<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Identification and characterization of a glycosulfatase-encoding gene cluster in Bifidobacterium breve UCC2003</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Egan, Muireann; Jiang, Hao; O’Connell Motherway, Mary; Oscarson, Stefan; van Sinderen, Douwe</td>
</tr>
<tr>
<td><strong>Publication date</strong></td>
<td>2016-09</td>
</tr>
<tr>
<td><strong>Type of publication</strong></td>
<td>Article (peer-reviewed)</td>
</tr>
</tbody>
</table>
| **Link to publisher's version** | http://aem.asm.org/content/early/2016/08/29/AEM.02022-16.abstract  
http://dx.doi.org/10.1128/AEM.02022-16  
Access to the full text of the published version may require a subscription. |
| **Rights** | © 2016, American Society for Microbiology. All Rights Reserved. |
| **Embargo information** | Access to this article is restricted until 6 months after publication by the request of the publisher. |
| **Embargo lift date** | 2017-03-02 |
| **Item downloaded from** | http://hdl.handle.net/10468/3130 |

Downloaded on 2019-11-30T03:49:45Z
Identification and characterization of a glycosulfatase-encoding gene cluster in

Bifidobacterium breve UCC2003

Muireann Egan¹, Hao Jiang², Mary O’Connell Motherway¹, Stefan Oscarson², Douwe van Sinderen¹

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

²Centre 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-acetylgalactosamine-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-acetylgalactosamine-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 cellooligosaccharides (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 Cys58 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 Ser72 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).
Chemical synthesis of sulfated monosaccharides. In brief, the 6-\(O\)-sulfated GlcNAc structure (Fig. 1A, structure 1) was synthesized in four steps from GlcNAc in an overall 40 % yield while the other three target structures, 3-\(O\)-sulfated GlcNAc (Fig. 1A, 2), 3-\(O\)-sulfated GalNAc and 6-\(O\)-sulfated GalNAc (Fig. 1B, 3 and 4, respectively), were synthesized from their corresponding benzyl \(\beta\)-glycoside, (Fig. 1A, 8 and Fig. 1B, 12), in three or four steps with an overall yield of about 60 %. The benzyl glycoside was obtained either by direct alkylation of a hemiacetal (Fig. 1A, 8, GlcNAc) or by glycosylation of a peracetylated precursor (Fig. 1B, 12, GalNAc). Sulfations were performed using a \(\text{SO}_3\cdot\text{NEt}_3\) complex in pyridine or DMF (yields 86-96 %). Direct regioselective 6-\(O\)-tritylation of GlcNAc followed by \textit{in situ} acetylation afforded compound 5 from which the trityl group was removed using aqueous acetic acid, without any acetyl migration detected, to yield the 6-OH derivative 6, sulfation of which gave compound 7 which was subsequently deacetylated using Zemplen conditions to afford target structure 1 (Fig. 1A). Benzylidenation of compounds 8 and 12 gave 3-OH compounds 9 and 13, respectively. Sulfation (\(\rightarrow\)10 and 14) followed by deprotection through catalytic hydrogenolysis yielded target structures 2 and 3. Isopropylidenation of compound 12 gave the 6-OH compound 15, which was sulfated (\(\rightarrow\)16) and then deprotected through acetal hydrolysis (\(\rightarrow\)17) followed by catalytic hydrogenolysis to afford target structure 4 (Fig. 1). The experimental methods are described in further detail in the supplementary material.

Nucleotide sequence analysis. Sequence data were obtained from the Artemis-mediated genome annotations of \textit{B. breve} UCC2003 (51, 52). Database searches were performed using the non-redundant sequence database accessible at the National Centre for Biotechnology.
DNA manipulations. Chromosomal DNA was isolated from *B. breve* UCC2003 as previously described (55). Plasmid DNA was isolated from *E. coli*, *L. lactis* and *B. breve* using the Roche High Pure plasmid isolation kit (Roche Diagnostics, Basel, Switzerland). An initial lysis step was performed using 30 mg ml\(^{-1}\) of lysozyme for 30 min at 37°C prior to plasmid isolation from *L. lactis* or *B. breve* (56). Single stranded oligonucleotide primers used in this study were synthesized by Eurofins (Ebersberg, Germany) (Table 2). Standard PCRs were performed using Taq PCR master mix (Qiagen GmBH, Hilden, Germany). *B. breve* colony PCRs were carried out as described previously (57). PCR fragments were purified using the Roche High Pure PCR purification kit (Roche Diagnostics).

