Identification and characterization of a glycosulfatase-encoding gene cluster in *Bifidobacterium breve* UCC2003

Muireann Egan[1], Hao Jiang[2], Mary O' Connell Motherway[1], Stefan Oscarson[2], Douwe van Sinderen[1]

[1]School of Microbiology and APC Microbiome Institute, University College Cork, Western Road, Cork, Ireland

[2]Centre for Synthesis and Chemical Biology, University College Dublin, Belfield, Dublin 4, Ireland.

Running title: Bifidobacterial metabolism of a sulfated monosaccharide

Keywords: Bifidobacteria, mucin, *N*-acetylglucosamine-6-sulfate, sulfatase, anaerobic sulfatase maturing enzyme

Corresponding author: Douwe van Sinderen, Room 4.05, School of Microbiology and APC Microbiome Institute, University College Cork, Western Road, Cork, Ireland. Tel: 00 353 21 490 1365, Fax: 00 353 21 490 3101. E-mail: d.vansinderen@ucc.ie.

Abstract

Bifidobacteria constitute a specific group of commensal bacteria, typically found in the gastrointestinal tract (GIT) of humans and other mammals. *Bifidobacterium breve* strains are numerically prevalent among the gut microbiota of many healthy breast-fed infants. In the current study, we investigated glycosulfatase activity in a bacterial nursling stool isolate, *B. breve* UCC2003. Two putative sulfatases were identified on the genome of *B. breve* UCC2003. The sulfated monosaccharide *N*-acetylglucosamine-6-sulfate (GlcNAc-6-S) was shown to support growth of *B. breve* UCC2003, while, *N*-acetylglucosamine-3-sulfate, *N*-acetylgalactosamine-3-sulfate and *N*-acetylgalactosamine-6-sulfate, did not support appreciable growth. Using a combination of transcriptomic and functional genomic approaches, a gene cluster, designated *ats2*, was shown to be specifically required for GlcNAc-6-S metabolism. Transcription of the *ats2* cluster is regulated by a ROK-family transcriptional repressor. This study represents the first description of glycosulfatase activity within the *Bifidobacterium* genus.

Importance

Bifidobacteria are saccharolytic organisms naturally found in the digestive tract of mammals and insects. *Bifidobacterium breve* strains utilize a variety of plant and host-derived carbohydrates which allow them to be present as prominent members of the infant gut microbiota as well as being present in the gastrointestinal tract of adults. In this study, we introduce a previously unexplored area of carbohydrate metabolism in bifidobacteria, namely the metabolism of sulfated carbohydrates. *B. breve* UCC2003 was shown to metabolize *N*-acetylglucosamine-6-sulfate (GlcNAc-6-S) through one of two sulfatase-encoding gene clusters identified on its genome. GlcNAc-6-S can be found in terminal or branched positions

of mucin oligosaccharides, the glycoprotein component of the mucous layer that covers the digestive tract. The results of this study provide further evidence of this species' ability to utilize mucin-derived sugars, a trait which may provide a competitive advantage in both the infant and adult gut.

Introduction

The *Bifidobacterium* genus represents one of the major components of the intestinal microbiota of breast-fed infants (1-5), while also typically constituting between 2 % and 10 % of the adult intestinal microbiota (6-11). Bifidobacteria are saccharolytic microorganisms whose ability to colonize and survive in the large intestine is presumed to depend on the ability to metabolize complex carbohydrates present in this environment (12, 13). Certain bifidobacterial species including *Bifidobacterium longum* subsp. *longum, Bifidobacterium adolescentis* and *Bifidobacterium breve* utilize a range of plant/diet-derived oligosaccharides such as raffinose, arabinoxylan, galactan and cellodextrins (14-20). Bifidobacterial metabolism of human milk oligosaccharides (HMOs) is also well-described, with the typically infant-derived species *B. longum* subsp. *infantis* and *Bifidobacterium bifidum* particularly well-adapted to utilize these carbon sources in the infant gut (21-23). However, the ability to utilize mucin, the glycoprotein component of the mucous layer that covers the epithelial cells of the gastrointestinal tract, is limited to members of the *B. bifidum* species (21, 24). Approximately 60 % of the predicted glycosyl hydrolases encoded by *B. bifidum* PRL2010 are predicted to be involved in mucin degradation, most of which are conserved exclusively within the *B. bifidum* species (21).

Host-derived glycoproteins such as mucin and proteoglycans (e.g. chondroitin sulfate and heparan sulfate), which are found in the colonic mucosa and/or human milk, are often highly sulfated (25-29). Human colonic mucin is heavily sulfated, which is in contrast to mucin from the stomach or small intestine, the presumed purpose of which is to protect mucin against degradation by bacterial glycosidases (30-32). Despite this apparent protective measure, glycosulfatase activity has been identified in various members of the gut microbiota, e.g. *Bacteroides thetaiotaomicron, Bacteroides ovatus* and *Prevotella* strain RS2 (33-38).

Prokaryotic and eukaryotic sulfatases uniquely require a 3-oxoalanine (typically called Cα-formylglycine or FGly) residue at their active site (39-41). Prokaryotic sulfatases carry either a conserved cysteine (Cys) or a serine (Ser) residue, which requires post-translational conversion to FGly in the cytosol in order to convert the enzyme to an active state (42-44). In bacteria, two distinct systems have been described for the post-translational modification of sulfatase enzymes. In *Mycobacterium tuberculosis*, the conversion of the Cys_{58} residue to FGly is catalyzed by an FGly-generating enzyme (FGE) which requires oxygen as a co-factor (45). In *Klebsiella pneumoniae,* the conversion of the Ser_{72} residue of the *atsA*-encoded sulfatase is catalysed by the AtsB enzyme, which is a member of the *S*-adenosyl-L-methionine (AdoMet)-dependent family of radical enzymes (43, 46). Similar enzymes have also been characterized from *Clostridium perfringens* and *Ba. thetaiotaomicron* which are active on both Cys and Ser-type sulfatases (37, 38, 47). Crucially, these enzymes are active under anaerobic conditions and were thus designated anaerobic sulfatase maturing enzymes (anSME) (38). Sulfatase activity has yet to be described in bifidobacteria. In the current study, we identify two predicted sulfatase and anSME-encoding gene clusters in *B. breve* UCC2003 (and other *B. breve* strains), and demonstrate that one such cluster is required for the metabolism of the sulfated monosaccharide *N*-acetylglucosamine-6-sulfate (GlcNAc-6-S).

Materials and methods

Bacterial strains, plasmids, media and culture conditions. Bacterial strains and plasmids used in this study are listed in Table 1. *B. breve* UCC2003 was routinely cultured in Reinforced Clostridial Medium (RCM; Oxoid Ltd., Basingstoke, Hampshire, United Kingdom). Carbohydrate utilization by bifidobacteria was examined in modified deMan Rogosa Sharpe (mMRS) medium made from first principles (48), excluding a carbohydrate source, supplemented with 0.05 % (wt/vol) L-cysteine HCl (Sigma Aldrich, Steinheim, Germany) and a particular carbohydrate source (0.5 % wt/vol). The carbohydrates used were lactose (Sigma Aldrich), GlcNAc-6-S (Dextra Laboratories, Reading, United Kingdom; see below), *N*-acetylglucosamine-3-sulfate (GlcNAc-3-S), *N*-acetylgalactosamine-3-sulfate (GalNAc-3-S) and *N*-acetylgalactosamine-6-sulfate (GalNAc-6-S) (see below). In order to determine bacterial growth profiles and final optical densities, 10 ml of a freshly prepared mMRS medium, supplemented with a particular carbohydrate, was inoculated with 100 µl (1 %) of a stationary-phase culture of a particular strain. Un-inoculated mMRS was used as a negative control. Cultures were incubated anaerobically for 24 h and the optical density (OD_{600nm}) was recorded. Bifidobacterial cultures were incubated under anaerobic conditions in a modular atmosphere-controlled system (Davidson and Hardy, Belfast, Ireland) at 37°C. *Escherichia coli* was cultured in Luria Bertani broth (LB) at 37°C with agitation (49). *Lactococcus lactis* strains were grown in M17 medium supplemented with 0.5 % (wt/vol) glucose at 30°C (50). Where appropriate, growth media contained tetracycline (Tet; 10 µg ml^{-1}), chloramphenicol (Cm; 5 µg ml^{-1} for *E. coli* and *L. lactis*, 2.5 µg ml^{-1} for *B. breve*), erythromycin (Em; 100 µg ml^{-1}) or kanamycin (Kan; 50 µg ml^{-1}). Recombinant *E. coli* cells containing pORI19 were selected on LB agar containing Em and Kan, and supplemented with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) (40 µg ml^{-1}) and 1 mM IPTG (isopropyl-β-D-galactopyranoside).

115

116 **Chemical synthesis of sulfated monosaccharides.** In brief, the 6-*O*-sulfated GlcNAc
117 structure (Fig. 1A, structure **1**) was synthesized in four steps from GlcNAc in an overall 40 %
118 yield while the other three target structures, 3-*O*-sulfated GlcNAc (Fig. 1A, **2**), 3-*O*-sulfated
119 GalNAc and 6-*O*-sulfated GalNAc (Fig. 1B, **3** and **4**, respectively), were synthesized from
120 their corresponding benzyl β-glycoside, (Fig. 1A, **8** and Fig. 1B, **12**), in three or four steps
121 with an overall yield of about 60 %. The benzyl glycoside was obtained either by direct
122 alkylation of a hemiacetal (Fig. 1A, **8**, GlcNAc) or by glycosylation of a peracetylated
123 precursor (Fig. 1B, **12**, GalNAc). Sulfations were performed using a $SO_3{\cdot}NEt_3$ complex in
124 pyridine or DMF (yields 86-96 %). Direct regioselective 6-*O*-tritylation of GlcNAc followed
125 by *in situ* acetylation afforded compound **5** from which the trityl group was removed using
126 aqueous acetic acid, without any acetyl migration detected, to yield the 6-OH derivative **6**,
127 sulfation of which gave compound **7** which was subsequently deacetylated using Zemplen
128 conditions to afford target structure **1** (Fig. 1A). Benzylidenation of compounds **8** and **12**
129 gave 3-OH compounds **9** and **13**, respectively. Sulfation (→**10** and **14**) followed by
130 deprotection through catalytic hydrogenolysis yielded target structures **2** and **3**.
131 Isopropylidenation of compound **12** gave the 6-OH compound **15**, which was sulfated (→**16**)
132 and then deprotected through acetal hydrolysis (→**17**) followed by catalytic hydrogenolysis
133 to afford target structure **4** (Fig. 1). The experimental methods are described in further detail
134 in the supplementary material.

135

136 **Nucleotide sequence analysis.** Sequence data were obtained from the Artemis-mediated
137 genome annotations of *B. breve* UCC2003 (51, 52). Database searches were performed using
138 the non-redundant sequence database accessible at the National Centre for Biotechnology

139 Information website (http://www.ncbi.nlm.nih.gov) using BLAST (53). Sequence analysis
140 was performed using the Seqbuilder and Seqman programs of the DNASTAR software
141 package (DNASTAR, Madison, WI, USA). Inverted repeats were identified using the
142 PrimerSelect program of the DNASTAR software package and a graphical representation of
143 the identified motifs was obtained using WebLogo software (54).

144

145 **DNA manipulations.** Chromosomal DNA was isolated from *B. breve* UCC2003 as
146 previously described (55). Plasmid DNA was isolated from *E. coli*, *L. lactis* and *B. breve*
147 using the Roche High Pure plasmid isolation kit (Roche Diagnostics, Basel, Switzerland). An
148 initial lysis step was performed using 30 mg ml^{-1} of lysozyme for 30 min at 37°C prior to
149 plasmid isolation from *L. lactis* or *B. breve* (56). Single stranded oligonucleotide primers
150 used in this study were synthesized by Eurofins (Ebersberg, Germany) (Table 2). Standard
151 PCRs were performed using Taq PCR master mix (Qiagen GmBH, Hilden, Germany). *B.*
152 *breve* colony PCRs were carried out as described previously (57). PCR fragments were
153 purified using the Roche High Pure PCR purification kit (Roche Diagnostics).
154 Electroporation of plasmid DNA into *E. coli*, *L. lactis* or *B. breve* was performed as
155 previously described (49, 58, 59).

