbr_0527) and *lntS* (locus tag
556 Bbr_0530), *nagB3* (locus tag Bbr_1248) and *nagK* (locus tag Bbr_1250), *lnpB* (locus tag
557 Bbr_1586) and *gltA* (locus tag Bbr_1590) and *nahS* (Bbr_1554) and *nahA* (Bbr_1556) were
558 prepared by PCR using IRD700-labelled primers pairs (Integrated DNA Technologies,
559 Coralville, Indiana, United States) (Table 2). EMSAs were performed essentially as described
560 previously (43, 77). In all cases, binding reactions were carried out in a final volume of 20 μl
561 in the presence of poly[d(I-C)] in binding buffer (20 mM Tris-HCl, 5 mM MgCl2, 0.5 mM
562 DTT, 1 mM EDTA, 100 mM KCl, 10 % glycerol). Varying amounts of crude protein extract,
563 ranging from 140 ng to 180 ng, of the LntR-, NahR-, or NagR1-(over)producing *L. lactis*

18

564 NZ9000 strain constructed, and a fixed amount of DNA probe (0.1 pmol) was mixed on ice

565 and subsequently incubated for 15 min at 37°C. In order to assess if the binding activity of

566 LntR, NahR, or NagR1 is modulated by a carbohydrate ligand, various carbohydrates

567 including galactose, galactose-1-phosphate, galactose-6-phosphate (all Sigma Aldrich,

568 Steinheim, Germany), LNT (Glycom, Lyngby, Denmark), LNnT (Glycom, Lyngby,

569 Denmark), LNB (Elicityl Oligotech, Crolles, France), glucose, N-acetylglucosamine, N-

570 acetylglucosamine-6-phosphate or lactose (all Sigma Aldrich, Steinheim, Germany) ranging

571 in concentration from 50 to 0.0625 mM, were included to the binding reaction buffer.

572 Samples were loaded onto a 6 % non-denaturing PAA gel prepared in TAE buffer (40 mM

573 Tris acetate [pH 8.0], 2 mM EDTA) and run in a 0.5-to-2.0 x gradient of TAE at 100 V for 90

574 min in an Atto Mini PAGE system (Atto Bioscience and Biotechnology, Tokyo, Japan).

575 Signals were detected using Odyssey Infrared Imaging System (Li-Cor Biosciences UK Ltd,

576 Cambridge, UK) and captured using the supplied software Odyssey V3.0.

577 **Primer extension analysis.** Total RNA was isolated from *B. breve* UCC2003, grown in

578 mMRS supplemented with 1 % LNnT or 1 % LNB, to early exponential phase, using a

579 previously described Macaloid method (78). RNA samples were treated with RNase-free

580 DNase (Ambion). Primer extension was performed by annealing 1 pmol of IRD700 synthetic

581 18-mer oligonucleotides to 15 μg of RNA as described previously (79). Sequence ladders of

582 the presumed promoter regions immediately upstream of *lntP1*, *lntS*, *nagB3*, *nagK*, *lnpB*,

583 *gltA*, *nahS* or *nahA*, amplified from both UCC2003 genomic DNA which were run alongside

584 the primer extension products, were produced using the same primer as the primer extension

585 reaction and employing the Thermo Sequenase Primer Cycle Sequencing Kit; Amersham.

586 Separation was achieved on a 6.5 % Li-Cor Matrix KB Plus acrylamide gel. Signal detection

587 and image capture was performed by means of a Li-Cor sequencing instrument (Li-Cor

588 Biosciences).

589

590 **Operator Consensus Sequence Prediction, using MEME and WEBLOGO online**

591 **software tools.** Co-regulated promoter regions were assessed for the presence of operator

592 sequences by the use of the MEME (Multiple Em for Motif Elicitation) online tool

593 (http://meme-suite.org/tools/meme) (80), which were the visualized by the WebLogo online

594 tool (http://weblogo.berkeley.edu/logo.cgi) (81, 82). Sequences used for consensus sequence

595 prediction are given in Supplemental Table S2.

19

596

Microarray data accession number. The microarray data obtained in this study have been deposited in NCBI's Gene Expression Omnibus database and are accessible through GEO Series accession number GSE105108.

600

601

602

603

604

605

606

607 **References:**

608 1. Álvarez-Martín P, O'Connell-Motherway M, van Sinderen D, Mayo B. 2007. Functional
609 analysis of the pBC1 replicon from Bifidobacterium catenulatum L48. Applied Microbiology
610 and Biotechnology 76:1395-1402.
611 2. Turroni F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, Kerr C, Hourihane J,
612 Murray D, Fuligni F, Gueimonde M, Margolles A, De Bellis G, O'Toole PW, van Sinderen D,
613 Marchesi JR, Ventura M. 2012. Diversity of Bifidobacteria within the Infant Gut Microbiota.
614 PLoS ONE 7:e36957.
615 3. Sivan A, Corrales L, Hubert N, Williams JB, Aquino-Michaels K, Earley ZM, Benyamin FW, Man
616 Lei Y, Jabri B, Alegre M-L, Chang EB, Gajewski TF. 2015. Commensal Bifidobacterium
617 promotes antitumor immunity and facilitates anti–PD-L1 efficacy. Science 350:1084-1089.
618 4. Maslowski KM, Mackay CR. 2011. Diet, gut microbiota and immune responses. Nat Immunol
619 12:5-9.
620 5. Milani C, Duranti S, Bottacini F, Casey E, Turroni F, Mahony J, Belzer C, Delgado Palacio S,
621 Arboleya Montes S, Mancabelli L, Lugli GA, Rodriguez JM, Bode L, de Vos W, Gueimonde M,
622 Margolles A, van Sinderen D, Ventura M. 2017. The First Microbial Colonizers of the Human
623 Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota.
624 Microbiology and Molecular Biology Reviews 81.
625 6. Gibson GR, Roberfroid MB. 1995. Dietary modulation of the human colonic microbiota:
626 introducing the concept of prebiotics. J Nutr 125:1401-12.
627 7. Kunz C, Rudloff S, Baier W, Klein N, Strobel S. 2000. OLIGOSACCHARIDES IN HUMAN MILK:
628 Structural, Functional, and Metabolic Aspects. Annual Review of Nutrition 20:699-722.
629 8. Brüssow H. 2015. Human microbiota: 'The philosophers have only interpreted the world in
630 various ways. The point, however, is to change it'. Microbial Biotechnology 8:11-12.
631 9. Asakuma S, Hatakeyama E, Urashima T, Yoshida E, Katayama T, Yamamoto K, Kumagai H,
632 Ashida H, Hirose J, Kitaoka M. 2011. Physiology of consumption of human milk
633 oligosaccharides by infant gut-associated bifidobacteria. J Biol Chem 286:34583-92.
634 10. Sela DA, Mills DA. 2010. Nursing our microbiota: molecular linkages between bifidobacteria
635 and milk oligosaccharides. Trends Microbiol 18:298-307.