Electroporation of plasmid DNA into *E. coli*, *L. lactis* or *B. breve* was performed as previously described (49, 58, 59).

*Construction of B. breve UCC2003 insertion mutants.* Internal fragments of Bbr_0849, designated here as *atsR2* (fragment encompasses 408 bp, representing codon numbers 134 through to 271 of the 395 codons of this gene), Bbr_0851, designated *atsT* (fragment encompasses 416 bp, representing codon numbers 149 through to 288 of the 476 codons of this gene) and Bbr_0852, designated *atsA2* (fragment encompasses 402 bp, representing codon numbers 148 through to 281 of the 509 codons of this gene) were amplified by PCR.
using *B. breve* UCC2003 chromosomal DNA as a template and primer pairs atsR2F and atsR2R, atsTF and atsTR, and atsA2F and atsA2R, respectively (Table 2). The insertion mutants were constructed as described previously (57). Site-specific recombination of potential Tet-resistant mutants was confirmed by colony PCR using primer combinations TetWF and TetWR to verify *tetW* gene integration, and the primers atsR2confirm, atsTconfirm and atsA2confirm (positioned upstream of the selected internal fragments of atsR2, atsT and atsA2, respectively) in combination with primer TetWF to confirm integration at the correct chromosomal location.

Analysis of global gene expression using *B. breve* DNA microarrays. Global gene expression was determined during log-phase growth (OD$_{600nm}$ of ~0.5) of *B. breve* UCC2003 in mMRS supplemented with 0.5 % GlcNAc-6-S and the obtained transcriptome was compared to that obtained from *B. breve* UCC2003 grown in mMRS supplemented with 0.5 % ribose. Similarly, global gene expression of the insertion mutant *B. breve* UCC2003-atsR2 was determined during log-phase (OD$_{600nm}$ of ~0.5) growth of the mutant in mMRS supplemented with 0.5 % ribose and the transcriptome was also compared to that from *B. breve* UCC2003 grown in 0.5 % ribose. DNA microarrays containing oligonucleotide primers representing each of the 1864 identified open reading frames on the genome of *B. breve* UCC2003 were designed and obtained from Agilent Technologies (Palo Alto, Ca., USA). RNA was isolated and purified from bifidobacterial cells using a combination of the “Macaloid” method and the Roche High Pure RNA isolation kit, as previously described (60). RNA was quantified spectrophotometrically as described by Sambrook *et al.* (49). Methods for complementary DNA synthesis and labelling were performed as described previously (61). Hybridization, washing of the slides and processing of the DNA-microarray data was also performed as previously described (62).
Plasmid Constructions. For the construction of plasmid pNZ-atsR2, a DNA fragment encompassing the complete coding region of the predicted transcriptional regulator atsR2 (Bbr_0849) was generated by PCR amplification from chromosomal DNA of B. breve UCC2003 using PfuUltra II DNA polymerase (Agilent Technologies) and the primer combination atsR2FOR and atsR2REV (Table 2). The generated amplicon was digested with NcoI and XbaI, and ligated into the similarly digested, nisin-inducible translational fusion plasmid pNZ8048 (63). The ligation mixture was introduced into L. lactis NZ9000 by electrotansformation and transformants were selected based on Cm resistance. The plasmid content of a number of Cmr transformants was screened by restriction analysis and the integrity of positively identified clones was verified by sequencing.