156

157 **Construction of *B. breve* UCC2003 insertion mutants.** Internal fragments of Bbr_0849,
158 designated here as *atsR2* (fragment encompasses 408 bp, representing codon numbers 134
159 through to 271 of the 395 codons of this gene), Bbr_0851, designated *atsT* (fragment
160 encompasses 416 bp, representing codon numbers 149 through to 288 of the 476 codons of
161 this gene) and Bbr_0852, designated *atsA2* (fragment encompasses 402 bp, representing
162 codon numbers 148 through to 281 of the 509 codons of this gene) were amplified by PCR

163 using *B. breve* UCC2003 chromosomal DNA as a template and primer pairs atsR2F and
164 atsR2R, atsTF and atsTR, and atsA2F and atsA2R, respectively (Table 2). The insertion
165 mutants were constructed as described previously (57). Site-specific recombination of
166 potential Tet-resistant mutants was confirmed by colony PCR using primer combinations
167 TetWF and TetWR to verify *tetW* gene integration, and the primers atsR2confirm,
168 atsTconfirm and atsA2confirm (positioned upstream of the selected internal fragments of
169 *atsR2, atsT* and *atsA2,* respectively) in combination with primer TetWF to confirm
170 integration at the correct chromosomal location.

171

172 **Analysis of global gene expression using *B. breve* DNA microarrays.** Global gene
173 expression was determined during log-phase growth (OD_{600nm} of ~0.5) of *B. breve* UCC2003
174 in mMRS supplemented with 0.5 % GlcNAc-6-S and the obtained transcriptome was
175 compared to that obtained from *B. breve* UCC2003 grown in mMRS supplemented with 0.5
176 % ribose. Similarly, global gene expression of the insertion mutant *B. breve* UCC2003-atsR2
177 was determined during log-phase (OD_{600nm} of ~0.5) growth of the mutant in mMRS
178 supplemented with 0.5 % ribose and the transcriptome was also compared to that from *B.*
179 *breve* UCC2003 grown in 0.5 % ribose. DNA microarrays containing oligonucleotide primers
180 representing each of the 1864 identified open reading frames on the genome of *B. breve*
181 UCC2003 were designed and obtained from Agilent Technologies (Palo Alto, Ca., USA).
182 RNA was isolated and purified from bifidobacterial cells using a combination of the
183 "Macaloid" method and the Roche High Pure RNA isolation kit, as previously described
184 (60). RNA was quantified spectrophotometrically as described by Sambrook *et al*. (49).
185 Methods for complementary DNA synthesis and labelling were performed as described
186 previously (61). Hybridization, washing of the slides and processing of the DNA-microarray
187 data was also performed as previously described (62).

188

189 **Plasmid Constructions.** For the construction of plasmid pNZ-atsR2, a DNA fragment
190 encompassing the complete coding region of the predicted transcriptional regulator *atsR2*
191 (Bbr_0849) was generated by PCR amplification from chromosomal DNA of *B. breve*
192 UCC2003 using PfuUltra II DNA polymerase (Agilent Technologies) and the primer
193 combination atsR2FOR and atsR2REV (Table 2). The generated amplicon was digested with
194 NcoI and XbaI, and ligated into the similarly digested, nisin-inducible translational fusion
195 plasmid pNZ8048 (63). The ligation mixture was introduced into *L. lactis* NZ9000 by
196 electrotransformation and transformants were selected based on Cm resistance. The plasmid
197 content of a number of Cm^r transformants was screened by restriction analysis and the
198 integrity of positively identified clones was verified by sequencing.

199 To clone the Bbr_0849 promoter region, a DNA fragment encompassing the intergenic
200 region between the Bbr_0849 and Bbr_0850 genes was generated by PCR amplification
201 employing *B. breve* UCC2003 chromosomal DNA as a template, and using PfuUltra II DNA
202 polymerase in combination with primer pair atsRPromF and atsRPromR (Table 2). The PCR
203 product was digested with HindIII and XbaI, and ligated to the similarly digested pBC1.2
204 (64). The ligation mixture was introduced into *E. coli* XL1-blue by electrotransformation and
205 transformants were selected based on Tet and Cm resistance. Transformants were checked for
206 plasmid content by restriction analysis and the integrity of several positively identified
207 recombinant plasmids was verified by sequencing. One of these verified recombinant
208 plasmids, designated pBC1.2-atsProm, was introduced into *B. breve* UCC2003-atsR2 by
209 electrotransformation and transformants were selected based on Tet and Cm resistance.

210

211 **Heterologous protein production.** For the heterologous expression of AtsR2, 25 ml of M17
212 broth supplemented with 0.5 % (wt/vol) glucose was inoculated with a 2 % inoculum of an
213 overnight culture grown for 16 h of *L. lactis* NZ9000 harbouring either pNZ-atsR2 or the
214 empty vector pNZ8048 (used as a negative control), followed by incubation at 30°C until an
215 OD_{600nm} of ~0.5 was reached, at which point protein expression was induced by addition of
216 cell-free supernatant of a nisin-producing strain (65), followed by continued incubation for a
217 further 2 h. Cells were harvested by centrifugation, resuspended in 10 mM Tris-HCl (pH 8.0),
218 and disrupted with glass beads in a mini-bead beater (BioSpec Products, Bartlesville, OK).
219 Cellular debris was removed by centrifugation to produce an AtsR2-containing crude cell
220 extract.

221

222 **Electrophoretic mobility shift assays (EMSA).** DNA fragments representing different
223 portions of each of the promoter regions upstream of the *atsR2* and *atsT* genes were prepared
224 by PCR using IRD-labelled primer pairs synthesized by Integrated DNA Technologies
225 (Coralville, IA) (Table 2). EMSAs were essentially performed as described previously (66).
226 In all cases, the binding reactions were performed in a final reaction volume of 20 µl in the
227 presence of poly (dI-dC) in binding buffer (20 mM Tris-HCl, 5 mM $MgCl_2$, 0.5 mM
228 dithiothreitol [DTT], 1 mM EDTA, 50 mM KCl, 10 % glycerol at pH 7.0). Various amounts
229 of *L. lactis* NZ9000 crude cell extract containing pNZ-atsR2 or pNZ8048 were mixed on ice
230 with a fixed amount of DNA probe (0.1 pmol) and subsequently incubated for 30 min at
231 37°C. Samples were loaded on a 6 % non-denaturing polyacrylamide (PAA) gel prepared in
232 TAE buffer (40 mM Tris acetate (pH 8.0), 2 mM EDTA) and run in a 0.5 to 2.0 x gradient of
233 TAE at 100 V for 120 min in an Atto Mini PAGE system (Atto Bioscience and
234 Biotechnology, Tokyo, Japan). Signals were detected using an Odyssey Infrared Imaging
235 System (Li-Cor Biosciences, United Kingdom Ltd., Cambridge, United Kingdom) and

images were captured using the supplied Odyssey software v3.0. To identify the effector molecule of AtsR2, either GlcNAc or GlcNAc-6-S was added to the binding reaction in concentrations ranging from 2.5 mM to 20 mM.

Primer extension analysis. Total RNA was isolated from exponentially growing cells of *B. breve* UCC2003-atsR2 or *B. breve* UCC2003-atsR2-pBC1.2-atsRProm in mMRS supplemented with 0.5 % ribose, as previously described (61). Primer extension was performed by annealing 1 pmol of an IRD-labelled synthetic oligonucleotide to 20 µg of RNA as previously described (67), using primers AtsR2R1F or AtsR2T1R (Table 2). Sequence ladders of the presumed *atsR2* and *atsT* promoter regions were produced using the same primer as in the primer extension reaction and a DNA cycle-sequencing kit (Jena Bioscience, Germany) and were run alongside the primer extension products to allow precise alignment of the transcriptional start site with the corresponding DNA sequence. Separation was achieved on a 6.5 % Li-Cor Matrix KB Plus acrylamide gel. Signal detection and image capture were performed with a Li-Cor sequencing instrument (Li-Cor Biosciences).

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 GSE81240.

Results

Genetic organisation of the sulfatase gene clusters in *B. breve* UCC2003. Based on the presence of a sulfatase-associated PFAM domain PF00884 and the previously described N-terminally located sulfatase signature (CxPxR) (68, 69), two putative Cys-type sulfatase-encoding genes were identified on the genome of *B. breve* UCC2003. The first, represented by the gene with the associated locus tag Bbr_0352 (and designated here as *atsA1*), is located in a cluster of four genes, designated the *ats1* cluster, which also includes a gene encoding a predicted hypothetical membrane spanning protein (Bbr_0349), a gene (Bbr_0350, designated here as *atsB1*) specifying a putative anSME which contains the signature motif CxxxCxxC characteristic of the radical AdoMet-dependent superfamily (70), and a gene specifying a predicted LacI-type transcriptional regulator (Bbr_0351, designated *atsR1*). Adjacent to these four genes, but oppositely oriented, three genes are present that encode a predicted ABC-type transport system (corresponding to locus tags Bbr_0353 through to Bbr_0355) (Fig. 2).

The second predicted sulfatase-encoding gene, Bbr _0852 (designated here as *atsA2*), is located in a cluster of four genes (Bbr_0851 through to Bbr_0854, designated here as *ats2*). Bbr_0851, designated *atsT*, encodes a predicted transporter from the major facilitator superfamily. Bbr_0853 (designated *atsB2*) encodes a putative anSME, which contains the signature CxxxCxxC motif. Bbr_0854 encodes a predicted membrane spanning protein, which shares 75 % amino acid identity with the deduced protein encoded by Bbr_0349 of the *ats1* gene cluster (Fig. 2). The AtsA1 and AtsA2 proteins share 28 % amino acid identity, while the AtsB1 and AtsB2 proteins exhibit 74 % identity between each other. Interestingly, the *ats2* gene cluster has a notably different GC content (63.96 %) compared to the *B. breve* UCC2003 genome average (58.73 %), whereas the GC content of the *ats1* cluster (57.6 %) is comparable to that of the genome.

280 Based on the comparative genome analysis presented in Figure 2, we found that the putative
281 sulfatase clusters are well conserved among the *B. breve* strains whose genomes were
282 recently published (71). Of the currently available complete *B. breve* genomes, *B. breve*
283 NCFB2258, *B. breve* 689B, *B. breve* 12L and *B. breve* S27 encode clear homologues of both
284 identified putative sulfatase gene clusters described above. In contrast, the genomes of *B.*
285 *breve* JCM7017, *B. breve* JCM7019 and *B. breve* ACS-071-V-Sch8b contain just a single,
286 but variable putative sulfatase cluster (Fig. 2). A clear homologue of the *ats1* gene cluster
287 was also identified in the recently published genome of *B. longum* subsp. *infantis* BT1
288 (Accession number CP010411). No other homologues of either sulfatase-encoding gene
289 clusters were identified by BLASTP analysis within the available bifidobacterial genome
290 sequences.

291

292 **Growth of *B. breve* UCC2003 on sulfated monosaccharides.** The presence of two putative
293 sulfatase-encoding clusters on the genome of *B. breve* UCC2003 suggests that this gut
294 commensal is capable of removing a sulfate ester from a sulfated compound, possibly a
295 sulfated carbohydrate. In mMRS supplemented with 0.5 % GlcNAc-6-S as the sole carbon
296 source, the strain was capable of substantial growth (final OD_{600nm} values following overnight
297 growth varied between 0.6 and 0.8). However, no appreciable growth was observed on
298 GlcNAc-3-S, GalNAc-3-S or GalNAc-6-S. On the positive control, 0.5 % lactose, the strain
299 reached an OD_{600nm} of almost 2, which is comparable to previous studies with this strain (17,
300 72, 73) (Fig. 3A).

301

302 **Genome response of *B. breve* UCC2003 to growth on GlcNAc-6-S.** In order to investigate
303 which genes are responsible for GlcNAc-6-S metabolism in *B. breve* UCC2003, global gene

304 expression was determined by microarray analysis during growth of the strain in mMRS
305 supplemented with GlcNAc-6-S and compared with gene expression when grown in mMRS
306 supplemented with ribose. Ribose was considered an appropriate carbohydrate for
307 comparative transcriptome analysis because the genes involved in ribose metabolism are
308 known, while it has furthermore successfully been used in a number of transcriptome studies
309 in this strain (17, 18, 72-74). Of the two predicted sulfatase and anSME-encoding gene
310 clusters of *B. breve* UCC2003 (see above), transcription of the *ats2* gene cluster was
311 significantly up-regulated (fold change >3.0, *P*-value <0.001) during growth on GlcNAc-6-S,
312 while no (significant) difference in the level of transcription was observed for the *ats1* gene
313 cluster (Table 3). Interestingly, three other gene clusters were also significantly up-regulated
314 (corresponding to locus tags Bbr_0846 through to Bbr_0849, Bbr_1585 through to Bbr_1590,
315 and Bbr_1247 through to Bbr_1249; see Fig. 4 and Table 3).