11. Kitaoka M. 2012. Bifidobacterial enzymes involved in the metabolism of human milk oligosaccharides. Adv Nutr 3:422s-9s.

12. Good M, Sodhi CP, Yamaguchi Y, Jia H, Lu P, Fulton WB, Martin LY, Prindle T, Nino DF, Zhou Q, Ma C, Ozolek JA, Buck RH, Goehring KC, Hackam DJ. 2016. The human milk oligosaccharide 2'-fucosyllactose attenuates the severity of experimental necrotising enterocolitis by enhancing mesenteric perfusion in the neonatal intestine. Br J Nutr 116:1175-1187.

13. Bode L. 2012. Human milk oligosaccharides: every baby needs a sugar mama. Glycobiology 22:1147-62.

14. Wada J, Ando T, Kiyohara M, Ashida H, Kitaoka M, Yamaguchi M, Kumagai H, Katayama T, Yamamoto K. 2008. Bifidobacterium bifidum lacto-N-biosidase, a critical enzyme for the degradation of human milk oligosaccharides with a type 1 structure. Appl Environ Microbiol 74:3996-4004.

15. Moller PL, Jorgensen F, Hansen OC, Madsen SM, Stougaard P. 2001. Intra- and extracellular beta-galactosidases from Bifidobacterium bifidum and B. infantis: molecular cloning, heterologous expression, and comparative characterization. Appl Environ Microbiol 67:2276-83.

16. Miwa M, Horimoto T, Kiyohara M, Katayama T, Kitaoka M, Ashida H, Yamamoto K. 2010. Cooperation of beta-galactosidase and beta-N-acetylhexosaminidase from bifidobacteria in assimilation of human milk oligosaccharides with type 2 structure. Glycobiology 20:1402-9.

17. Ashida H, Miyake A, Kiyohara M, Wada J, Yoshida E, Kumagai H, Katayama T, Yamamoto K. 2009. Two distinct alpha-L-fucosidases from Bifidobacterium bifidum are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates. Glycobiology 19:1010-7.

18. Kiyohara M, Tanigawa K, Chaiwangsri T, Katayama T, Ashida H, Yamamoto K. 2011. An exo-alpha-sialidase from bifidobacteria involved in the degradation of sialyloligosaccharides in human milk and intestinal glycoconjugates. Glycobiology 21:437-47.

19. Wada J, Suzuki R, Fushinobu S, Kitaoka M, Wakagi T, Shoun H, Ashida H, Kumagai H, Katayama T, Yamamoto K. 2007. Purification, crystallization and preliminary X-ray analysis of the galacto-N-biose-/lacto-N-biose I-binding protein (GL-BP) of the ABC transporter from Bifidobacterium longum JCM1217. Acta Crystallogr Sect F Struct Biol Cryst Commun 63:751-3.

20. Sakurama H, Kiyohara M, Wada J, Honda Y, Yamaguchi M, Fukiya S, Yokota A, Ashida H, Kumagai H, Kitaoka M, Yamamoto K, Katayama T. 2013. Lacto-N-biosidase Encoded by a Novel Gene of Bifidobacterium longum Subspecies longum Shows Unique Substrate Specificity and Requires a Designated Chaperone for Its Active Expression. The Journal of Biological Chemistry 288:25194-25206.

21. Sela DA, Chapman J, Adeuya A, Kim JH, Chen F, Whitehead TR, Lapidus A, Rokhsar DS, Lebrilla CB, German JB, Price NP, Richardson PM, Mills DA. 2008. The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. Proc Natl Acad Sci U S A 105:18964-9.

22. Yoshida E, Sakurama H, Kiyohara M, Nakajima M, Kitaoka M, Ashida H, Hirose J, Katayama T, Yamamoto K, Kumagai H. 2012. Bifidobacterium longum subsp. infantis uses two different beta-galactosidases for selectively degrading type-1 and type-2 human milk oligosaccharides. Glycobiology 22:361-8.

23. Garrido D, Ruiz-Moyano S, Mills DA. 2012. Release and utilization of N-acetyl-D-glucosamine from human milk oligosaccharides by Bifidobacterium longum subsp. infantis. Anaerobe 18:430-5.

24. Xiao JZ, Takahashi S, Nishimoto M, Odamaki T, Yaeshima T, Iwatsuki K, Kitaoka M. 2010. Distribution of in vitro fermentation ability of lacto-N-biose I, a major building block of human milk oligosaccharides, in bifidobacterial strains. Appl Environ Microbiol 76:54-9.