To clone the Bbr_0849 promoter region, a DNA fragment encompassing the intergenic region between the Bbr_0849 and Bbr_0850 genes was generated by PCR amplification employing B. breve UCC2003 chromosomal DNA as a template, and using PfuUltra II DNA polymerase in combination with primer pair atsRPromF and atsRPromR (Table 2). The PCR product was digested with HindIII and XbaI, and ligated to the similarly digested pBC1.2 (64). The ligation mixture was introduced into E. coli XL1-blue by electrotansformation and transformants were selected based on Tet and Cm resistance. Transformants were checked for plasmid content by restriction analysis and the integrity of several positively identified recombinant plasmids was verified by sequencing. One of these verified recombinant plasmids, designated pBC1.2-atsProm, was introduced into B. breve UCC2003-atsR2 by electrotansformation and transformants were selected based on Tet and Cm resistance.
**Heterologous protein production.** For the heterologous expression of AtsR2, 25 ml of M17 broth supplemented with 0.5 % (wt/vol) glucose was inoculated with a 2 % inoculum of an overnight culture grown for 16 h of *L. lactis* NZ9000 harbouring either pNZ-atsR2 or the empty vector pNZ8048 (used as a negative control), followed by incubation at 30°C until an OD<sub>600nm</sub> of ~0.5 was reached, at which point protein expression was induced by addition of cell-free supernatant of a nisin-producing strain (65), followed by continued incubation for a further 2 h. Cells were harvested by centrifugation, resuspended in 10 mM Tris-HCl (pH 8.0), and disrupted with glass beads in a mini-bead beater (BioSpec Products, Bartlesville, OK). Cellular debris was removed by centrifugation to produce an AtsR2-containing crude cell extract.

**Electrophoretic mobility shift assays (EMSA).** DNA fragments representing different portions of each of the promoter regions upstream of the *atsR2* and *atsT* genes were prepared by PCR using IRD-labelled primer pairs synthesized by Integrated DNA Technologies (Coralville, IA) (Table 2). EMSAs were essentially performed as described previously (66). In all cases, the binding reactions were performed in a final reaction volume of 20 μl in the presence of poly (dI-dC) in binding buffer (20 mM Tris-HCl, 5 mM MgCl₂, 0.5 mM dithiothreitol [DTT], 1 mM EDTA, 50 mM KCl, 10 % glycerol at pH 7.0). Various amounts of *L. lactis* NZ9000 crude cell extract containing pNZ-atsR2 or pNZ8048 were mixed on ice with a fixed amount of DNA probe (0.1 pmol) and subsequently incubated for 30 min at 37°C. Samples were loaded on a 6 % non-denaturing polyacrylamide (PAA) gel prepared in TAE buffer (40 mM 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 120 min in an Atto Mini PAGE system (Atto Bioscience and Biotechnology, Tokyo, Japan). Signals were detected using an Odyssey Infrared Imaging 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.
Based on the comparative genome analysis presented in Figure 2, we found that the putative sulfatase clusters are well conserved among the *B. breve* strains whose genomes were recently published (71). Of the currently available complete *B. breve* genomes, *B. breve* NCFB2258, *B. breve* 689B, *B. breve* 12L and *B. breve* S27 encode clear homologues of both identified putative sulfatase gene clusters described above. In contrast, the genomes of *B. breve* JCM7017, *B. breve* JCM7019 and *B. breve* ACS-071-V-Sch8b contain just a single, but variable putative sulfatase cluster (Fig. 2). A clear homologue of the *ats1* gene cluster was also identified in the recently published genome of *B. longum* subsp. *infantis* BT1 (Accession number CP010411). No other homologues of either sulfatase-encoding gene clusters were identified by BLASTP analysis within the available bifidobacterial genome sequences.

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

**Genome response of *B. breve* UCC2003 to growth on GlcNAc-6-S.** In order to investigate which genes are responsible for GlcNAc-6-S metabolism in *B. breve* UCC2003, global gene
expression was determined by microarray analysis during growth of the strain in mMRS supplemented with GlcNAc-6-S and compared with gene expression when grown in mMRS supplemented with ribose. Ribose was considered an appropriate carbohydrate for comparative transcriptome analysis because the genes involved in ribose metabolism are known, while it has furthermore successfully been used in a number of transcriptome studies in this strain (17, 18, 72-74). Of the two predicted sulfatase and anSME-encoding gene clusters of \textit{B. breve} UCC2003 (see above), transcription of the \textit{ats2} gene cluster was significantly up-regulated (fold change $>3.0$, $P$-value $<0.001$) during growth on GlcNAc-6-S, while no (significant) difference in the level of transcription was observed for the \textit{ats1} gene cluster (Table 3). Interestingly, three other gene clusters were also significantly up-regulated (corresponding to locus tags Bbr$_{0846}$ through to Bbr$_{0849}$, Bbr$_{1585}$ through to Bbr$_{1590}$, and Bbr$_{1247}$ through to Bbr$_{1249}$; see Fig. 4 and Table 3).