316 Within the Bbr_0846-0849 gene cluster, which is separated from the *ats2* cluster by a single
317 gene (Fig. 3), Bbr_0846 (*nagA1*) and Bbr_0847 (*nagB2*) are predicted to encode an *N*-
318 acetylglucosamine-6-phosphate deacetylase and a glucosamine-6-phosphate deaminase,
319 respectively. Bbr_0848 (designated here as *nagK*) encodes a predicted ROK-family kinase,
320 which contains the characteristic DxGxT motif at its N-terminal end (75). The *B. breve*
321 UCC2003-encoded NagK protein exhibits 42 % similarity at protein level with the previously
322 characterized *E. coli* K-12-encoded, ROK-family NagK protein, which phosphorylates
323 GlcNAc to produce *N*-acetylglucosamine-6-phosphate (GlcNAc-6-P) (76). Therefore this
324 cluster is predicted to encode enzymes for the complete GlcNAc catabolic pathway as
325 previously described in *E. coli*, whereby GlcNAc is first phosphorylated by NagK, producing
326 GlcNAc-6-P, followed by NagA-mediated deacetylation to produce glucosamine-6-
327 phosphate, and the NagB-mediated deamination and isomerisation to produce fructose-6-

328 phosphate (76, 77). Bbr_0849 encodes a predicted transcriptional regulator from the ROK
329 family (designated here as *atsR2*).

330 The Bbr_1585-1590 cluster includes a predicted UDP-glucose-4-epimerase (Bbr_1585,
331 *galE*), a predicted *N*-acetylhexosamine-1-kinase (Bbr_1586, *nahK*) and a predicted lacto-*N*-
332 biose phosphorylase (Bbr_1586, *lnbP*), representing three of the four enzymes required for
333 the degradation of galacto-*N*-biose (Galβ1-3GalNAc; GNB), which is found in mucin, or
334 lacto-*N*-biose (Galβ1-3GlcNAc; LNB), a known HMO (78, 79). The other three genes of this
335 cluster, Bbr_1588-1590, encode a predicted ABC transport system, including two predicted
336 permease proteins and a solute binding protein, respectively (Fig. 4). This gene cluster was
337 previously shown to be transcriptionally up-regulated when *B. breve* UCC2003 was grown in
338 co-culture with *B. bifidum* PRL2010 in mucin (80).

339 Finally, the Bbr_1247-1249 cluster contains a gene specifying an *N*-acetylglucosamine-6-
340 phosphate deacetylase (Bbr_1247) and a glucosamine-6-phosphate deaminase (Bbr_1248)-
341 encoding gene, designated *nagA2* and *nagB3*, respectively. These genes were previously
342 shown to be up-regulated during *B. breve* UCC2003 growth on sialic acid (72). The NagA1
343 protein shares a 74 % identity with NagA2, while the NagB2 protein shares 84 % identity
344 with NagB1 of the *nan/nag* cluster for sialic acid metabolism (72) and 84 % identity with
345 NagB3. Bbr_1249 encodes a predicted transcriptional ROK family regulator (Fig. 4).

346

347 **Disruption of the *atsT* and *atsA2* genes.** In order to investigate if disruption of individual
348 genes from the *ats2* gene cluster would affect the ability of *B. breve* UCC2003 to utilize
349 GlcNAc-6-S, insertion mutants were constructed in the *atsT* and *atsA2* genes, resulting in
350 strains *B. breve* UCC2003-atsT and *B. breve* UCC2003-atsA2, respectively (see Materials
351 and Methods). The insertion mutants were analyzed for their ability to grow in mMRS

352 supplemented with GlcNAc-6-S as compared to *B. breve* UCC2003. As expected, and in
353 contrast to the wild type, there was a complete lack of growth of *B. breve* UCC2003-atsT and
354 *B. breve* UCC2003-atsA2 in media containing GlcNAc-6-S as the sole carbon source (Fig.
355 3B), thus demonstrating the involvement of the disrupted genes in GlcNAc-6-S metabolism.
356 Growth of the insertion mutants was not impaired on lactose, where all strains reached final
357 OD_{600nm} levels comparable to that reached by the wild type strain (Fig. 3B).

358

359 **Transcriptome of *B. breve* UCC2003-atsR2.** The Bbr_0846-0849 gene cluster, which is up-
360 regulated when *B. breve* UCC2003 is grown on GlcNAc-6-S, and the *ats2* gene cluster are
361 separated by just a single gene (Fig. 2). An insertion mutant was constructed in the predicted
362 ROK-type transcriptional regulator-encoding Bbr_0849 gene (*atsR2*). It was hypothesized
363 that if this gene encoded a repressor, mutation of the gene would lead to increased
364 transcription of the genes it controls even in the absence of the inducing carbohydrate.
365 Microarray data revealed that in comparison to *B. breve* UCC2003, the genes of the *ats2*
366 cluster were indeed significantly up-regulated (>3.0 fold change; P <0.001) in the mutant
367 strain, thus identifying *atsR2* as a transcriptional repressor (Table 4). Transcription of the
368 Bbr_0846-0849 gene cluster was down-regulated in the mutant strain as compared to the wild
369 type, when both strains were grown on ribose. It is speculated that, since *atsR2* represents the
370 first gene of this presumed operon (Fig. 2), the insertion mutation caused a (negative) polar
371 effect on the transcription of the downstream located genes.

372

373 **Electrophoretic mobility shift assays.** In order to determine if the AtsR2 protein directly
374 interacts with promoter regions of the *ats2* gene cluster, crude cell extracts of *L. lactis*
375 NZ9000-pNZ-atsR2 were used to perform EMSAs, with crude cell extracts of *L. lactis*

NZ9000-pNZ8048 (empty vector) used as a negative control. As expected, the negative control did not alter the electrophoretic behaviour of any of the tested DNA fragments (Fig. 5B). The results obtained with crude cell extract expressing AtsR2 demonstrate that this presumed regulator specifically binds to DNA fragments encompassing the upstream regions of *atsR2* and *atsT* (Fig. 5A and 5B). Dissection of the promoter region of *atsR2* showed that AtsR2 binding required a 184 bp region within which a 21 bp imperfect inverted repeat was identified. Similarly, dissection of the *atsT* promoter region revealed that AtsR2 binding required a 192 bp region which also includes a 21 bp imperfect repeat, similar to that identified upstream of *atsR2*. When either of the inverted repeats were excluded, binding of AtsR2 to such DNA fragments was abolished, suggesting that these inverted repeats contained the operator sequence of AtsR2 (Fig. 5A and 5B).

To demonstrate if AtsR2 binding to its DNA target is affected by the presence of a carbohydrate effector molecule, GlcNAc and GlcNAc-6-S were tested for their effects on the formation of the AtsR2-DNA complex. The ability of AtsR2 to bind to the promoter regions of *atsR2* or *atsT* was eliminated in the presence of 2.5 mM GlcNAc-6-S, the lowest concentration used in this assay. The presence of GlcNAc was shown to inhibit binding of AtsR2 to the *atsR2* and *atsT* promoter regions, yet only at GlcNAc concentrations above 5 mM (Fig. 5C). This suggests that while GlcNAc-6-S has the highest affinity for the regulator and is therefore the most likely effector of this repressor protein, the structurally similar GlcNAc is also able to bind this regulator, yet at concentrations that are probably not physiologically relevant.

Identification of the transcription start sites of *atsR2* and *atsT*. Based on the EMSA results and the transcriptome of *B. breve* UCC2003-atsR2, it was deduced that an AtsR2-

dependent promoter is located upstream of both *atsR2* and *atsT* (Fig. 1). In order to determine the transcriptional start site of these presumed promoters, primer extension analysis was performed using RNA extracted from *B. breve* UCC2003-atsR2 grown in mMRS supplemented with 0.5 % ribose. Microarray analysis had shown that the expression levels of *atsT* were high when the *B. breve* UCC2003-atsR2 strain was grown on ribose (Table 4). For this reason, the mutant strain was considered most suitable for primer extension analyis. For the *atsR2* promoter region, initial attempts to attain a primer extension product from mRNA isolated from *B. breve* UCC2003-atsR2 cells were unsuccessful. In an attempt to increase the amount of mRNA transcripts of this promoter region, a DNA fragment encompassing the deduced promoter region was cloned into pBC1.2 and introduced into *B. breve* UCC2003-atsR2, generating strain *B. breve* UCC2003-atsR2-pBC1.2-atsRProm. A primer extension product was obtained for the *atsT* promoter region using mRNA isolated from *B. breve* UCC2003-atsR2, therefore it was not necessary to clone this promoter. Single extension products were identified upstream of *atsR2* and *atsT* (Fig. 6). Potential promoter recognition sequences resembling consensus -10 and -35 hexamers were identified upstream of each of the transcription start sites (Fig. 6). The deduced operator sequences of AtsR2 overlap with the respective -35 or -10 sequences, consistent with our findings that AtsR2 acts as a transcriptional repressor.

Discussion

A large-scale metagenomic analysis of fecal samples from 13 individuals of various ages has revealed that genes predicted to encode anSMEs are enriched in the gut microbiomes of humans as compared to non-gut microbial communities (81). Interestingly, in the same study it was found that such genes are more commonly found in members of the gut microbiota of adults and weaned children, as compared to unweaned infants. The current study describes two gene clusters in an infant-isolated bacterium, namely *B. breve* UCC2003, each encoding a (predicted) sulfatase and accompanying anSME, as well as an associated transport system and transcriptional regulator. The *ats2* gene cluster was shown to be required for the metabolism of GlcNAc-6-S, while GlcNAc-3-S, GalNAc-3-S and GalNAc-6-S did not support growth of *B. breve* UCC2003. The substrate(s) for the sulfatase encoded by the *ats1* gene cluster is as yet unknown. However, as recently shown in a study of sulfatases from *Ba. thetaiotaomicron*, these enzymes can vary quite significantly in their substrate specificity. It is therefore possible that, similar to the BT_3349 and BT_1596 enzymes recently characterised from *Ba. thetaiotaomicron*, the AtsA1 sulfatase might be active on sulfated di- or oligosaccharides rather than monosaccharides (35) or that the transport system encoded by the *ats1* cluster is specific for an as yet unknown sulfated substrate. However, at the current time this is mere speculation and further study is required to expand this premise.

Interestingly, the two gene clusters, *ats1* and *ats2*, are quite dissimilar in terms of their genetic organization. The gene order and composition of the *ats1* cluster resembles that of a typical bifidobacterial carbohydrate utilization cluster as it includes genes encoding a predicted ABC-type transport system, a LacI-type repressor (*atsR1*) and the carbohydrate-active *atsA1*-encoded sulfatase and *atsB*-encoded anSME, which in this case replace the typical glycosyl hydrolase-encoding gene(s) (16, 82). In the *ats2* cluster, the *atsT* gene encodes a predicted transporter of the major facilitator superfamily, while the *atsA2* and

444 *atsB2* genes are adjacent, as is also the case for their homologous genes in *K. pneumoniae*
445 and *Prevotella* strain RS2 (83, 84). We obtained compelling evidence that the *ats2* cluster is
446 co-regulated with the Bbr_0846-0849 cluster by the ROK-family transcriptional repressor
447 AtsR2. The only previously characterised bifidobacterial ROK-family transcriptional
448 regulator is RafA, the transcriptional activator of the raffinose utilisation cluster in *B. breve*
449 UCC2003 (73). The Bbr_0846-0848 genes are presumed to be involved in the metabolism of
450 GlcNAc following the removal of the sulfate residue from GlcNAc-6-S. The fructose-6-
451 phosphate produced from GlcNAc by the combined activities of NagK, NagA and NagB is
452 expected to enter the fructose-6-phosphate phosphoketolase pathway or bifid shunt, the
453 central metabolic pathway of bifidobacteria (85). It is interesting that *B. breve* UCC2003 is
454 capable of growth on GlcNAc-6-S as a sole carbon source, but apparently not on GlcNAc
455 (16). Since the *B. breve* UCC2003 genome seems to encode the enzymes required to
456 metabolise GlcNAc, it suggests that the *atsT* transporter has (high) affinity for only the
457 sulfated form of this N-acetylated carbohydrate.

458 A novel method of desulfating mucin which does not require a sulfatase enzyme has been
459 characterised from *Prevotella* strain RS2, whereby a sulfoglycosidase removes GlcNAc-6-S
460 from purified porcine gastric mucin (86). The presence of a signal sequence on this
461 glycosulfatase (86), thus indicating extracellular activity, is interesting in relation to the
462 current study, as it presents a source of GlcNAc-6-S to *B. breve* strains, suggestive of a cross-
463 feeding opportunity for members of this species. This is particularly noteworthy when it is
464 considered that the sulfatase enzymes produced by *B. breve* UCC2003 are intracellular,
465 implying that *B. breve* UCC2003 is reliant on the extracellular glycosyl hydrolase activity of
466 other members of the gut microbiota in order to gain access to mucin-derived sulfated
467 monosaccharides. Recent studies have shown that *B. breve* UCC2003 employs a cross-
468 feeding strategy to great effect, as it can utilize components of 3' sialyllactose (a HMO) and

469 mucin following the degradation of these sugars by *B. bifidum* PRL2010, whereas in the
470 absence of *B. bifidum* PRL2010, it is not capable of utilising either of these sugars as a sole
471 carbon source (72, 80). A recent study has further provided transcriptomic evidence for
472 carbohydrate cross-feeding between bifidobacterial species. Four bifidobacterial strains,
473 namely *B. bifidum* PRL2010, *B. breve* 12L, *B. adolescentis* 22L and *B. longum* subsp.
474 *infantis* ATCC25697, were cultivated either in pairs (bi-association) or a combination of all
475 four strains (multi-association), under *in vivo* conditions in a murine model. In all strains,
476 transcription of predicted glycosyl hydrolase-encoding genes, particularly those involved in
477 xylose or starch utilization, were affected by co- or multi-association. In relation to xylose
478 metabolism, the authors speculated that in co- or multi-association, the combined glycosyl
479 hydrolase activities of the strains may allow them to degrade xylose-containing
480 polysaccharides which would otherwise be inaccessible (87).