686   25.   Sela DA, Li Y, Lerno L, Wu S, Marcobal AM, German JB, Chen X, Lebrilla CB, Mills DA. 2011. An
687   Infant-associated Bacterial Commensal Utilizes Breast Milk Sialyloligosaccharides. The
688   Journal of Biological Chemistry 286:11909-11918.
689   26.   Sela DA, Garrido D, Lerno L, Wu S, Tan K, Eom H-J, Joachimiak A, Lebrilla CB, Mills DA. 2012.
690   Bifidobacterium longum subsp. infantis ATCC 15697 α-Fucosidases Are Active on Fucosylated
691   Human Milk Oligosaccharides. Applied and Environmental Microbiology 78:795-803.
692   27.   Bunesova V, Lacroix C, Schwab C. 2016. Fucosyllactose and L-fucose utilization of infant
693   Bifidobacterium longum and Bifidobacterium kashiwanohense. BMC Microbiology 16:248.
694   28.   James K, Motherway MO, Bottacini F, van Sinderen D. 2016. Bifidobacterium breve UCC2003
695   metabolises the human milk oligosaccharides lacto-N-tetraose and lacto-N-neo-tetraose
696   through overlapping, yet distinct pathways. Sci Rep 6:38560.
697   29.   Egan M, Motherway MO, Kilcoyne M, Kane M, Joshi L, Ventura M, van Sinderen D. 2014.
698   Cross-feeding by Bifidobacterium breve UCC2003 during co-cultivation with Bifidobacterium
699   bifidum PRL2010 in a mucin-based medium. BMC Microbiol 14:282.
700   30.   Egan M, O'Connell Motherway M, Ventura M, van Sinderen D. 2014. Metabolism of sialic
701   acid by Bifidobacterium breve UCC2003. Appl Environ Microbiol 80:4414-26.
702   31.   Egan M, Jiang H, O'Connell Motherway M, Oscarson S, van Sinderen D. 2016. Glycosulfatase-
703   Encoding Gene Cluster in Bifidobacterium breve UCC2003. Appl Environ Microbiol 82:6611-
704   6623.
705   32.   Titgemeyer F, Reizer J, Reizer A, Saier MH, Jr. 1994. Evolutionary relationships between
706   sugar kinases and transcriptional repressors in bacteria. Microbiology 140 (Pt 9):2349-54.
707   33.   O'Connell Motherway M, Fitzgerald GF, Neirynck S, Ryan S, Steidler L, van Sinderen D. 2008.
708   Characterization of ApuB, an extracellular type II amylopullulanase from Bifidobacterium
709   breve UCC2003. Appl Environ Microbiol 74:6271-9.
710   34.   Ruiz L, Motherway MOC, Lanigan N, van Sinderen D. 2013. Transposon Mutagenesis in
711   Bifidobacterium breve: Construction and Characterization of a Tn5 Transposon Mutant
712   Library for Bifidobacterium breve UCC2003. PLoS ONE 8:e64699.
713   35.   Ruiz L, Zomer A, O'Connell-Motherway M, van Sinderen D, Margolles A. 2012. Discovering
714   novel bile protection systems in Bifidobacterium breve UCC2003 through functional
715   genomics. Appl Environ Microbiol 78:1123-31.
716   36.   Hawley DK, McClure WR. 1983. Compilation and analysis of Escherichia coli promoter DNA
717   sequences. Nucleic Acids Research 11:2237-2255.
718   37.   Bottacini F, Zomer A, Milani C, Ferrario C, Lugli GA, Egan M, Ventura M, van Sinderen D.
719   2017. Global transcriptional landscape and promoter mapping of the gut commensal
720   Bifidobacterium breve UCC2003. BMC Genomics 18:991.
721   38.   Pokusaeva K, O'Connell-Motherway M, Zomer A, Macsharry J, Fitzgerald GF, van Sinderen D.
722   2011. Cellodextrin utilization by bifidobacterium breve UCC2003. Appl Environ Microbiol
723   77:1681-90.
724   39.   O'Connell KJ, Motherway MO, Liedtke A, Fitzgerald GF, Paul Ross R, Stanton C, Zomer A, van
725   Sinderen D. 2014. Transcription of two adjacent carbohydrate utilization gene clusters in
726   Bifidobacterium breve UCC2003 is controlled by LacI- and repressor open reading frame
727   kinase (ROK)-type regulators. Appl Environ Microbiol 80:3604-14.
728   40.   Egan M, O'Connell Motherway M, van Sinderen D. 2015. A GntR-type transcriptional
729   repressor controls sialic acid utilization in Bifidobacterium breve UCC2003. FEMS Microbiol
730   Lett 362.
731   41.   Ravcheev DA, Khoroshkin MS, Laikova ON, Tsoy OV, Sernova NV, Petrova SA, Rakhmaninova
732   AB, Novichkov PS, Gelfand MS, Rodionov DA. 2014. Comparative genomics and evolution of
733   regulons of the LacI-family transcription factors. Frontiers in Microbiology 5.
734   42.   Gilbert W, Maxam A. 1973. The Nucleotide Sequence of the lac Operator. Proceedings of the
735   National Academy of Sciences of the United States of America 70:3581-3584.