Within the Bbr$_{0846}$-0849 gene cluster, which is separated from the \textit{ats2} cluster by a single gene (Fig. 3), Bbr$_{0846}$ (\textit{nagA1}) and Bbr$_{0847}$ (\textit{nagB2}) are predicted to encode an N-acetylglucosamine-6-phosphate deacetylase and a glucosamine-6-phosphate deaminase, respectively. Bbr$_{0848}$ (designated here as \textit{nagK}) encodes a predicted ROK-family kinase, which contains the characteristic DxGxT motif at its N-terminal end (75). The \textit{B. breve} UCC2003-encoded NagK protein exhibits 42 \% similarity at protein level with the previously characterized \textit{E. coli} K-12-encoded, ROK-family NagK protein, which phosphorylates GlcNAc to produce \textit{N}-acetylg glucosamine-6-phosphate (GlcNAc-6-P) (76). Therefore this cluster is predicted to encode enzymes for the complete GlcNAc catabolic pathway as previously described in \textit{E. coli}, whereby GlcNAc is first phosphorylated by NagK, producing GlcNAc-6-P, followed by NagA-mediated deacetylation to produce glucosamine-6-phosphate, and the NagB-mediated deamination and isomerisation to produce fructose-6-
phosphate (76, 77). Bbr_0849 encodes a predicted transcriptional regulator from the ROK family (designated here as atsR2).

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

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

**Disruption of the atsT and atsA2 genes.** In order to investigate if disruption of individual genes from the *ats2* gene cluster would affect the ability of *B. breve* UCC2003 to utilize GlcNAc-6-S, insertion mutants were constructed in the *atsT* and *atsA2* genes, resulting in strains *B. breve* UCC2003-atsT and *B. breve* UCC2003-atsA2, respectively (see Materials and Methods). The insertion mutants were analyzed for their ability to grow in mMRS
supplemented with GlcNAc-6-S as compared to *B. breve* UCC2003. As expected, and in contrast to the wild type, there was a complete lack of growth of *B. breve* UCC2003-atsT and *B. breve* UCC2003-atsA2 in media containing GlcNAc-6-S as the sole carbon source (Fig. 3B), thus demonstrating the involvement of the disrupted genes in GlcNAc-6-S metabolism. Growth of the insertion mutants was not impaired on lactose, where all strains reached final OD_{600nm} levels comparable to that reached by the wild type strain (Fig. 3B).

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

Electrophoretic mobility shift assays. In order to determine if the AtsR2 protein directly interacts with promoter regions of the ats2 gene cluster, crude cell extracts of *L. lactis* 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 \textit{atsR2} and \textit{atsT} (Fig. 1). In order to determine the transcriptional start site of these presumed promoters, primer extension analysis was performed using RNA extracted from \textit{B. breve} UCC2003-\textit{atsR2} grown in mMRS supplemented with 0.5 % ribose. Microarray analysis had shown that the expression levels of \textit{atsT} were high when the \textit{B. breve} UCC2003-\textit{atsR2} strain was grown on ribose (Table 4). For this reason, the mutant strain was considered most suitable for primer extension analysis. For the \textit{atsR2} promoter region, initial attempts to attain a primer extension product from mRNA isolated from \textit{B. breve} UCC2003-\textit{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 \textit{B. breve} UCC2003-\textit{atsR2}, generating strain \textit{B. breve} UCC2003-\textit{atsR2}-pBC1.2-\textit{atsRProm}. A primer extension product was obtained for the \textit{atsT} promoter region using mRNA isolated from \textit{B. breve} UCC2003-\textit{atsR2}, therefore it was not necessary to clone this promoter. Single extension products were identified upstream of \textit{atsR2} and \textit{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...
atsB2 genes are adjacent, as is also the case for their homologous genes in K. pneumoniae and Prevotella strain RS2 (83, 84). We obtained compelling evidence that the ats2 cluster is co-regulated with the Bbr_0846-0849 cluster by the ROK-family transcriptional repressor AtsR2. The only previously characterised bifidobacterial ROK-family transcriptional regulator is RafA, the transcriptional activator of the raffinose utilisation cluster in B. breve UCC2003 (73). The Bbr_0846-0848 genes are presumed to be involved in the metabolism of GlcNAc following the removal of the sulfate residue from GlcNAc-6-S. The fructose-6-phosphate produced from GlcNAc by the combined activities of NagK, NagA and NagB is expected to enter the fructose-6-phosphate phosphoketolase pathway or bifid shunt, the central metabolic pathway of bifidobacteria (85). It is interesting that B. breve UCC2003 is capable of growth on GlcNAc-6-S as a sole carbon source, but apparently not on GlcNAc (16). Since the B. breve UCC2003 genome seems to encode the enzymes required to metabolise GlcNAc, it suggests that the atsT transporter has (high) affinity for only the sulfated form of this N-acetylated carbohydrate.