481

482 In *Ba. thetaiotaomicron*, the *in vivo* contribution of sulfatase activity towards bacterial fitness
483 has been well-established. In previous studies of chondroitin sulfate and heparan sulfate
484 metabolism by this species, mutagenesis of a gene designated *chuR*, which was first predicted
485 to encode a regulatory protein but then later identified as an anSME, resulted in the inability
486 to compete with wild type *Ba. thetaiotaomicron* in germ-free mice (37, 88). In a recent study,
487 28 predicted sulfatase-encoding genes were identified on the genome of *Ba.*
488 *thetaiotaomicron*, 20 of which are predicted extracellular enzymes, yet the previously
489 described *chuR* gene is the sole anSME-encoding gene (36, 89, 90). Recently, this anSME
490 was shown to be of significant importance in this strain's ability to colonize the gut, as an
491 isogenic derivative of this strain (designated ΔanSME) carrying a deletion in the anSME-
492 encoding gene displayed reduced fitness *in vivo* (36). The authors have speculated that
493 anSME activity and associated sulfatase activities are important as the bacterium adapts to

494 the gut environment (36). Given that sulfatase activity within the *Bifidobacterium* genus is (at
495 least based on currently available genome sequences) limited to the *B. breve* species and a
496 single member of the *B. longum* subsp. *infantis* subspecies, it is interesting to speculate on the
497 effect this activity may have on bacterial fitness in the large intestine. It is intriguing to note
498 that human intestinal mucins increase in acidity along the intestinal tract, with more than half
499 of mucin oligosaccharide structures in the distal colon containing either sialic and/or sulfate
500 residues (91). We recently showed that 11 of 14 strains of *B. breve* tested were capable of
501 growth on sialic acid, while sialic acid utilization genes can also be found on the genomes of
502 *B. longum* subsp. *infantis* strains (20, 22, 72). The ability of *B. breve* strains and possibly
503 certain *B. longum* subsp. *infantis* strains, to utilise both sialic acid and sulfated GlcNAc-6-S
504 may provide them with a competitive advantage over other members of the *Bifidobacterium*
505 genus and other members of the gut microbiota, thus contributing to their successful
506 colonization ability in this highly competitive environment.

507

508 **Funding Information**

509 The APC Microbiome Institute is funded by Science Foundation Ireland (SFI), through the
510 Irish Government's National Development Plan. The authors and their work were supported
511 by SFI (Grant Nos. 07/CE/B1368 and SFI/12/RC/2273 for ME, MOCM and DvS, and Grant
512 No. 13/IA/1959 for HJ and SO) and a HRB postdoctoral fellowship (Grant No.
513 PDTM/20011/9) awarded to MOCM.

514

515 **References:**

516

517 1. **Turroni F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, Kerr C,**
518 **Hourihane J, Murray D, Fuligni F, Gueimonde M, Margolles A, De Bellis G, O'Toole**
519 **PW, van Sinderen D, Marchesi JR, Ventura M.** 2012. Diversity of bifidobacteria within
520 the infant gut microbiota. PLoS ONE **7:**e36957.

521 2. **Vaishampayan PA, Kuehl JV, Froula JL, Morgan JL, Ochman H, Francino MP.**
522 2010. Comparative metagenomics and population dynamics of the gut microbiota in mother
523 and infant. Genome Biol Evol **2:**53-66.

524 3. **Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y,**
525 **Xie H, Zhong H, Khan Muhammad T, Zhang J, Li J, Xiao L, Al-Aama J, Zhang D, Lee**
526 **Ying S, Kotowska D, Colding C, Tremaroli V, Yin Y, Bergman S, Xu X, Madsen L,**
527 **Kristiansen K, Dahlgren J, Wang J.** 2015. Dynamics and stabilization of the human gut
528 microbiome during the first year of life. Cell Host Microbe **17:**690-703.

529 4. **Jost T, Lacroix C, Braegger CP, Rochat F, Chassard C.** 2014. Vertical mother–neonate
530 transfer of maternal gut bacteria via breastfeeding. Environ Microbiol **16:**2891-2904.

531 5. **Tannock GW, Lawley B, Munro K, Gowri Pathmanathan S, Zhou SJ, Makrides M,**
532 **Gibson RA, Sullivan T, Prosser CG, Lowry D, Hodgkinson AJ.** 2013. Comparison of the
533 compositions of the stool microbiotas of infants fed goat milk formula, cow milk-based
534 formula, or breast milk. Appl Environ Microbiol **79:**3040-3048.

535 6. **Turroni F, Ribbera A, Foroni E, van Sinderen D, Ventura M.** 2008. Human gut
536 microbiota and bifidobacteria: from composition to functionality. Antonie van Leeuwenhoek
537 **94:**35-50.

538 7. **Andersson AF, Lindberg M, Jakobsson H, Bäckhed F, Nyrén P, Engstrand L.** 2008.
539 Comparative analysis of human gut microbiota by barcoded pyrosequencing. PloS one
540 **3:**e2836.

541 8. **Zwielehner J, Liszt K, Handschur M, Lassl C, Lapin A, Haslberger AG.** 2009.
542 Combined PCR-DGGE fingerprinting and quantitative-PCR indicates shifts in fecal
543 population sizes and diversity of *Bacteroides*, bifidobacteria and *Clostridium* cluster IV in
544 institutionalized elderly. Exp Gerontol **44:**440-446.

545 9. **Agans R, Rigsbee L, Kenche H, Michail S, Khamis HJ, Paliy O.** 2011. Distal gut
546 microbiota of adolescent children is different from that of adults. FEMS Microbiol Ecol
547 **77:**404-412.

548 10. **Nishijima S, Suda W, Oshima K, Kim S-W, Hirose Y, Morita H, Hattori M.** 2016.
549 The gut microbiome of healthy Japanese and its microbial and functional uniqueness. DNA
550 Research doi:10.1093/dnares/dsw002.

551 11. **Milani C, Mancabelli L, Lugli GA, Duranti S, Turroni F, Ferrario C, Mangifesta M,**
552 **Viappiani A, Ferretti P, Gorfer V, Tett A, Segata N, van Sinderen D, Ventura M.** 2015.
553 Exploring vertical transmission of Bifidobacteria from mother to child. Appl Environ
554 Microbiol **81:**7078-7087.

555 12. **Milani C, Andrea Lugli G, Duranti S, Turroni F, Mancabelli L, Ferrario C,**
556 **Mangifesta M, Hevia A, Viappiani A, Scholz M, Arioli S, Sanchez B, Lane J, Ward DV,**
557 **Hickey R, Mora D, Segata N, Margolles A, van Sinderen D, Ventura M.** 2015.

558 Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut.
559 Scientific Reports **5:**15782.

560 13. **Milani C, Turroni F, Duranti S, Lugli GA, Mancabelli L, Ferrario C, van Sinderen**
561 **D, Ventura M.** 2015. Genomics of the genus *Bifidobacterium* reveals species-specific
562 adaptation to the glycan-rich gut environment. Appl Environ Microbiol
563 doi:10.1128/aem.03500-15.

564 14. **Margolles A, de los Reyes-Gavilan CG.** 2003. Purification and functional
565 characterization of a novel alpha-L-arabinofuranosidase from *Bifidobacterium longum* B667.
566 Appl Environ Microbiol **69:**5096-5103.

567 15. **Pokusaeva K, O'Connell-Motherway M, Zomer A, MacSharry J, Fitzgerald GF, van**
568 **Sinderen D.** 2011. Cellodextrin Utilization by *Bifidobacterium breve* UCC2003. Appl
569 Environ Microbiol **77:**1681-1690.

570 16. **Pokusaeva K, Fitzgerald G, Sinderen D.** 2011. Carbohydrate metabolism in
571 Bifidobacteria. Genes Nutr **6:**285-306.

572 17. **O'Connell Motherway M, Kinsella M, Fitzgerald GF, van Sinderen D.** 2013.
573 Transcriptional and functional characterization of genetic elements involved in galacto-
574 oligosaccharide utilization by *Bifidobacterium breve* UCC2003. Microb Biotechnol **6:**67-79.

575 18. **O'Connell KJ, O'Connell Motherway M, O'Callaghan J, Fitzgerald GF, Ross RP,**
576 **Ventura M, Stanton C, van Sinderen D.** 2013. Metabolism of four α-glycosidic linkage-

577 containing oligosaccharides by *Bifidobacterium breve* UCC2003. Appl Environ Microbiol
578 **79:**6280-6292.

579 19. **Rivière A, Moens F, Selak M, Maes D, Weckx S, De Vuyst L.** 2014. The ability of
580 bifidobacteria to degrade arabinoxylan oligosaccharide constituents and derived
581 oligosaccharides is strain dependent. Appl Environ Microbiol **80:**204-217.

582 20. **O'Callaghan A, Bottacini F, O'Connell Motherway M, van Sinderen D.** 2015.
583 Pangenome analysis of *Bifidobacterium longum* and site-directed mutagenesis through by-
584 pass of restriction-modification systems. BMC Genomics **16:**1-19.

585 21. **Turroni F, Bottacini F, Foroni E, Mulder I, Kim J-H, Zomer A, Sanchez B, Bidossi**
586 **A, Ferrarini A, Giubellini V, Delledonne M, Henrissat B, Coutinho P, Oggioni M,**
587 **Fitzgerald G, Mills D, Margolles A, Kelly D, van Sinderen D, Ventura M.** 2010. Genome
588 analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived
589 glycan foraging. Proc Natl Acad Sci U S A **107:**19514-19519.

590 22. **Sela DA, Chapman J, Adeuya A, Kim JH, Chen F, Whitehead TR, Lapidus A,**
591 **Rokhsar DS, Lebrilla CB, German JB, Price NP, Richardson PM, Mills DA.** 2008. The
592 genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk
593 utilization within the infant microbiome. Proc Natl Acad Sci U S A **105:**18964-18969.

594 23. **Garrido D, Ruiz-Moyano S, Lemay DG, Sela DA, German JB, Mills DA.** 2015.
595 Comparative transcriptomics reveals key differences in the response to milk oligosaccharides
596 of infant gut-associated bifidobacteria. Sci Rep **5:**13517.

597 24. **Ruas-Madiedo P, Gueimonde M, Fernández-García M, de los Reyes-Gavilán CG,**
598 **Margolles A.** 2008. Mucin degradation by *Bifidobacterium* strains isolated from the human
599 intestinal microbiota. Appl Environ Microbiol **74:**1936-1940.

600 25. **Newburg D, Linhardt R, Ampofo S, Yolken R.** 1995. Human milk glycosaminoglycans
601 inhibit HIV glycoprotein gp120 binding to its host cell CD4. J Nutr **125:**419.

602 26. **Oshiro M, Ono K, Suzuki Y, Ota H, Katsuyama T, Mori N.** 2001.
603 Immunohistochemical localization of heparan sulfate proteoglycan in human gastrointestinal
604 tract. Histochem Cell Biol **115:**373-380.

605 27. **Eliakim R, Gilead L, Ligumsky M, Okon E, Rachmilewitz D, Razin E.** 1986.
606 Histamine and chondroitin sulfate E proteoglycan released by cultured human colonic
607 mucosa: indication for possible presence of E mast cells. Proc Natl Acad Sci U S A **83:**461-
608 464.

609 28. **Larsson JM, Karlsson H, Sjovall H, Hansson GC.** 2009. A complex, but uniform O-
610 glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn.
611 Glycobiology **19:**756-766.

612 29. **Guérardel Y, Morelle W, Plancke Y, Lemoine J, Strecker G.** 1999. Structural analysis
613 of three sulfated oligosaccharides isolated from human milk. Carbohydr Res **320:**230-238.

614 30. **Filipe M.** 1978. Mucins in the human gastrointestinal epithelium: a review. Invest Cell
615 Pathol **2:**195-216.

616 31. **Brockhausen I.** 2003. Sulphotransferases acting on mucin-type oligosaccharides.
617 Biochem Soc Trans **31:**318-325.

618 32. **Robertson AM, Wright DP.** 1997. Bacterial glycosulphatases and sulphomucin
619 degradation. Can J Gastroenterol **11:**361-366.