736 43. O'Connell Motherway M, Fitzgerald GF, van Sinderen D. 2011. Metabolism of a plant derived
737   galactose-containing polysaccharide by Bifidobacterium breve UCC2003. Microbial
738   Biotechnology 4:403-416.
739 44. Pokusaeva K, Neves AR, Zomer A, O'Connell-Motherway M, MacSharry J, Curley P, Fitzgerald
740   GF, van Sinderen D. 2010. Ribose utilization by the human commensal Bifidobacterium breve
741   UCC2003. Microb Biotechnol 3:311-23.
742 45. Matthews KS, Nichols JC. 1998. Lactose repressor protein: functional properties and
743   structure. Prog Nucleic Acid Res Mol Biol 58:127-64.
744 46. Lewis M. 2005. The lac repressor. Comptes Rendus Biologies 328:521-548.
745 47. Plumbridge J. 2001. DNA binding sites for the Mlc and NagC proteins: regulation of nagE,
746   encoding the N-acetylglucosamine-specific transporter in Escherichia coli. Nucleic Acids
747   Research 29:506-514.
748 48. Brechemier-Baey D, Dominguez-Ramirez L, Oberto J, Plumbridge J. 2015. Operator
749   recognition by the ROK transcription factor family members, NagC and Mlc. Nucleic Acids
750   Res 43:361-72.
751 49. Khoroshkin MS, Leyn SA, Van Sinderen D, Rodionov DA. 2016. Transcriptional Regulation of
752   Carbohydrate Utilization Pathways in the Bifidobacterium Genus. Frontiers in Microbiology
753   7:120.
754 50. Weickert MJ, Adhya S. 1992. A family of bacterial regulators homologous to Gal and Lac
755   repressors. J Biol Chem 267:15869-74.
756 51. Weickert MJ, Adhya S. 1993. The galactose regulon of Escherichia coli. Mol Microbiol 10:245-
757   51.
758 52. O'Connell Motherway M, Kinsella M, Fitzgerald GF, van Sinderen D. 2013. Transcriptional
759   and functional characterization of genetic elements involved in galacto-oligosaccharide
760   utilization by Bifidobacterium breve UCC2003. Microb Biotechnol 6:67-79.
761 53. Barboza M, Sela DA, Pirim C, LoCascio RG, Freeman SL, German JB, Mills DA, Lebrilla CB.
762   2009. Glycoprofiling Bifidobacterial Consumption of Galacto-Oligosaccharides by Mass
763   Spectrometry Reveals Strain-Specific, Preferential Consumption of Glycans. Applied and
764   Environmental Microbiology 75:7319-7325.
765 54. Coulier L, Timmermans J, Bas R, Van Den Dool R, Haaksman I, Klarenbeek B, Slaghek T, Van
766   Dongen W. 2009. In-depth characterization of prebiotic galacto-oligosaccharides by a
767   combination of analytical techniques. J Agric Food Chem 57:8488-95.
768 55. Pokusaeva K, Fitzgerald GF, van Sinderen D. 2011. Carbohydrate metabolism in
769   Bifidobacteria. Genes & Nutrition 6:285-306.
770 56. De Man JC RM, Sharpe ME. 1960. A medium for the cultivation of 688 lactobacilli. JAppl
771   Bacteriol 23:130-135.
772 57. Watson D, O'Connell Motherway M, Schoterman MH, van Neerven RJ, Nauta A, van
773   Sinderen D. 2013. Selective carbohydrate utilization by lactobacilli and bifidobacteria. J Appl
774   Microbiol 114:1132-46.
775 58. Terzaghi BE, Sandine WE. 1975. Improved medium for lactic streptococci and their
776   bacteriophages. Appl Microbiol 29:807-13.
777 59. Sambrook J FE, Maniatis T. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold
778   Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
779 60. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis:
780   sequence visualization and annotation. Bioinformatics 16:944-5.
781 61. O'Connell Motherway M, Zomer A, Leahy SC, Reunanen J, Bottacini F, Claesson MJ, O'Brien
782   F, Flynn K, Casey PG, Moreno Munoz JA, Kearney B, Houston AM, O'Mahony C, Higgins DG,
783   Shanahan F, Palva A, de Vos WM, Fitzgerald GF, Ventura M, O'Toole PW, van Sinderen D.
784   2011. Functional genome analysis of Bifidobacterium breve UCC2003 reveals type IVb tight
785   adherence (Tad) pili as an essential and conserved host-colonization factor. Proceedings of
786   the National Academy of Sciences 108:11217-11222.

787 62. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool.
788   J Mol Biol 215:403-10.
789 63. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped
790   BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids
791   Res 25:3389-402.
792 64. O'Riordan K, Fitzgerald GF. 1999. Molecular characterisation of a 5.75-kb cryptic plasmid
793   from Bifidobacterium breve NCFB 2258 and determination of mode of replication. FEMS
794   Microbiol Lett 174:285-94.
795 65. O'Connell Motherway M, O'Driscoll J, Fitzgerald GF, Van Sinderen D. 2009. Overcoming the
796   restriction barrier to plasmid transformation and targeted mutagenesis in Bifidobacterium
797   breve UCC2003. Microb Biotechnol 2:321-32.
798 66. Maze A, O'Connell-Motherway M, Fitzgerald GF, Deutscher J, van Sinderen D. 2007.
799   Identification and characterization of a fructose phosphotransferase system in
800   Bifidobacterium breve UCC2003. Appl Environ Microbiol 73:545-53.
801 67. Law J, Buist G, Haandrikman A, Kok J, Venema G, Leenhouts K. 1995. A system to generate
802   chromosomal mutations in Lactococcus lactis which allows fast analysis of targeted genes. J
803   Bacteriol 177:7011-8.
804 68. Egan M, O'Connell Motherway M, Ventura M, van Sinderen D. 2014. Metabolism of sialic
805   acid by Bifidobacterium breve UCC2003. Applied and Environmental Microbiology
806   doi:10.1128/aem.01114-14.
807 69. Zomer A, Fernandez M, Kearney B, Fitzgerald GF, Ventura M, van Sinderen D. 2009. An
808   interactive regulatory network controls stress response in Bifidobacterium breve UCC2003. J
809   Bacteriol 191:7039-49.
810 70. Garcia de la Nava J, Santaella DF, Cuenca Alba J, Maria Carazo J, Trelles O, Pascual-Montano
811   A. 2003. Engene: the processing and exploratory analysis of gene expression data.
812   Bioinformatics 19:657-8.
813 71. van Hijum S, de Jong A, Baerends R, Karsens H, Kramer N, Larsen R, den Hengst C, Albers C,
814   Kok J, Kuipers O. 2005. A generally applicable validation scheme for the assessment of
815   factors involved in reproducibility and quality of DNA-microarray data. BMC Genomics 6:1-
816   10.
817 72. van Hijum SA, Garcia de la Nava J, Trelles O, Kok J, Kuipers OP. 2003. MicroPreP: a cDNA
818   microarray data pre-processing framework. Appl Bioinformatics 2:241-4.
819 73. Long AD, Mangalam HJ, Chan BY, Tolleri L, Hatfield GW, Baldi P. 2001. Improved statistical
820   inference from DNA microarray data using analysis of variance and a Bayesian statistical
821   framework. Analysis of global gene expression in Escherichia coli K12. J Biol Chem
822   276:19937-44.
823 74. Mierau I, Kleerebezem M. 2005. 10 years of the nisin-controlled gene expression system
824   (NICE) in Lactococcus lactis. Appl Microbiol Biotechnol 68:705-17.
825 75. O'Connell KJ, O'Connell Motherway M, O'Callaghan J, Fitzgerald GF, Ross RP, Ventura M,
826   Stanton C, van Sinderen D. 2013. Metabolism of four alpha-glycosidic linkage-containing
827   oligosaccharides by Bifidobacterium breve UCC2003. Appl Environ Microbiol 79:6280-92.
828 76. de Ruyter PG, Kuipers OP, de Vos WM. 1996. Controlled gene expression systems for
829   Lactococcus lactis with the food-grade inducer nisin. Applied and Environmental
830   Microbiology 62:3662-3667.
831 77. Hamoen LW, Van Werkhoven AF, Bijlsma JJ, Dubnau D, Venema G. 1998. The competence
832   transcription factor of Bacillus subtilis recognizes short A/T-rich sequences arranged in a
833   unique, flexible pattern along the DNA helix. Genes Dev 12:1539-50.
834 78. Kuipers OP, Beerthuyzen MM, Siezen RJ, De Vos WM. 1993. Characterization of the nisin
835   gene cluster nisABTCIPR of Lactococcus lactis. Requirement of expression of the nisA and nisI
836   genes for development of immunity. Eur J Biochem 216:281-91.