A novel method of desulfating mucin which does not require a sulfatase enzyme has been characterised from Prevotella strain RS2, whereby a sulfoglycosidase removes GlcNAc-6-S from purified porcine gastric mucin (86). The presence of a signal sequence on this glycosulfatase (86), thus indicating extracellular activity, is interesting in relation to the current study, as it presents a source of GlcNAc-6-S to B. breve strains, suggestive of a cross-feeding opportunity for members of this species. This is particularly noteworthy when it is considered that the sulfatase enzymes produced by B. breve UCC2003 are intracellular, implying that B. breve UCC2003 is reliant on the extracellular glycosyl hydrolase activity of other members of the gut microbiota in order to gain access to mucin-derived sulfated monosaccharides. Recent studies have shown that B. breve UCC2003 employs a cross-feeding strategy to great effect, as it can utilize components of 3′ sialyllactose (a HMO) and...
mucin following the degradation of these sugars by *B. bifidum* PRL2010, whereas in the absence of *B. bifidum* PRL2010, it is not capable of utilising either of these sugars as a sole carbon source (72, 80). A recent study has further provided transcriptomic evidence for carbohydrate cross-feeding between bifidobacterial species. Four bifidobacterial strains, namely *B. bifidum* PRL2010, *B. breve* 12L, *B. adolescentis* 22L and *B. longum* subsp. *infantis* ATCC25697, were cultivated either in pairs (bi-association) or a combination of all four strains (multi-association), under *in vivo* conditions in a murine model. In all strains, transcription of predicted glycosyl hydrolase-encoding genes, particularly those involved in xylose or starch utilization, were affected by co- or multi-association. In relation to xylose metabolism, the authors speculated that in co- or multi-association, the combined glycosyl hydrolase activities of the strains may allow them to degrade xylose-containing polysaccharides which would otherwise be inaccessible (87).

In *B. thetaiotaomicron*, the *in vivo* contribution of sulfatase activity towards bacterial fitness has been well-established. In previous studies of chondroitin sulfate and heparan sulfate metabolism by this species, mutagenesis of a gene designated *chuR*, which was first predicted to encode a regulatory protein but then later identified as an anSME, resulted in the inability to compete with wild type *B. thetaiotaomicron* in germ-free mice (37, 88). In a recent study, 28 predicted sulfatase-encoding genes were identified on the genome of *Ba. thetaiotaomicron*, 20 of which are predicted extracellular enzymes, yet the previously described *chuR* gene is the sole anSME-encoding gene (36, 89, 90). Recently, this anSME was shown to be of significant importance in this strain’s ability to colonize the gut, as an isogenic derivative of this strain (designated ∆anSME) carrying a deletion in the anSME-encoding gene displayed reduced fitness *in vivo* (36). The authors have speculated that anSME activity and associated sulfatase activities are important as the bacterium adapts to
the gut environment (36). Given that sulfatase activity within the *Bifidobacterium* genus is (at least based on currently available genome sequences) limited to the *B. breve* species and a single member of the *B. longum* subsp. *infantis* subspecies, it is interesting to speculate on the effect this activity may have on bacterial fitness in the large intestine. It is intriguing to note that human intestinal mucins increase in acidity along the intestinal tract, with more than half of mucin oligosaccharide structures in the distal colon containing either sialic and/or sulfate residues (91). We recently showed that 11 of 14 strains of *B. breve* tested were capable of growth on sialic acid, while sialic acid utilization genes can also be found on the genomes of *B. longum* subsp. *infantis* strains (20, 22, 72). The ability of *B. breve* strains and possibly certain *B. longum* subsp. *infantis* strains, to utilise both sialic acid and sulfated GlcNAc-6-S may provide them with a competitive advantage over other members of the *Bifidobacterium* genus and other members of the gut microbiota, thus contributing to their successful colonization ability in this highly competitive environment.