620 33. **Salyers AA, Vercellotti JR, West SE, Wilkins TD.** 1977. Fermentation of mucin and
621 plant polysaccharides by strains of *Bacteroides* from the human colon. Appl Environ
622 Microbiol **33:**319-322.

623 34. **Roberton A, McKenzie C, Sharfe N, Stubbs L.** 1993. A glycosulphatase that removes
624 sulphate from mucus glycoprotein. Biochem J **293:**683-689.

625 35. **Ulmer JE, Vilén EM, Namburi RB, Benjdia A, Beneteau J, Malleron A, Bonnaffé D,**
626 **Driguez P-A, Descroix K, Lassalle G, Le Narvor C, Sandström C, Spillmann D, Berteau**
627 **O.** 2014. Characterization of glycosaminoglycan (GAG) sulfatases from the human gut
628 symbiont *Bacteroides thetaiotaomicron* reveals the first GAG-specific bacterial
629 endosulfatase. J Biol Chem **289:**24289-24303.

630 36. **Benjdia A, Martens EC, Gordon JI, Berteau O.** 2011. Sulfatases and a radical S-
631 adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the
632 prominent human gut symbiont, *Bacteroides thetaiotaomicron*. J Biol Chem **286:**25973-
633 25982.

634 37. **Benjdia A, Subramanian S, Leprince J, Vaudry H, Johnson MK, Berteau O.** 2008.
635 Anaerobic sulfatase-maturating enzymes, first dual substrate radical S-adenosylmethionine
636 enzymes. J Biol Chem **283:**17815-17826.

637 38. **Berteau O, Guillot A, Benjdia A, Rabot S.** 2006. A new type of bacterial sulfatase
638 reveals a novel maturation pathway in prokaryotes. J Biol Chem **281:**22464-22470.

639 39. **Bond CS, Clements PR, Ashby SJ, Collyer CA, Harrop SJ, Hopwood JJ, Guss JM.**
640 1997. Structure of a human lysosomal sulfatase. Structure **5:**277-289.

641 40. **Lukatela G, Krauss N, Theis K, Selmer T, Gieselmann V, Von Figura K, Saenger W.**
642 1998. Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at
643 the active site suggest a novel mechanism for sulfate ester hydrolysis. Biochemistry **37:**3654-
644 3664.

645 41. **Schmidt B, Selmer T, Ingendoh A, Figurat Kv.** 1995. A novel amino acid modification
646 in sulfatases that is defective in multiple sulfatase deficiency. Cell **82:**271-278.

647 42. **Beil S, Kehrli H, James P, Staudenmann W, Cook AM, Leisinger T, Kertesz MA.**
648 1995. Purification and characterization of the arylsulfatase synthesized by *Pseudomonas*
649 *aeruginosa* PAO during growth in sulfate-free medium and cloning of the arylsulfatase gene
650 (*atsA*). Eur J Biochem **229:**385-394.

651 43. **Szameit C, Miech C, Balleininger M, Schmidt B, von Figura K, Dierks T.** 1999. The
652 iron sulfur protein AtsB is required for posttranslational formation of formylglycine in the
653 *Klebsiella* sulfatase. J Biol Chem **274:**15375-15381.

654 44. **Miech C, Dierks T, Selmer T, von Figura K, Schmidt B.** 1998. Arylsulfatase from
655 *Klebsiella pneumoniae* carries a formylglycine generated from a serine. J Biol Chem
656 **273:**4835-4837.

657 45. **Carlson BL, Ballister ER, Skordalakes E, King DS, Breidenbach MA, Gilmore SA,**
658 **Berger JM, Bertozzi CR.** 2008. Function and structure of a prokaryotic formylglycine-
659 generating enzyme. J Biol Chem **283:**20117-20125.

660 46. **Fang Q, Peng J, Dierks T.** 2004. Post-translational formylglycine modification of
661 bacterial sulfatases by the radical S-adenosylmethionine protein AtsB. J Biol Chem
662 **279:**14570-14578.

663 47. **Grove TL, Ahlum JH, Qin RM, Lanz ND, Radle MI, Krebs C, Booker SJ.** 2013.
664 Further characterization of Cys-type and Ser-type anaerobic sulfatase maturating enzymes
665 suggests a commonality in mechanism of catalysis. Biochemistry **52:**2874-2887.

666 48. **De Man JC, Rogosa M, Sharpe ME.** 1960. A medium for the cultivation of lactobacilli.
667 J Appl Bacteriol **23:**130-135.

668 49. **Sambrook J, Fritsch, E. F., and Maniatis, T.** 1989. Molecular cloning a laboratory
669 manual, 2nd ed. , 2nd Edition ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

670 50. **Terzaghi BE, Sandine WE.** 1975. Improved medium for lactic streptococci and their
671 bacteriophages. Appl Microbiol**:**807-813.

672 51. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B.**
673 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944-945.

674 52. **O'Connell Motherway M, Zomer A, Leahy SC, Reunanen J, Bottacini F, Claesson**
675 **MJ, O'Brien F, Flynn K, Casey PG, Moreno Munoz JA, Kearney B, Houston AM,**
676 **O'Mahony C, Higgins DG, Shanahan F, Palva A, de Vos WM, Fitzgerald GF, Ventura**
677 **M, O'Toole PW, van Sinderen D.** 2011. Functional genome analysis of *Bifidobacterium*
678 *breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved
679 host-colonization factor. Proc Natl Acad Sci USA **108:**11217-11222.

680 53. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment
681 search tool. J Mol Biol **215:**403-410.

682 54. **Crooks GE, Hon G, Chandonia J-M, Brenner SE.** 2004. WebLogo: A Sequence Logo
683 Generator. Genome Research **14:**1188-1190.

684 55. **Riordan O.** 1998. Studies on antimicrobial activity and genetic diversity of
685 *Bifidobacterium* species: molecular characterization of a 5.75 kb plasmid and a
686 chromosomally encoded *recA* gene homologue from *Bifidobacterium breve*.National
687 University of Ireland, Cork, Cork.

688 56. **Pokusaeva K, O'Connell-Motherway M, Zomer A, Fitzgerald GF, van Sinderen D.**
689 2009. Characterization of two novel alpha-glucosidases from *Bifidobacterium breve*
690 UCC2003. Appl Environ Microbiol **75:**1135-1143.

691 57. **O'Connell Motherway M, O'Driscoll J, Fitzgerald GF, Van Sinderen D.** 2009.
692 Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in
693 *Bifidobacterium breve* UCC2003. Microb Biotechnol **2:**321-332.

694 58. **Maze A, O'Connell-Motherway M, Fitzgerald G, Deutscher J, van Sinderen D.** 2007.
695 Identification and characterization of a fructose phosphotransferase system in
696 *Bifidobacterium breve* UCC2003. Appl Environ Microbiol **73:**545-553.

697 59. **Wells JM, Wilson PW, Le Page RWF.** 1993. Improved cloning vectors and
698 transformation procedure for *Lactococcus lactis*. J Appl Bacteriol **74:**629-636.

699 60. **van Hijum S, de Jong A, Baerends R, Karsens H, Kramer N, Larsen R, den Hengst**
700 **C, Albers C, Kok J, Kuipers O.** 2005. A generally applicable validation scheme for the
701 assessment of factors involved in reproducibility and quality of DNA-microarray data. BMC
702 Genomics **6:**77.

703 61. **Zomer A, Fernandez M, Kearney B, Fitzgerald GF, Ventura M, van Sinderen D.**
704 2009. An interactive regulatory network controls stress response in *Bifidobacterium breve*
705 UCC2003. J Bacteriol **191:**7039-7049.

706 62. **Pokusaeva K, Neves AR, Zomer A, O'Connell-Motherway M, MacSharry J, Curley**
707 **P, Fitzgerald GF, van Sinderen D.** 2010. Ribose utilization by the human commensal
708 *Bifidobacterium breve* UCC2003. Microb Biotechnol **3:**311-323.

709 63. **Mierau I, Kleerebezem M.** 2005. 10 years of the nisin-controlled gene expression
710 system (NICE) in *Lactococcus lactis*. Appl Microbiol Biotechnol **68:**705-717.

711 64. **Álvarez-Martín P, O'Connell-Motherway M, van Sinderen D, Mayo B.** 2007.
712 Functional analysis of the pBC1 replicon from *Bifidobacterium catenulatum* L48. Appl
713 Microbiol and Biotechnol **76:**1395-1402.

714 65. **de Ruyter PG, Kuipers OP, de Vos WM.** 1996. Controlled gene expression systems for
715 *Lactococcus lactis* with the food-grade inducer nisin. Appl Environ Microbiol **62:**3662-3667.

716 66. **Hamoen LW, Van Werkhoven AF, Bijlsma JJE, Dubnau D, Venema G.** 1998. The
717 competence transcription factor of *Bacillus subtilis* recognizes short A/T-rich sequences
718 arranged in a unique, flexible pattern along the DNA helix. Genes Dev **12:**1539-1550.

719 67. **Ventura M, Zink R, Fitzgerald GF, van Sinderen D.** 2005. Gene structure and
720 transcriptional organization of the *dnaK* operon of *Bifidobacterium breve* UCC2003 and
721 application of the operon in bifidobacterial tracing. Appl Environ Microbiol **71:**487-500.

722 68. **Dierks T, Lecca MR, Schlotterhose P, Schmidt B, von Figura K.** 1999. Sequence
723 determinants directing conversion of cysteine to formylglycine in eukaryotic sulfatases.
724 EMBO J **18:**2084-2091.

725 69. **Sardiello M, Annunziata I, Roma G, Ballabio A.** 2005. Sulfatases and sulfatase
726 modifying factors: an exclusive and promiscuous relationship. Hum Mol Gen **14:**3203-3217.

727 70. **Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, Miller NE.** 2001. Radical SAM, a
728 novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with
729 radical mechanisms: functional characterization using new analysis and information
730 visualization methods. Nucleic acids res **29:**1097-1106.

731 71. **Bottacini F, OConnell Motherway M, Kuczynski J, OConnell K, Serafini F, Duranti**
732 **S, Milani C, Turroni F, Lugli G, Zomer A, Zhurina D, Riedel C, Ventura M, Sinderen**
733 **D.** 2014. Comparative genomics of the *Bifidobacterium breve* taxon. BMC Genomics **15:**170.

734 72. **Egan M, O'Connell Motherway M, Ventura M, van Sinderen D.** 2014. Metabolism of
735 sialic acid by *Bifidobacterium breve* UCC2003. Appl Environ Microbiol **80:**4414-4426.

736 73. **O'Connell KJ, O'Connell Motherway M, Liedtke A, Fitzgerald GF, Ross RP,**
737 **Stanton C, Zomer A, van Sinderen D.** 2014. Transcription of two adjacent carbohydrate
738 utilization gene clusters in *Bifidobacterium breve* UCC2003 is controlled by LacI- and
739 Repressor Open reading frame Kinase (ROK)-type regulators. Appl Environ Microbiol
740 **80:**3604-3614.

741 74. **Egan M, O'Connell Motherway M, van Sinderen D.** 2015. A GntR-type transcriptional
742 repressor controls sialic acid utilization in *Bifidobacterium breve* UCC2003. FEMS
743 Microbiol Lett **362:**1-9.

744 75. **Conejo M, Thompson S, Miller B.** 2010. Evolutionary bases of carbohydrate
745 recognition and substrate discrimination in the ROK protein family. J Mol Evol **70:**545-556.

746 76. **Uehara T, Park JT.** 2004. The *N*-acetyl-D-glucosamine kinase of *Escherichia coli* and
747 its role in murein recycling. J Bacteriol **186:**7273-7279.

748 77. **White R.** 1968. Control of amino sugar metabolism in *Escherichia coli* and isolation of
749 mutants unable to degrade amino sugars. Biochem J **106:**847-858.

750 78. **Nishimoto M, Kitaoka M.** 2007. Identification of *N*-acetylhexosamine 1-kinase in the
751 complete lacto-*N*-biose I/galacto-*N*-biose metabolic pathway in *Bifidobacterium longum*.
752 Appl Environ Microbiol **73:**6444-6449.

753 79. **Kitaoka M, Tian J, Nishimoto M.** 2005. Novel putative galactose operon involving
754 lacto-*N*-biose phosphorylase in *Bifidobacterium longum*. Appl Environ Microbiol **71:**3158-
755 3162.

756 80. **Egan M, O'Connell Motherway M, Kilcoyne M, Kane M, Joshi L, Ventura M, van**
757 **Sinderen D.** 2014. Cross-feeding by *Bifidobacterium breve* UCC2003 during co-cultivation
758 with *Bifidobacterium bifidum* PRL2010 in a mucin-based medium. BMC Microbiol **14:**1-14.