24

837   79.   Ventura M, Zink R, Fitzgerald GF, van Sinderen D. 2005. Gene structure and transcriptional
838   organization of the dnaK operon of Bifidobacterium breve UCC 2003 and application of the
839   operon in bifidobacterial tracing. Appl Environ Microbiol 71:487-500.
840   80.   Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover
841   motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2:28-36.
842   81.   Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator.
843   Genome Res 14:1188-90.
844   82.   Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display consensus
845   sequences. Nucleic Acids Research 18:6097-6100.

846

847

848   **Figure 1.** Schematic representation of HMO metabolism-associated loci in *B. breve*
849   UCC2003, as identified preciously [22]. (a) The genes of the *lnt* locus. (b) The genes of the
850   *nah* locus. (c) The genes of the *nag* locus, and adjacent genes *nagR2* and *nagK2*. (d) The
851   genes of the *lnp*/*glt* locus. The length of the arrows is proportional to the size of the open
852   reading. Genes shown in red possess a predicted promoter in their upstream intergenic region.
853   Genes shown in green are predicted to encode a regulator protein. Genes shown in blue were
854   identified as not possessing a predicted promoter in their upstream intergenic region.

855

856   **Figure 2.** EMSA images showing (a) LntR, (b) NahR, and (c) and (d) NagR1 interactions
857   with a range of DNA fragments from the regions in the proximity of their predicted target
858   promoters, in order to identify their approximate locations. The locations and sizes of
859   fragments used, in relation to the promoter regions' respective transcription start sites, are
860   given in Supplemental Table S1. The panels below schematically represent the locations of
861   the DNA fragments used in relation to the locations of the putative operator sequences (red
862   boxes), transcription start sites (green arrow) and genes (blue box arrows). In each panel, '-'
863   indicates a negative control, where an equivalent amount crude cell extract from NZ9000
864   harbouring empty pNZ8150 was added instead of crude extract from the regulator-expressing
865   NZ9000 strain.

866

867   **Figure 3.** WebLogo representation of the operator motif consensus sequences for (a) the
868   LacI-type regulator LntR, (b) the NahC/XylR-typr regulator NahR and (c) the ROK/NagC-
869   type regulator NagR1, predicted using *in silico* analysis. Predicted operator sequences
870   identified in the intergenic regions containing the co-regulated promoters for each regulator
871   using the MEME online tool. Motif consensuses were generated by inputting these predicted
872   operator sequences to the WebLogo online tool. The locations and sequences of each operator
873   are shown alongside their respective consensus sequence.

874

875   **Figure 4.** EMSA images showing (a) LntR, (b) NahR, and (c) and (d) NagR1 interactions
876   with promoter-containing DNA fragments, with the addition of a gradient of their respective
877   inducers, ranging from 0mM-20mM. In each panel, 'C' indicates a negative control, where an
878   equivalent amount crude cell extract from NZ9000 harbouring empty pNZ8150 was added
879   instead of crude extract from the regulator-expressing NZ9000 strain.

26

880

Figure 5. Schematic representation of the proposed model for transcriptomic regulation of
LNT, LNnT and LNB metabolism by *B. breve* UCC2003. LNT and LNnT are internalised
and intracellularly degraded through sequential hydrolysis and release of monosaccharides
from their non-reducing ends. These released monosaccharides act as effectors relieving
transcriptional repression of the loci encoding the cellular components responsible for
liberation of these glycans. As such, liberated galactose relieves transcriptional repression of
the *lnt* locus, and N-acetylglucosamine relieves transcriptional repression of the *nah* locus.
Similarly, intracellular degradation of LNB (derived from the extracellular hydrolysis of
complex HMO structures by other infant GIT microbes) releases N-acetylglucosamine and
galactose-6-phosphate, the latter of which relieves transcriptional repression of the *lnp*/*glt* and
nag loci.

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

915 **Author Contributions Statement:** D.v.S., K.J. and M.O.C.M. conceived the experiments.

916 K.J., with the assistance of C.P. and R.O'B., conducted the experiments. All authors analysed

917 the results and contributed to writing the manuscript.

918 **Additional Information:** The microarray data obtained in this study have been deposited in

919 NCBI's Gene Expression Omnibus database and are accessible through GEO Series

920 accession number GSE105108.