**Funding Information**

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

**References:**


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


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

45. Carlson BL, Ballister ER, Skordalakes E, King DS, Breidenbach MA, Gilmore SA,
Berger JM, Bertozzi CR. 2008. Function and structure of a prokaryotic formylglycine-

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

Further characterization of Cys-type and Ser-type anaerobic sulfatase maturating enzymes
suggests a commonality in mechanism of catalysis. Biochemistry 52:2874-2887.


50. Terzaghi BE, Sandine WE. 1975. Improved medium for lactic streptococci and their


Figure 1: (A) Synthesis of 6-O- and 3-O-sulfate-2-acetamido-2-deoxy-D-glucose 1 and 2 (i): BnBr, NaH, LiBr, DMF; (ii): Ac₂O, Py; (iii): NaOMe, MeOH; (iv): PhCH(OMe)₂, HCOOH; (v): SO₃·NEt₃, Py, 85°C; (vi): 10% Pd/C, EtOH, 15 bar H₂; (vii): (1) TrCl, CaSO₄, Py, 100 °C, (2) Ac₂O; (viii): AcOH, HBr; (ix): SO₃·NEt₃, DMF, 55 °C; (x): NaOMe, MeOH. (B) Synthesis of 3-O- and 6-O-sulfate-2-acetamido-2-deoxy-D-galactose 3 and 4. Key (i): Ac₂O, Py; (ii): BnOH, BF₃·OEt₂, CH₂Cl₂, 3 A MS; (iii): NaOMe, MeOH; (iv): PhCH(OMe)₂, HCOOH; (v): SO₃·NEt₃, DMF, 55 °C; (vi): 10% Pd/C, EtOH, 15 bar H₂; (vii): NaOMe, MeOH; (viii): Me₂C(OMe)₂, p-TSA, DMF, 65°C; (x): SO₃·NEt₃, DMF, 55°C; (x): CF₃COOH, H₂O; (xi): 10% Pd/C, EtOH, 10 bar H₂.

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

Figure 3: (A) Final OD₆₀₀nm values obtained following 24 h growth of B. breve UCC2003 on mMRS without supplementation with a carbon source (negative control) or containing 0.5 % (wt/vol) lactose, GlcNAc-6-S, GlcNAc-3-S, GalNAc-6-S or GalNAc-3-S as the sole carbon source. (B) Final OD₆₀₀nm values obtained following 24 h growth of B. breve UCC2003, B. breve UCC2003-atsT and B. breve UCC2003-atsA2 in modified MRS without
supplementation with a carbon source (negative control, horizontally striped bars) or containing 0.5 % (wt/vol) lactose (diagonally striped bars) or GlcNAc-6-S (solid grey filled bars) as the sole carbon source. The results are the mean values obtained from two separate experiments. Error bars represent the standard deviation.

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

Figure 5: (A) Schematic representation of the ats2 gene cluster of B. breve UCC2003 and DNA fragments used in EMSAs for the atsR2 and atsT promoter regions, together with Weblogo representation of the predicted operator of AtsR2. Plus or minus signs indicate ability or inability of AtsR2 to bind to the DNA fragment. The bent arrows represent the position and direction of the proven promoter sequences (see Fig. 6). (B) EMSAs showing the interactions of (I) crude cell extract containing pNZ-AtsR2 with the DNA fragments R1, R2, R3, T1, T2 and T3 and (II) crude cell extract containing pNZ8048 (empty vector) with the DNA fragments R1 and T1. The minus symbol indicates reactions to which no crude cell extract was added, while the remaining lanes represent binding reactions with the respective DNA probes incubated with increasing amounts of crude cell extract. Each successive lane from right to left represents a doubling of the amount of crude cell extract. (C) EMSAs showing AtsR2 interaction with the DNA fragments R1 and T1 with the addition of GlcNAc or GlcNAc-6-S in concentrations ranging from 2.5 mM to 20 mM.
Figure 6: Schematic representation of the atsR2 (panel A), atsT (panel B), promoter regions. Boldface type and underlining indicate -10 and -35 hexamers (as deduced from the primer extension results) and ribosomal binding site (RBS); the transcriptional start site is indicated by an asterisk. The arrows underneath the indicated DNA sequences indicate the inverted repeats that represent the presumed AtsR2 binding site. The arrows in the right panels indicate the primer extension products.
Table 1: Bacterial strains and plasmids used in this study.