759 81. **Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, Takami H, Morita**
760 **H, Sharma VK, Srivastava TP, Taylor TD, Noguchi H, Mori H, Ogura Y, Ehrlich DS,**
761 **Itoh K, Takagi T, Sakaki Y, Hayashi T, Hattori M.** 2007. Comparative metagenomics
762 revealed commonly enriched gene sets in human gut microbiomes. DNA Res **14:**169-181.

763 82. **Schell M, Karmirantzou M, Snel B, Vilanova D, Berger B, Pessi G, Zwahlen M-C,**
764 **Desiere F, Bork P, Delley M, Pridmore D, Arigoni F.** 2002. The genome sequence of
765 *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. Proc Natl
766 Acad Sci U S A **99:**14422-14427.

767 83. **Murooka Y, Ishibashi K, Yasumoto M, Sasaki M, Sugino H, Azakami H, Yamashita**
768 **M.** 1990. A sulfur- and tyramine-regulated *Klebsiella aerogenes* operon containing the
769 arylsulfatase (*atsA*) gene and the *atsB* gene. J Bacteriol **172:**2131-2140.

770 84. **Wright DP, Knight CG, Parkar SG, Christie DL, Roberton AM.** 2000. Cloning of a
771 mucin-desulfating sulfatase gene from *Prevotella* strain RS2 and its expression using a
772 *Bacteroides* recombinant system. J Bacteriol **182:**3002-3007.

773 85. **Scardovi V, Trovatelli ID.** 1965. The fructose-6-phosphate shunt as a peculiar pattern of
774 hexose degradation in the genus *Bifidobacterium*. Ann Microbiol**:**19-29.

775 86. **Rho JH, Wright DP, Christie DL, Clinch K, Furneaux RH, Roberton AM.** 2005. A
776 novel mechanism for desulfation of mucin: identification and cloning of a mucin-desulfating
777 glycosidase (sulfoglycosidase) from *Prevotella* strain RS2. J Bacteriol **187:**1543-1551.

778 87. **Turroni F, Milani C, Duranti S, Mancabelli L, Mangifesta M, Viappiani A, Andrea**
779 **Lugli G, Ferrario C, Gioiosa L, Ferrarini A, Li J, Palanza P, Delledonne M, van**
780 **Sinderen D, Ventura M.** 2016. Deciphering bifidobacterial-mediated metabolic interactions
781 and their impact on gut microbiota by a multi-omics approach. ISME J
782 doi:10.1038/ismej.2015.236.

783 88. **Cheng Q, Hwa V, Salyers AA.** 1992. A locus that contributes to colonization of the
784 intestinal tract by *Bacteroides thetaiotaomicron* contains a single regulatory gene (*chuR*) that
785 links two polysaccharide utilization pathways. J Bacteriol **174:**7185-7193.

786 89. **Sonnenburg ED, Sonnenburg JL, Manchester JK, Hansen EE, Chiang HC, Gordon**
787 **JI.** 2006. A hybrid two-component system protein of a prominent human gut symbiont
788 couples glycan sensing in vivo to carbohydrate metabolism. Proc Natl Acad Sci USA
789 **103:**8834-8839.

790 90. **Sonnenburg JL, Xu J, Leip DD, Chen C-H, Westover BP, Weatherford J, Buhler**
791 **JD, Gordon JI.** 2005. Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont.
792 Science **307:**1955-1959.

793 91. **Robbe C, Capon C, Maes E, Rousset M, Zweibaum A, Zanetta J-P, Michalski J-C.**
794 2003. Evidence of Regio-specific Glycosylation in Human Intestinal Mucins: Presence of an
795 acidic gradient along the intestinal tract. J Biol Chem **278:**46337-46348.

796 92. **Law J, Buist G, Haandrikman A, Kok J, Venema G, Leenhouts K.** 1995. A system to
797 generate chromosomal mutations in *Lactococcus lactis* which allows fast analysis of targeted
798 genes. J Bacteriol **177:**7011-7018.

799

800 **Figure 1: (A)** Synthesis of 6-*O*- and 3-*O*-sulfate-2-acetamido-2-deoxy-D-glucose **1** and **2** (i):
801 BnBr, NaH, LiBr, DMF; (ii): Ac_2O, Py; (iii): NaOMe, MeOH; (iv): $PhCH(OMe)_2$, HCOOH;
802 (v): $SO_3{\cdot}NEt_3$, Py, 85°C; (vi): 10% Pd/C, EtOH, 15 bar H_2 ; (vii): (1) TrCl, $CaSO_4$, Py, 100
803 °C, (2) Ac_2O; (viii): AcOH, HBr; (ix): $SO_3{\cdot}NEt_3$, DMF, 55 °C; (x): NaOMe, MeOH. **(B)**
804 Synthesis of 3-*O*- and 6-*O*-sulfate-2-acetamido-2-deoxy-D-galactose **3** and **4**. Key (i): Ac_2O,
805 Py; (ii): BnOH, $BF_3{\cdot}OEt_2$, CH_2Cl_2, 3 A MS; (iii): NaOMe, MeOH; (iv): $PhCH(OMe)_2$,
806 HCOOH; (v) $SO_3{\cdot}NEt_3$, DMF, 55 °C; (vi): 10% Pd/C, EtOH, 15 bar H_2; (vii): NaOMe,
807 MeOH; (viii): $Me_2C(OMe)_2$, p-TSA, DMF, 65°C; (xi): $SO_3{\cdot}NEt_3$, DMF, 55°C; (x):
808 CF_3COOH, H_2O; (xi): 10% Pd/C, EtOH, 10 bar H_2.

809

810 **Figure 2:** Comparison of the sulfatase and anSME-encoding gene clusters of *B. breve*
811 UCC2003 with corresponding loci in the currently available complete *B. breve* genome
812 sequences and *B. longum* subsp. *infantis* BT1. Each solid arrow represents an open reading
813 frame. The length of the arrows (which contain the locus tag number) is proportional to the
814 size of the open reading frame. The corresponding gene name, which is indicative of putative
815 function, is given above relevant arrows at the top of the figure. Orthologs are marked with
816 the same colour. The amino acid identity of each predicted protein to its equivalent protein
817 encoded by *B. breve* UCC2003, expressed as a percentage, is given above each arrow.

818

819 **Figure 3: (A)** Final OD_{600nm} values obtained following 24 h growth of *B. breve* UCC2003 on
820 mMRS without supplementation with a carbon source (negative control) or containing 0.5 %
821 (wt/vol) lactose, GlcNAc-6-S, GlcNAc-3-S, GalNAc-6-S or GalNAc-3-S as the sole carbon
822 source. **(B)** Final OD_{600nm} values obtained following 24 h growth of *B. breve* UCC2003, *B.*
823 *breve* UCC2003-atsT and *B. breve* UCC2003-atsA2 in modified MRS without

824 supplementation with a carbon source(negative control, horizontally striped bars) or
825 containing 0.5 % (wt/vol) lactose (diagonally striped bars) or GlcNAc-6-S (solid grey filled
826 bars) as the sole carbon source. The results are the mean values obtained from two separate
827 experiments. Error bars represent the standard deviation.
828

829 **Figure 4:** Schematic representation of the four *B. breve* UCC2003 gene clusters up-regulated
830 during growth on GlcNAc-6-S as the sole carbon source. The length of the arrows (which
831 contain the locus tag number) is proportional to the size of the open reading frame and the
832 gene locus name, which is indicative of its putative function, is given at the top. Genes are
833 grouped by colour based on their predicted function in carbohydrate metabolism.
834

835 **Figure 5: (A)** Schematic representation of the *ats2* gene cluster of *B. breve* UCC2003 and
836 DNA fragments used in EMSAs for the *atsR2* and *atsT* promoter regions, together with
837 Weblogo representation of the predicted operator of AtsR2. Plus or minus signs indicate
838 ability or inability of AtsR2 to bind to the DNA fragment. The bent arrows represent the
839 position and direction of the proven promoter sequences (see Fig. 6). **(B)** EMSAs showing
840 the interactions of (I) crude cell extract containing pNZ-AtsR2 with the DNA fragments R1,
841 R2, R3,T1, T2 and T3 and (II) crude cell extract containing pNZ8048 (empty vector) with the
842 DNA fragments R1 and T1. The minus symbol indicates reactions to which no crude cell
843 extract was added, while the remaining lanes represent binding reactions with the respective
844 DNA probes incubated with increasing amounts of crude cell extract. Each successive lane
845 from right to left represents a doubling of the amount of crude cell extract. **(C)** EMSAs
846 showing AtsR2 interaction with the DNA fragments R1 and T1 with the addition of GlcNAc
847 or GlcNAc-6-S in concentrations ranging from 2.5 mM to 20 mM.

848

849 **Figure 6:** Schematic representation of the *atsR2* (panel A), *atsT* (panel B), promoter regions.
850 Boldface type and underlining indicate -10 and -35 hexamers (as deduced from the primer
851 extension results) and ribosomal binding site (RBS); the transcriptional start site is indicated
852 by an asterisk. The arrows underneath the indicated DNA sequences indicate the inverted
853 repeats that represent the presumed AtsR2 binding site. The arrows in the right panels
854 indicate the primer extension products.

855

856

857

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

Strains and plasmids	Relevant features	Reference or source
Strains		
***Escherichia coli* strains**		
***E.coli* EC101**	Cloning host; *repA*$^+$ *kmr*	(92)
***E.coli* EC101-pNZ-M.Bbrll+Bbr1**	EC101 harboring a pNZ8048 derivative containing bbrllM and bbrlllM	(57)
***E. coli* XL1-blue**	*(supE44 hsdR17 recA1 gyrA96 thi relA1 lac* F′ [*proAB*$^+$ *lacI*^{q} *lacZΔM15* Tn*10*(Tet^r)])	Stratagene
***E. coli* XL1-blue-pBC1.2-atsProm**	XL1-blue harboring pBC1.2-atsProm	This study
***L. lactis* strains**		
***L. lactis* NZ9000**	MG1363, *pep*N::*nis*RK, nisin inducible overexpression host	(65)
***L. lactis* NZ9000-pNZ8048**	NZ9000 containing pNZ8048	This study
***L. lactis* NZ9000-pNZ-atsR2**	NZ9000 containing pNZ-atsR2	This study
***L. lactis* NZ97000**	Nisin-A producing strain	(65)
***Bifidobacterium* sp. strains**		
***B. breve* UCC2003**	Isolate from a nursling stool	(58)
***B. breve* UCC2003-atsR2**	pORI19-tetW-atsR2 insertion mutant of *B. breve* UCC2003	This study
***B. breve* UCC2003-atsT**	pORI19-tetW-atsT insertion mutant of *B. breve* UCC2003	This study
***B. breve* UCC2003-atsA2**	pORI19-tetW-atsA2 insertion mutant of *B. breve* UCC2003	This study
***B. breve* UCC2003-atsR2-pBC1.2-atsProm**	pORI19-tetW-atsR2 insertion mutant of UCC2003 containing pBC1.2-atsProm	This study
Plasmids		
pAM5	pBC1-pUC19-Tc^r	(64)
pNZ8048	Cm^r, nisin-inducible translational fusion vector	(65)
pNZ-atsR2	Cm^r, pNZ8048 derivative containing translational fusion of *atsR2* encoding DNA fragment to nisin-inducible promoter	This study
pORI19	Em^r, repA$^-$, ori$^+$, cloning vector	(92)
pORI19-tetW-atsR2	Internal 408 bp fragment of *atsR2* and *tetW* cloned in pORI19	This study
pORI19-tetW-atsT	Internal 416 bp fragment of *atsT* and *tetW* cloned in pORI19	This study
pORI19-tetW-atsA2	Internal 402 bp fragment of *atsA2* and *tetW* cloned in pORI19	This study
pBC1.2	pBC1-pSC101-Cm^r	(64)
pBC1.2-atsProm	AtsR2 promoter region cloned in pBC1.2	This study

Table 2: Oligonucleotide primers used in this study.