921 The authors declare that, to the best of their knowledge, there are no competing financial

922 interests.

923

924

925

926

927

928

929

930

931

932

933

934

935

28

936

Strain or plasmid	Relevant Features	Reference or Source
Strains		
Escherichia coli strains		
E. coli EC101	Cloning host, repA$^+$ km$^+$	(43)
E. coli EC101-pNZ-M.BbrII + M.BbrIII	EC101 harbouring pNZ8048 derivative containing *bbrIIM* and *bbrIIIM*	(41)
Lactococcus lactis strains		
L. lactis NZ9000	MG1363, pepN::nisRK, nisin-inducible overexpression host	(55)
L. lactis NZ9700	Nisin-producing strain	(55)
L. lactis NZ9000-pNZ-lntR	NZ9000 conatining pNZ-lntR	This study
L. lactis NZ9000-pNZ-nahR	NZ9000 conatining pNZ-nahR	This study
L. lactis NZ9000-pNZ-nagR1	NZ9000 conatining pNZ-nagR1	This study
Bifidobacterium sp. Strains		
B. breve UCC2003	Isolate from nursling stool	(42)
B. breve UCC2003-lntR	pORI19-tet-bbr_0526 insertion mutant of UCC2003	This study
B. breve UCC2003-nahR	pORI19-tet-bbr_1555 insertion mutant of UCC2003	This study
B. breve UCC2003-nagR1	pORI19-tet-bbr_1249 insertion mutant of UCC2003	This study
B. breve UCC2003-nagR2	pORI19-tet-bbr_1251 insertion mutant of UCC2003	This study
Plasmids		
pAM5	pBC1-puC19-Tet^r	(1)
pORI19	Em^r, repA$^-$, ori$^+$, cloning vector	(43)
pORI19-tet-lntR	Internal 367 bp fragment of *bbr_0526* and tetW cloned in pORI19	This study
pORI19-tet-nahR	Internal 448 bp fragment of *bbr_1554* and tetW cloned in pORI19	This study
pORI19-tet-nagR1	Internal 502 bp fragment of *bbr_1249* and tetW cloned in pORI19	This study
pORI19-tet-nagR2	Internal 507 bp fragment of *bbr_1251* and tetW cloned in pORI19	This study
pNZ8150	Cm^r, nisin inducible translational fusion vector	(51)
pNZ-lntR	Cm^r, pNZ8150 derivative containing translational fusion of Bbr_0526 encoding DNA fragment to nisin inducible promoter	This study
pNZ-nahR	Cm^r, pNZ8150 derivative containing translational fusion of Bbr_1555 encoding DNA fragment to nisin inducible promoter	This study
pNZ-nagR1	Cm^r, pNZ8150 derivative containing translational fusion of Bbr_1249 encoding DNA fragment to nisin inducible promoter	This study

Table 1: Bacterial plasmids and strains used in this work.

Cm^r, Em^r,Km^r and Tet^r, resistance to chloramphenicol, erythromycin, kanamycin and tetracycline, respectively.

UCC, University College Cork Culture Collection.

Purpose	Primer	Sequence (5'-3')
Cloning of Bbr_0526 in pNZ8150	526F	tgcatc**cccggg**atg*catcaccatcaccatcaccatcaccatcac*gcgagaccaacacaggtttcc
	526R	tgcgca**tctaga**cgtttcccgtataccattaatcag
Cloning of Bbr_1555 in pNZ8150	1555F	tgcatc**gatatc**atg*catcaccatcaccatcaccatcaccatcac*tacgctaaatccaatccc
	1555R	tgcgca**tctaga**cggcggcacggtgatctg
Cloning of Bbr_1249 in pNZ8150	1249F	tgcatc**cagctg**atg*catcaccatcaccatcaccatcaccatcac*tcgtatcccggtcttgcc
	1249R	tgcatc**cagctg**atgtcgtatcccggtcttgcc
Cloning of internal 465bp fragment of Bbr_0526 in pORI19	IM526F	ctggtc**aagctt**cgttgaagccgcgatgga
	IM526R	ctggtc**tctaga**gtcaacggtggggcagtg
Cloning of internal 488bp fragment of Bbr_1555 in pORI19	IM1555F	ctggtc**aagctt**gctggccatcgatacggac
	IM1555R	ctggtc**tctaga**ctcgtcgttcagcagcac
Cloning of internal 443bp fragment of Bbr_1249 in pORI19	IM1249F	ctggtc**aagctt**cgaagaaggcctattgcg
	IM1249R	ctggtc**tctaga**cagcagaatcgccgaacc
Cloning of internal 488bp fragment of Bbr_1251 in pORI19	IM1251F	ctggtc**aagctt**gaagagaccggcgacctgg
	IM1251R	ctggtc**tctaga**gcccattgtcgatgacgcc
Amplification of *tet*W	tetWFw	tcagctgtcgacatgctcatgtacggtaaggaagca
	tetWRv	gcgacggtcgaccataacttctgattgttgccg
Confirmation of site specific homologous recombination	526confirm1	gcgctagctgttacaatggtc
	526confirm2	gccatttccaaccctctc
	1555confirm1	tacgctaaatccaatccc
	1555confirm2	gacgcaagggccaacaaccgc
	1249confirm1	catacagccgccacggcac
	1249confirm2	tcgtatcccggtcttgcc
	1251confirm1	gcagacgatactgcacgcg
	1251confrim2	gtcaagcatctctaccac
Amplification of Bbr_0527 promoter fragments with IRD700-labelled oligonucleotides	527IRDfa	ctcgcccctcgcttgtctctc
	527IRDra	gcataggcacggcagcgac
	527IRDfb	attgttttcgtgaccattg
	527IRDrb	gaataatgaacacgaacacg
	527IRDfc	caattttggtcaaccttcg
	527IRDrc	cgcgcgtagttctcgac
Amplification of Bbr_0530 promoter fragments with IRD700-labelled oligonucleotides	530IRDfa	gccgaacggtgtgctggtgg
	530IRDra	cttcatcgttctgttctccttc
	530IRDfb	cgataacacgcccgccatc
	530IRDrb	gctggacttgccgctatc
	530IRDfc	cttcatagagccacttc
	530IRDrc	ctcgaagtccttggcaac
Amplification of Bbr_1554 promoter fragments with IRD700-labelled oligonucleotides	1554IRDfa	gtcgctgggattggatttagcg
	1554IRDra	gtggctatgactgcgcgc
	1554IRDfb	cggcttcaggataacacccca
	1554IRDrb	ggatttggcggcgcgatc
	1554IRDfc	ccaaacaaaatagttgctacggc
	1554IRDrc	gttgagtgcggtgtaggtctcc
Amplification of Bbr_1556 promoter fragments with IRD700-labelled oligonucleotides	1556IRDfa	ctggacggctgctcaaagc
	1556IRDra	gcagagatgtttgaccgttcat
	1556IRDfb	ggtgacgacgccactctgc
	1556IRDrb	gtggtttgcggttgccct
	1556IRDfc	gccatctcaggaccgaacg
	1556IRDrc	ggtcgtcaaggtgatgaatcc
Amplification of Bbr_1248 promoter fragments with IRD700-labelled oligonucleotides	1248IRDfa	cctcctgcctgaacgatg
	1248IRDra	gacaatgatgatttccggc
	1248IRDfb	gttagggaacttcactaatacattcc
	1248IRDrb	cctggagagatgtcgatcgactc
	1248IRDfc	gtccgtacgtccataattgtaagtag
	1248IRDrc	cctgggagatgtcgatcgactc
	1248IRDfd	gatggcggctttgggcag
	1248IRDrd	cctgggagatgtcgatcgactc