<table>
<thead>
<tr>
<th>Strains and plasmids</th>
<th>Relevant features</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Escherichia coli</em> strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli EC101</td>
<td>Cloning host; repC+ low</td>
<td>(92)</td>
</tr>
<tr>
<td>E. coli EC101-pNZ-M.Bbrl+Bbrl</td>
<td>EC101 harboring a pNZ8048 derivative containing bbrllM and bbrlllM</td>
<td>(57)</td>
</tr>
<tr>
<td>E. coli XL1-blue</td>
<td>(supEl4 hsdR17 recA1 gvy46 thi relA1 lac F[proAB lacZΔM15 Tn10(TetR)]) Stratagene</td>
<td></td>
</tr>
<tr>
<td>E. coli XL1-blue-pBC1.2-atsProm</td>
<td>XL1-blue harboring pBC1.2-atsProm</td>
<td></td>
</tr>
<tr>
<td>L. lactis strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. lactis NZ9000</td>
<td>MG1363, pepN::nisRK, nisin inducible overexpression host</td>
<td>(65)</td>
</tr>
<tr>
<td>L. lactis NZ9000-pNZ8048</td>
<td>NZ9000 containing pNZ8048</td>
<td>This study</td>
</tr>
<tr>
<td>L. lactis NZ9000-pNZ-atsR2</td>
<td>NZ9000 containing pNZ-atsR2</td>
<td>This study</td>
</tr>
<tr>
<td>L. lactis NZ97000</td>
<td>Nisin-A producing strain</td>
<td>(65)</td>
</tr>
<tr>
<td>Bifidobacterium sp. strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. breve UCC2003</td>
<td>Isolate from a nursling stool</td>
<td>(58)</td>
</tr>
<tr>
<td>B. breve UCC2003-atsR2</td>
<td>pORI19-tetW-atsR2 insertion mutant of B. breve UCC2003</td>
<td>This study</td>
</tr>
<tr>
<td>B. breve UCC2003-atsT</td>
<td>pORI19-tetW-atsT insertion mutant of B. breve UCC2003</td>
<td>This study</td>
</tr>
<tr>
<td>B. breve UCC2003-atsA2</td>
<td>pORI19-tetW-atsA2 insertion mutant of B. breve UCC2003</td>
<td>This study</td>
</tr>
<tr>
<td>B. breve UCC2003-atsR2-pBC1.2-atsProm</td>
<td>pORI19-tetW-atsR2 insertion mutant of UCC2003 containing pBC1.2-atsProm</td>
<td>This study</td>
</tr>
<tr>
<td>Plasmids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAM5</td>
<td>pBC1-pUC19-Te2</td>
<td>(64)</td>
</tr>
<tr>
<td>pNZ8048</td>
<td>Cm', nisin-inducible translational fusion vector</td>
<td>(65)</td>
</tr>
<tr>
<td>pNZ-atsR2</td>
<td>Cm', pNZ8048 derivative containing translational fusion of atsR2 encoding DNA</td>
<td>This study</td>
</tr>
<tr>
<td>pORI19</td>
<td>Em', repA, ori', cloning vector</td>
<td>(92)</td>
</tr>
<tr>
<td>pORI19-tetW-atsR2</td>
<td>Internal 408 bp fragment of atsR2 and tetW cloned in pORI19</td>
<td>This study</td>
</tr>
<tr>
<td>pORI19-tetW-atsT</td>
<td>Internal 416 bp fragment of atsT and tetW cloned in pORI19</td>
<td>This study</td>
</tr>
<tr>
<td>pORI19-tetW-atsA2</td>
<td>Internal 402 bp fragment of atsA2 and tetW cloned in pORI19</td>
<td>This study</td>
</tr>
<tr>
<td>pBC1.2</td>
<td>pBC1-pSC101-Cmr</td>
<td>(64)</td>
</tr>
<tr>
<td>pBC1.2-atsProm</td>
<td>AtsR2 promoter region cloned in pBC1.2</td>
<td>This study</td>
</tr>
</tbody>
</table>
Table 2: Oligonucleotide primers used in this study.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloning of 408 bp fragment of atsR2 in pORI19</td>
<td>AtsR2F</td>
<td>GACTAGATCCATTAACAGTGAAAGC</td>
</tr>
<tr>
<td>Cloning of 416 bp fragment of atsT in pORI19</td>
<td>AtsT</td>
<td>GACTAGATCCATTAACAGTGAAAGC</td>
</tr>
<tr>
<td>Cloning of 402 bp fragment of atsA2 in pORI19</td>
<td>AtsA2F</td>
<td>GACTAGATCCATTAACAGTGAAAGC</td>
</tr>
<tr>
<td>Amplification of tetW</td>
<td>TetWF</td>
<td>TGCACTGAATTCGCGTACGTCGATG</td>
</tr>
<tr>
<td>Confirmation of site-specific homologous recombination</td>
<td>AtsR2confirm</td>
<td>GACGGCGACGAGAAC</td>
</tr>
<tr>
<td>Cloning of atsR2 in pNZ8048</td>
<td>AtsR2FOR</td>
<td>GACTAGATCCATTAACAGTGAAAGC</td>
</tr>
<tr>
<td>Cloning of atsR2 promoter in pBC1.2</td>
<td>AtsR2PromF</td>
<td>GACTAGATCCATTAACAGTGAAAGC</td>
</tr>
<tr>
<td>IRD-labelled primers</td>
<td>AtsR2R1F</td>
<td>GACTAGATCCATTAACAGTGAAAGC</td>
</tr>
<tr>
<td></td>
<td>AtsR2R1R</td>
<td>GACGGCGACGAGAAC</td>
</tr>
<tr>
<td></td>
<td>AtsR2R2F</td>
<td>GACTAGATCCATTAACAGTGAAAGC</td>
</tr>
<tr>
<td></td>
<td>AtsR2R2R</td>
<td>GACGGCGACGAGAAC</td>
</tr>
<tr>
<td></td>
<td>AtsR2T1F</td>
<td>GACTAGATCCATTAACAGTGAAAGC</td>
</tr>
<tr>
<td></td>
<td>AtsR2T1R</td>
<td>GACGGCGACGAGAAC</td>
</tr>
<tr>
<td></td>
<td>AtsR2T2F</td>
<td>GACGGCGACGAGAAC</td>
</tr>
<tr>
<td></td>
<td>AtsR2T2R</td>
<td>GACGGCGACGAGAAC</td>
</tr>
<tr>
<td></td>
<td>AtsR2T3F</td>
<td>GACGGCGACGAGAAC</td>
</tr>
<tr>
<td></td>
<td>AtsR2T3R</td>
<td>GACGGCGACGAGAAC</td>
</tr>
</tbody>
</table>