Purpose	Primer	Sequence
Cloning of 408 bp fragment of *atsR2* in pORI19	AtsR2F	GACTAG*AAGCTT*GCCATCACGATCGACGACG
	AtsR2R	TAGCAT*TCTAGA*GCATCCCGGACGTCCACAG
Cloning of 416 bp fragment of *atsT* in pORI19	AtsTF	GACTAG*AAGCTT*GATCTCCTTCCGCCAGCTC
	AtsTR	TAGCAT*TCTAGA*CGTTGGTGCCGGTCAGCTG
Cloning of 402 bp fragment of *atsA2* in pORI19	AtsA2F	GACTAG*AAGCTT*GAATACGTCGCCTGGCTCAAG
	AtsA2R	TAGCAT*TCTAGA*CCTCCACTGGTCGTTGTCG
Amplification of *tetW*	TetWF	TCAGCT*GTCGAC*ATGCTCATGTACGGTAAG
	TetWR	GCGACG*GTCGAC*CATTACCTTCTGAAACATA
Confirmation of site-specific homologous recombination	AtsR2confirm	CATCGACACGGCATACTGG
	AtsTconfirm	CATCTTCGGCGCGTTATG
	AtsA2confirm	GGAAACCGACTGGACCTACAC
Cloning of *atsR2* in pNZ8048	AtsR2FOR	TACGTA*CCATGG*TGCATTTCGCATCGG
	AtsR2REV	GCTAGC*TCTAGA*GTGGAATATGCGGTGCGTG
Cloning of *atsR2* promoter in pBC1.2	AtsRPromF	GTACTA*AAGCTT*CCAGTATGCCGTGTCGATG
	AtsRPromR	TAGCTA*TCTAGA*CGCAATGCCAGAAACTCAGC
IRD-labelled primers	AtsR2R1F	CATCGTGTTATTGGCGCGG
	AtsR2R1R	GACGCCATATCACAGAGGGTTG
	AtsR2R2F	GCATGCGGCGTGAACTCC
	AtsR2R2R	CGCAATGCCAGAAACTCAGC
	AtsR2R3F	GATGTTGCCTTGCGGTATG
	AtsR2R3R	CAACGGCTGCCCACTGG
	AtsR2T1F	GGTCCTCCTTCGTCTGTGTGG
	AtsR2T1R	GTCGTGGCATATCGTTCGG
	AtsR2T2F	GGGCCGACGAAGTTGTTG
	AtsR2T2R	CGATGAGACCGCCGATG
	AtsR2T3F	CTAGCGGCATTCAGTATCGAG
	AtsR2T3R	GCGGCAGAACAGCAGGAAC

Restriction sites incorporated into oligonucleotide primer sequences are indicated in italics.

Table 3: Effect of GlcNAc-6-S on the transcriptome of *B. breve* UCC2003

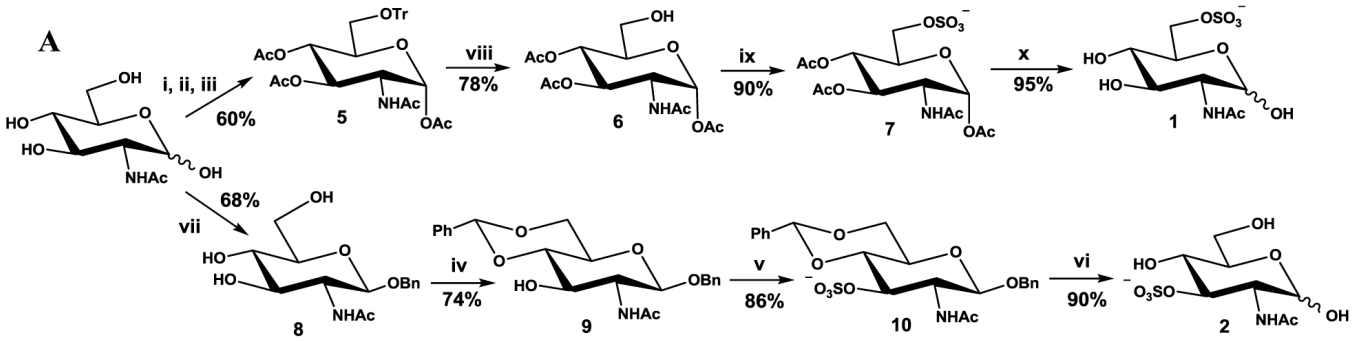
Locus tag (gene name)	Predicted Function	Level of up-regulation
Bbr_0846 (*nagA1*)	*N*-acetylglucosamine-6-phosphate deacetylase	12.63
Bbr_0847 (*nagB2*)	Glucosamine-6-phosphate isomerase	6.17
Bbr_0848 (*nagK*)	Sugar kinase, ROK family	9.85
Bbr_0849 (*atsR2*)	Transcriptional regulator, ROK family	8.58
Bbr_0851 (*atsT*)	Carbohydrate transport protein	96.75
Bbr_0852 (*atsA2*)	Sulfatase	35.36
Bbr_0853 (*atsB2*)	anSME	31.25
Bbr_0854	Hypothetical membrane spanning protein	4.175
Bbr_1247 (*nagA2*)	*N*-acetylglucosamine-6-phosphate deacetylase	10.84
Bbr_1248 (*nagB3*)	Glucosamine-6-phosphate isomerase	11.88
Bbr_1249	Transcriptional regulator, ROK family	3.07
Bbr_1585 (*galE*)	UDP-glucose 4-epimerase	3.09
Bbr_1586 (*nahK*)	*N*-acetylhexosamine kinase	4.96
Bbr_1587 (*lnbP*)	lacto-*N*-biose phosphorylase	6.58
Bbr_1588	Permease protein of ABC transporter system	6.24
Bbr_1589	Permease protein of ABC transporter system	8.27
Bbr_1590	Solute-binding protein of ABC transporter system	23.97

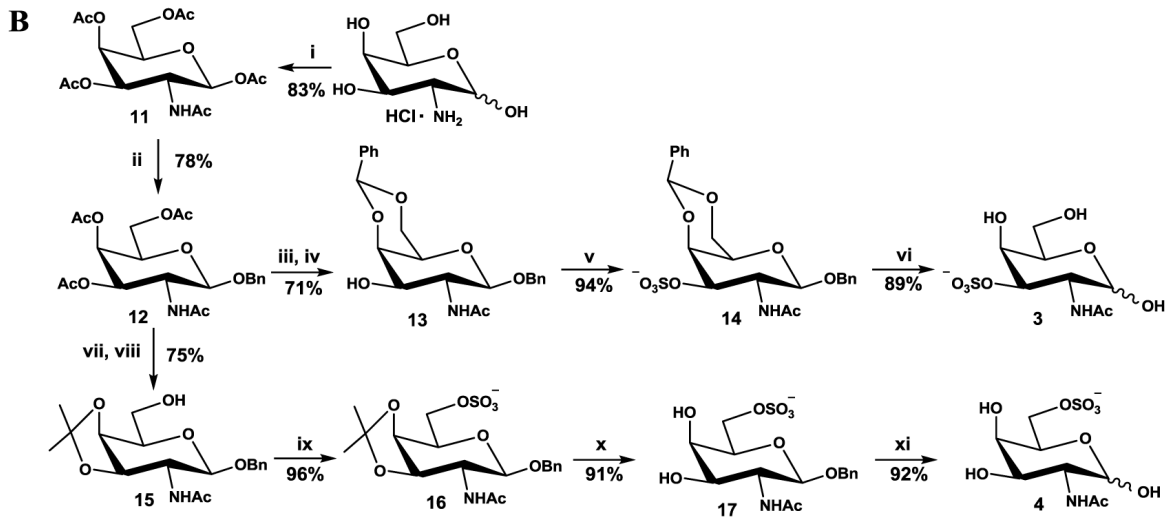
The cutoff point is 3- fold with a *P*-value of <0.001.

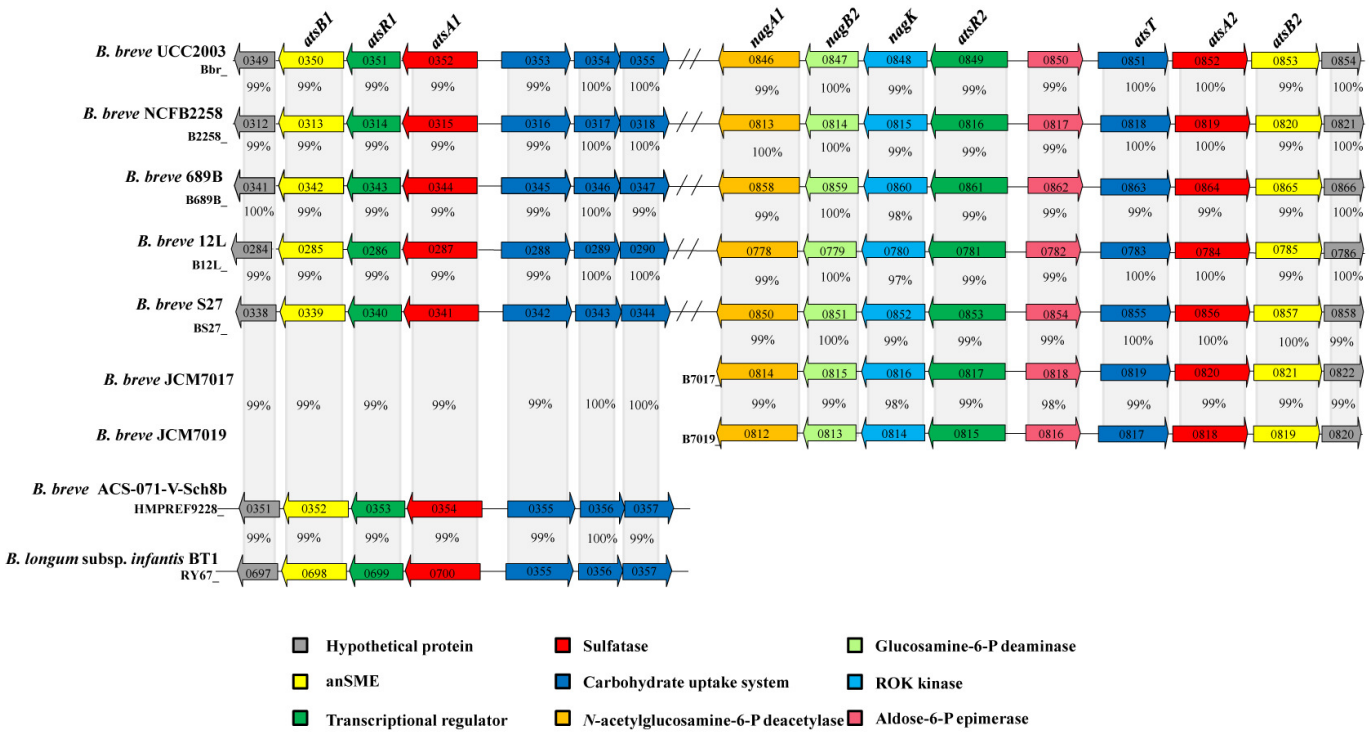
Table 4: Transcriptome analysis of *B. breve* UCC2003-atsR2 as compared to *B. breve* UCC2003 grown on 0.5 % (wt/vol) ribose.

Locus tag (gene name)	Predicted Function	Fold up-regulation	Fold down-regulation
Bbr_0846 (*nagA1*)	*N*-acetylglucosamine-6-phosphate deacetylase	-	3.77
Bbr_0847 (*nagB2*)	Glucosamine-6-phosphate isomerase	-	3.35
Bbr_0848 (*nagK*)	Sugar kinase, ROK family	-	4.45
Bbr_0850	Aldose-1-epimerase	4.58	-
Bbr_0851 (*atsT*)	Carbohydrate transport protein	106.28	-
Bbr_0852 (*atsA2*)	Sulfatase	59.58	-
Bbr_0853 (*atsB2*)	anSME	15.57	-
Bbr_0854	Hypothetical membrane spanning protein	9.09	-

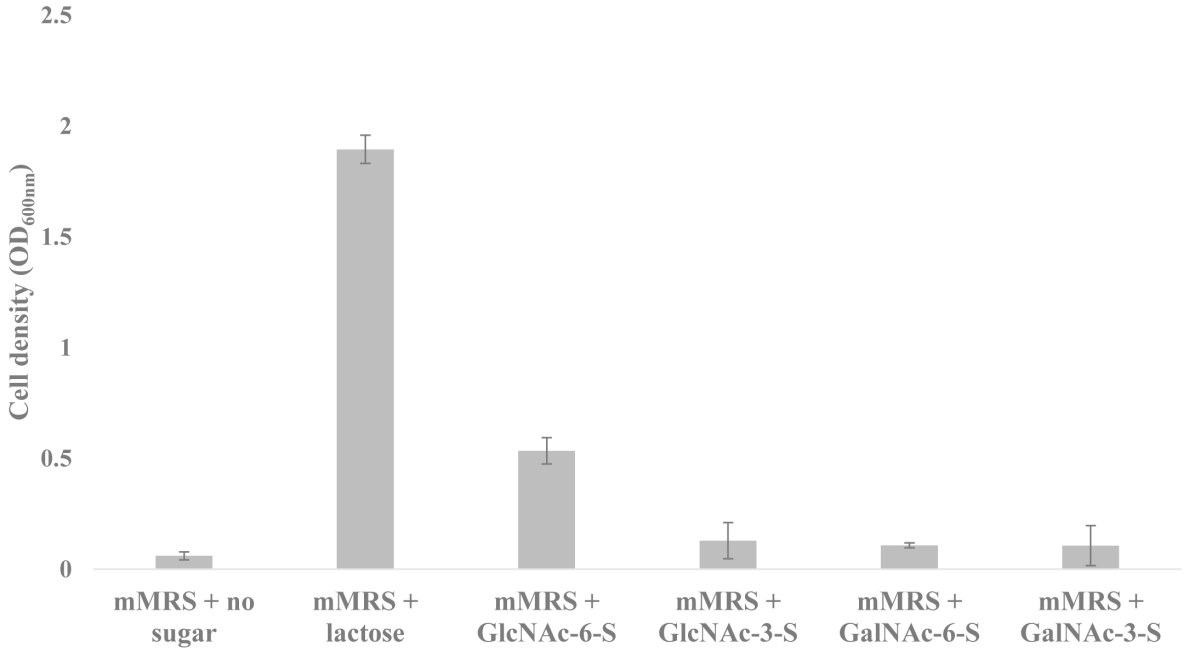
The cutoff point is 3- fold with a *P*-value of <0.001; values below the cutoff are indicated by a minus.