Table 2: Oligonucleotide primers used in this work.

Restriction sites incorporated into oligonucleotide primer sequences are indicated in bold, and His-tag sequences incorporated into nucleotide primer sequences are indicated in italics.

Table 2: Oligonucleotide primers used in this work.

Purpose	Primer	Sequence (5'-3')
Amplification of Bbr_1250 promoter fragments with IRD700-labelled oligonucleotides	1250IRDfa 1250IRDra 1250IRDfb 1250IRDrb	gatgccgttgtggtagagatg ggtgttatcagtcattgcctatcc gcgtgtcgcgtatgaggc ggtgttatcagtcattgcctatcc
Amplification of Bbr_1586 promoter fragments with IRD700-labelled oligonucleotides	1586IRDfa 1586IRDra 1586IRDfb 1586IRDrb 1586IRDfc 1586IRDrc	cggttcgtcgaaaatccaag cagtgcgaagtgtgaggcg gccgcttattgcggctttatag cttttgagggcagaagtaactagttc gtatgcgcgttcgtccac cttttgagggcagaagtaactagttc
Amplification of Bbr_1590 promoter fragments with IRD700-labelled oligonucleotides	1590IRDfa 1590IRDra 1590IRDfb 1590IRDrb 1590IRDfc 1590IRDrc	ggcccgctggcagattag ggcaagagcagccacgatg gacagatgtctgagcggtc gaatcgggcagacggtgc cgcgcagaaattgttagttagg gaatcgggcagacggtgc
Amplification of region containing Bbr_0527 promoter region for sequencing ladders	527promF 527promR	gcattgctgtcattcgccacac gaataatgaacacgaacacg
Amplification of region containing Bbr_0530 promoter region for sequencing ladders	530promF 530promR	gcgtgcggatgaaactgg gtctggaacggcttggcgc
Amplification of region containing Bbr_1554 promoter region for sequencing ladders	1554promF 1554promR	cgtttcctcgaccccagttc gaatgtgtccttgagcttggc
Amplification of region containing Bbr_1556 promoter region for sequencing ladders	1556promF 1556promR	ctggacggctgctcaaagc ggtcgtcaaggtgatgaatcc
Amplification of region containing Bbr_1248 promoter region for sequencing ladders	1248promF 1248promR	ggaggctttggcggtacgg cctgggagatgtcgatcgactc
Amplification of region containing Bbr_1250 promoter region for sequencing ladders	1250promF 1250promR	gatgccgttgtggtagagatg ggtgccacccacatcaacac
Amplification of region containing Bbr_1586 promoter region for sequencing ladders	1586promF 1586promR	gcgagaccttcgaccttcagcc cggcacgagattgtaagacac
Amplification of region containing Bbr_1590 promoter region for sequencing ladders	1590promF 1590promR	ggcccgctggcagattag gaatcgggcagacggtgc
527 promoter for primer extension analysis	527PE	gcataggcacggcagcgac
530 promoter for primer extension analysis	530PE	cttcatcgttctgttctccttc
1554 promoter for primer extension analysis	1554PE	gttcatgttggtcttctttcc
1556 promoter for primer extension analysis	1556PE	gcagagatgtttgaccgttcat
1248 promoter for primer extension analysis	1248PE	ctgcccaaagccgcccatc
1250 promoter for primer extension analysis	1250PE	ggtgttatcagtcattgcctatcc
1586 promoter for primer extension analysis	1586PE	gcgatgtcaaatagtgtttcc
1590 promoter for primer extension analysis	1590PE	ggcaagagcagccacgatg

Restriction sites incorporated into oligonucleotide primer sequences are indicated in bold, and His-tag sequences incorporated into nucleotide primer sequences are indicated in italics.

1   **Table 3.** *B. breve* UCC2003 regulator mutant genes upregulated in transcription during
2   growth in mMRS medium supplemented with 1 % ribose as the sole carbohydrate, as
3   compared to the wild type (control).