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

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

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.

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

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

\[
\begin{align*}
&\text{ATTGTATCCATCCCGACATATTTAGGCAACCATCTTGA}^{\text{CTAGTAATTAATT}}^{	ext{-10}} \\
&CACC\text{TGTGCTATATT}{^*}\text{TATCAATTGAAAGGCTATCAACGAGCGGATTAGCC} \\
&GCG\text{CACAAGAAGCTCGAAGAAAAAGGAGTTTCACGCACGATGGCCCGCCCGCA} \\
&AAG\text{ATGACACAATAAGGATTTGGCATCCTCGAAGAGCCGCTTCTTCAAGC} \\
&RBS \\
&GCT\text{GACGGGAGCTTTTCTAG}
\end{align*}
\]

B

\[
\begin{align*}
&\text{TTCTTCACCGTGAGGACCATTGATATTTCAGTAATCTTTGAGTCTTTTGCG} \\
&\text{TGC\text{CTTATTCTCATAATGTCAACAAGGTTGACGAATGATGTATACT}}^{	ext{-35}} \\
&\text{GAATCAGCGCCAGCTATCAGCGCGACATTCTAGTGAACTAGGATTCAAAGG} \\
&RBS \\
&\text{AGAAAAGGATACTTTTCATG}
\end{align*}
\]