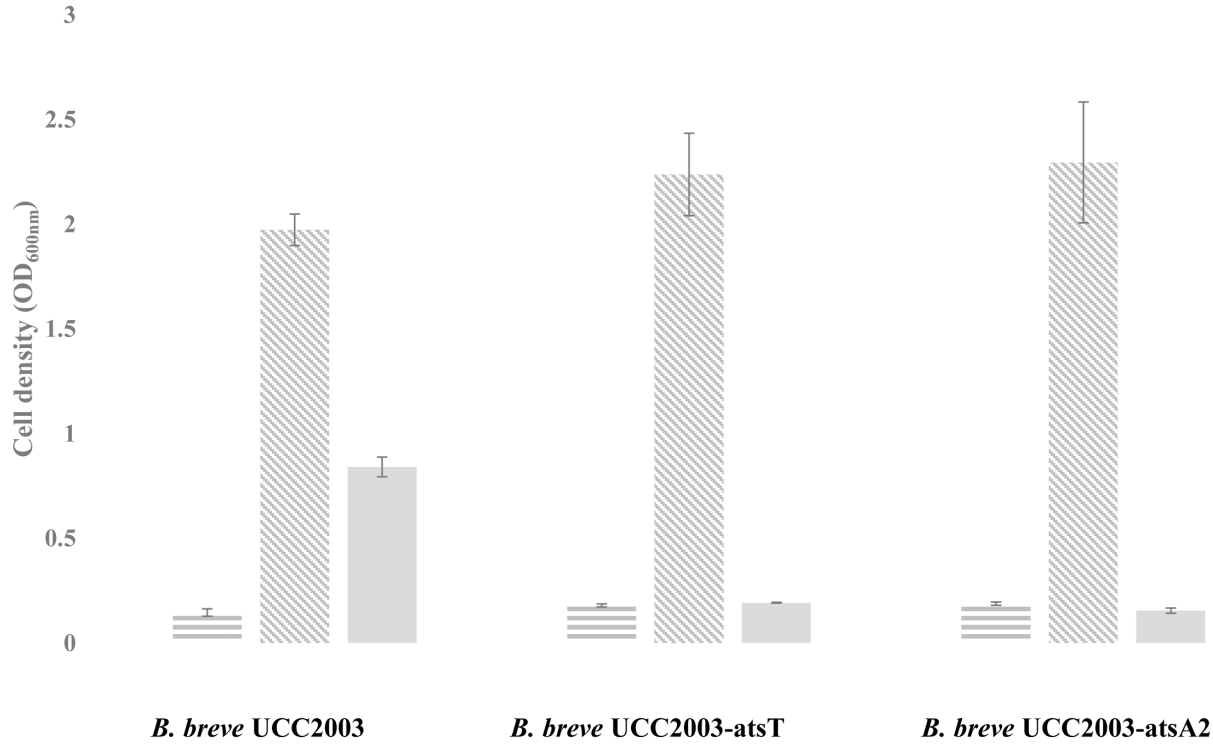


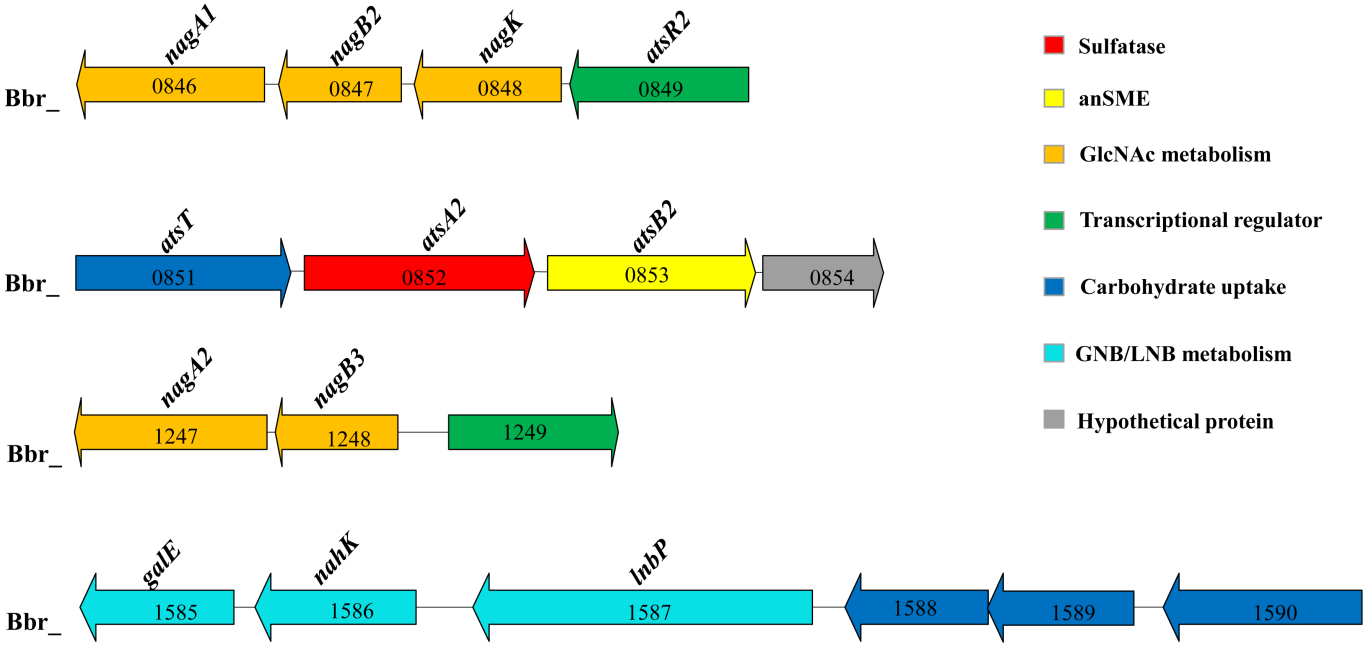


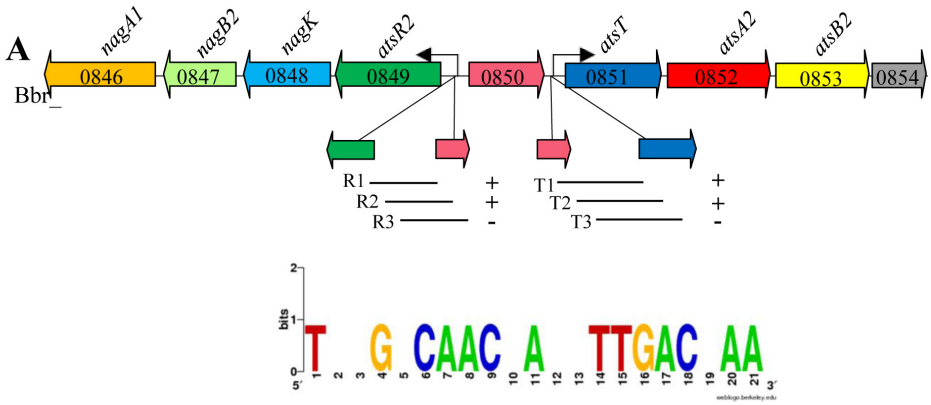
A

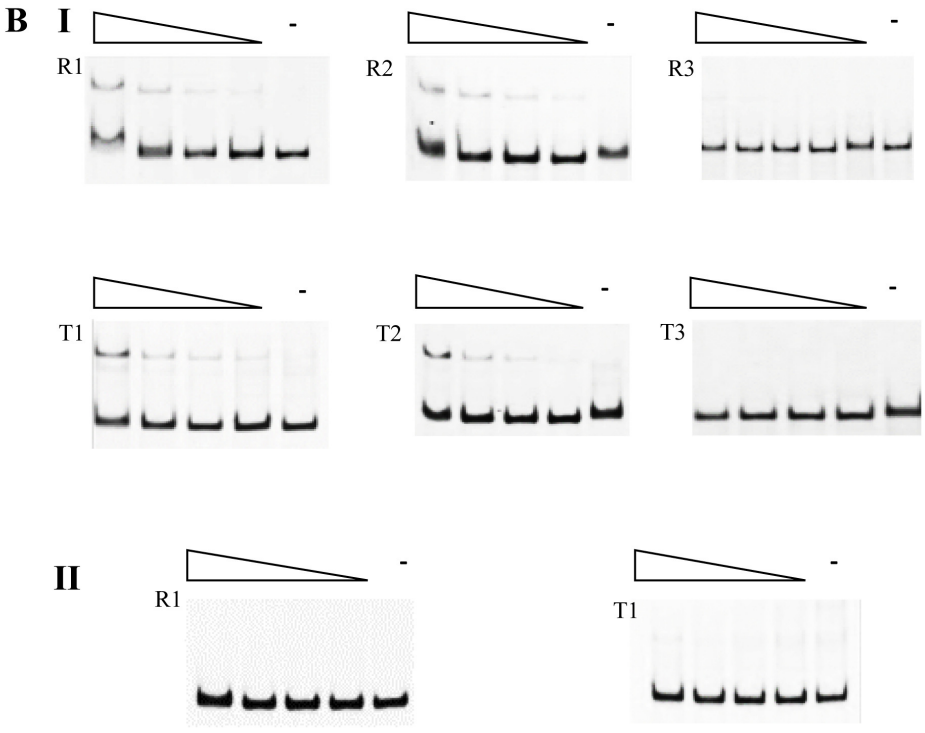


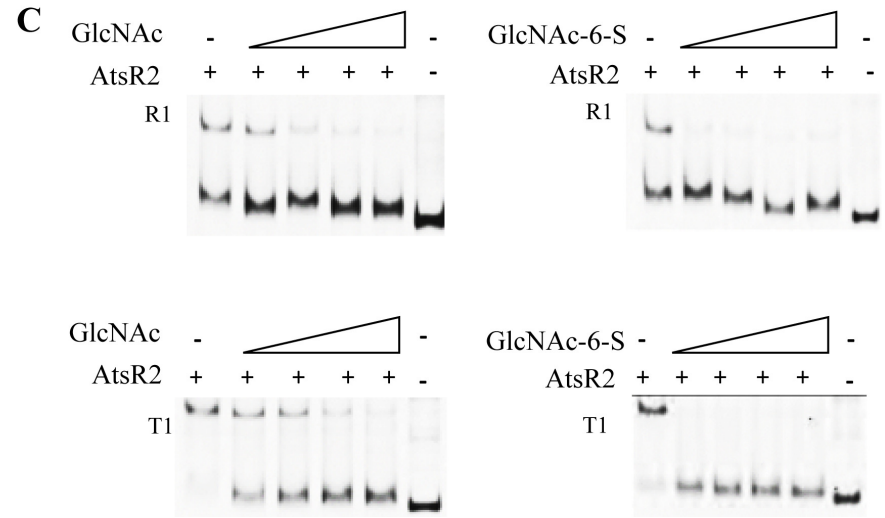
B











A

ATTGTATCCATCCCGACATATTTAGGCAACCATC**TTGACT**AATTAATTT

AtsR2 -35

-10

CACCATGC**TATATT**TATCAATTGAAGCCTATCAACGAGCCGATTAGCC

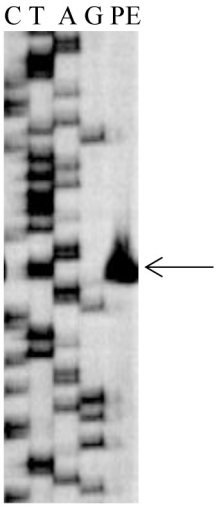
*

GCGCACAAGACTCGAAAAAGGAGTTCACGCCGCATGCCCGCCCGCA

AAGATGACACAATAGGTATTGGCATCATCGAAAAGCGCCTCTTCAAGC

RBS

GCTGACGGGAGCTTTTCATG



B

TTCTTCACCGTGAACCATTGATATTTCAGTAATCTTTGAGTCTTTTGCG

-35 AtsR2 -10

TGCGTTATTCCTAAATATG**TCAACA**AGGTTGACGAAATGATG**TATACT**

*

GAATCACGCCAGCTATCACTAGCGGCATTCAGTATCGAGATTCAAAGG

RBS

AGAAAAGGATACTTTCATG