Gene ID	Gene name	Function	Fold upregulation[a,b] during growth on:		
			UCC2003-lntR	UCC2003-nahR	UCC2003-nagR1
Bbr_0526	lntR	Transcriptional regulator, LacI family	N/A	-	-
Bbr_0527	lntP1	Permease protein of ABC transporter system for sugars	3.84	-	-
Bbr_0528	lntP2	Permease protein of ABC transporter system for sugars	3.77	-	-
Bbr_0529	lntA	GH42 Beta-galactosidase	2.78	-	-
Bbr_0530	lntS	Solute-binding protein of ABC transporter system for sugars	5.66	-	-
Bbr_1247	nagA2	CE9 nagA2 N-acetylglucosamine-6-phosphate deacetylase	-	-	6.70
Bbr_1248	nagB3	nagB3 Glucosamine-6-phosphate isomerase	-	-	9.11
Bbr_1249	nagR1	Transcriptional regulator, ROK family	-	-	N/A
Bbr_1250	nagK	Sugar kinase, ROK family	-	-	2.29
Bbr_1251	nagR2	Transcriptional regulator, ROK family	-	-	-
Bbr_1252	nagK2	Sugar kinase, pfkB family	-	-	-
Bbr_1554	nahS	Solute-binding protein of ABC transporter system (lactose)	-	17.44	-
Bbr_1555	nahR	NagC/XylR-type transcriptional regulator	-	N/A	-
Bbr_1556	nahA	GH20 nagZ Beta-N-acetylhexosaminidase	-	-	-
Bbr_1558	nahP	Permease protein of ABC transporter system	-	-	-
Bbr_1559	nahT1	ATP-binding protein of ABC transporter system	-	-	-
Bbr_1560	nahT2	ATP-binding protein of ABC transporter system	-	-	-
Bbr_1585	lnpD	UDP-glucose 4-epimerase	-	-	3.06
Bbr_1586	lnpB	Phosphotransferase family protein	-	-	3.36
Bbr_1587	lnpA	GH112 lacto-N-biose phorylase	-	-	2.90
Bbr_1588	gltC	Permease protein of ABC transporter system for sugars	-	-	2.91
Bbr_1589	gltB	Permease protein of ABC transporter system for sugars	-	-	3.02
Bbr_1590	gltA	Solute-binding protein of ABC transporter system for sugars	-	-	5.07

4   The level of transcription is shown as a fold-value of increase in transcription on each carbohydrate, as compared to a ribose control, with a
5   cut-off of a minimum 2.0-fold increase in transcription.

6   a Based on comparative transcriptome analysis using B. breve UCC2003-lntR, B. breve UCC2003-nahR and B. breve UCC2003-nagR1
7   grown on 1% ribose, as compared to wild-type B. breve UCC2003 grown under the same conditions as a control. Two independent
8   biological replicates were used for each array using a Cy3/Cy5 dye-swap.

9   b The cutoff point is 2.0-fold, with a P value of _0.001. —, value below the cutoff.

10 N/A indicates that the fold-value for increase in transcription for this gene is not included, as this is the gene in which the mutation was
11 made, and thus does not accurately represent its natural transcription under these conditions.

12 The level of transcription is not given for the regulator-encoding genes containing the mutations in their respective arrays, as their
13 transcription has been interrupted, and thus cannot be considered as reliable.

1

A

IntR ← | IntP1 → | IntP2 → | IntA → | IntS →

B

← nahS | nahR → | nahA → | nahP → | nahT1 → | nahT2 →

C

← nagA2 | ← nagB3 | nagR1 → | ← nagK | ┊ | nagR2 → | nagK2 →

D

← InpD | ← InpB | ← InpA | ← gltC | ← gltB | ← gltA

(a)

bits 2 1 0

5′ TG T T A CG A CA 3′
1 2 3 4 5 6 7 8 9 10 11 12 13 14

weblogo.berkeley.edu

-80   -60
ATAACATCATTGTTTCGTGACCATTGTAACAGCTAGCGCA
        IntP1ir1

-20   -10   * +1
GACAATGTTATGAATGGTGTTATGATGATAACGTAAACAA
-35   -10   IntP1ir2

IntP1
Upstream region

-80   -60
TTCAAATGCGATGGCGGGCGTGTTAGCGATAACACGCCCGCC
              IntSir1

-10   * +1   +10
TGTTTTGAGATGTGTATACTAAAAGATGTTATCGCAACCATAG
              -10   IntSir2

IntS
upstream region

(b)

bits 2 1 0

5′ CAT A ATG 3′
1 2 3 4 5 6 7 8 9 10

weblogo.berkeley.edu

-40 -35   -10
CGGTGTTTGACAAGTTTCCCATCCCTGTGCATAATCATG
        -35   -10
                              nahSir1

nahS
upstream region

-140   -120
GCCTGATAGTCATCACGATGCCGTCGGTGATCTGGGACTTCA
        nahAir1

nahA
upstream region

(c)

bits 2 1 0

5′ A T TGT A A T T T ACAA T 3′
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

weblogo.berkeley.edu

-10   * +1   +10
CCATAATGTAAGTAGTATTAACAAATAACTGGTTAGGTTGAT
-10   nagB3ir1

nagB3
upstream region

-40 -35
ATGGTTTTCTGTAATTGTTAAGATAGTTGTCAATACATAA
              -35
                  nagKir1

nagK
upstream region

-40 -35   -10
TCATATCAATGTCTATTTATGTATATATACTTAACAAAT
        -35   -10
                              lnpBir1

lnpB
upstream region

-40 -35
GCGCAGAAATTGTTAGTTAGGTTGACAATAAAGATTTGAGCGC
                      -35
                  gltAir1

gltA
upstream region

(a)

*IntP1*a   *IntS*a

Galactose − 20 15 10 5 2 1 0 mM

LntR − + + + + + + +   − + + + + + + + +

(b)

*nahS*a   *nahA*a

GlcNAc

NahR − + + + + + + + +   − + + + + + + + +

(c)

*nagB3*a   *nagK*a

Gal-6P

NagR1 − + + + + +   − + + + + +

(d)

*lnpB*a   *gltA*a

Gal-6P

NagR1 − + + + + +   − + + + + +