Supporting Information

Motherway et al. 10.1073/pnas.1105380108

SI Materials and Methods

Sequence Assembly and Quality Control. The *Bifidobacterium breve* UCC2003 genome sequence was determined using a whole-genome shotgun approach followed by Roche 454 sequencing. Two shotgun libraries were constructed: a small-insert library (2–4 kb) using pGEM-T easy vector (Promega) and a large cosmid library (40-45 kb) using pWEB vector (Epicentre Technologies). Sequence sampling from these banks generated just over 7 million bp of useable sequence data, representing about threefold coverage of the *B. breve* UCC2003 genome (performed by MWG-Biotech). For Roche 454 sequencing, genomic DNA was extracted as described previously (1), after which a 3-kb fragment library was prepared and sequenced using Roche standard procedures (performed by Beckman Coulter). Sequence reads were assembled into 376 contigs using Phred (2, 3), Phrap (P. Green, University of Washington, Seattle, WA; http://www.phrap.org), RepeatMasker (A.F.A. Smit, R. Hubley, and P. Green; http://www.repeatmasker.org) and the Staden package (4) software. Gaps were closed using additional sequencing from small-insert clones (2,000 sequencing reads) and direct and inverse PCR (850 sequencing reads), resulting in a single contig (2.42 Mb). Quality improvement of the genome sequence involved sequencing of more than 400 PCR products (3,600 sequencing reads) across the entire genome to ensure correct assembly, double stranding, and the resolution of any remaining base conflicts. This process resulted in a total of 374,449 sequences with an average read length of 381 bp covering ~134 Mb, roughly constituting a 43-fold genome coverage. Roche 454 reads were mapped onto the *B. breve* UCC2003 genome and used to correct base-calling errors manually.

Sequence Annotation. Protein-coding sequences were predicted using a combination of the methods Glimmer (5, 6), FrameD (7), and Prodigal (8), as well as comparative analysis involving Orpheus (9) and BLASTX (10). Results from the gene-finder programs were combined, and a preliminary identification of ORFs was made on the basis of BLASTP (11) analysis against a nonredundant protein database provided by the National Center for Biotechnology Information (12). Artemis (13) was used to inspect the results of the combined gene finders and associated BLASTP results. BlastxBrowser was used to examine the BLASTX results (14). A manual inspection was performed to verify or, if necessary, redefine the start of each predicted coding region. Annotation made use of the GC frame plot feature of Artemis, ribosome-binding site information obtained from RBSfinder (15), alignments with similar protein-coding sequences from other organisms, and guanine-cytosine (G+C) content analysis.

Assignment of protein function to predicted coding regions of the *B. breve* UCC2003 genome was performed manually using the ERGO annotation tool (16). Primary functional classification of the *B. breve* gene products was performed according to the Riley rules with some modifications to reflect the *B. breve* gene content (17, 18). HMMTOP (19) and TMHMM (20) were used to predict transmembrane sequences, and SignalP (21) was used for the prediction of signal peptides. Ribosomal RNA genes were detected on the basis of BLASTN searches and were annotated manually. Transfer RNA genes were identified using tRNAscan-SE (22). Miscellaneous-coding RNAs were identified using the Rfam database (23) using the INFERNAL software package (24). Insertion sequence elements were identified using Repeatfinder (25), Reputer (26), and BLAST (11) and were annotated manually. IS families were assigned using ISFinder (http://www-is.biotoul.fr/is.html). Carbohydrate-active enzymes were identified based on similarity to the carbohydrate-active enzyme (CAZy) database entries (27), and transporter classification was performed according to the TC-DB scheme (28).

Comparative Analysis. Whole-genome comparisons of *B. breve* UCC2003 were performed with BLASTN alignments against *Bifidobacterium longum* subsp. *longum* NCC2705 (AE014295), *B. longum* subsp. *longum* DJO10A (CP000605), *B. longum* subsp. *infantis* ATCC15697 (CP001095), *Bifidobacterium adolescentis* ATCC15703 (NC_008618), *Bifidobacterium dentium* Bd1 (NC_013714), *Bifidobacterium bifidum* PRL2010 (CP001840), and *Bifidobacterium animalis* subsp. *lactis* DSM10140 (NC_012815). Genome synteny comparisons were performed using the set of genomes as mentioned above. Whole genomes were compared at the nucleotide level using Dotter (29) and MUMmer programs (30) with default values.

Bacterial strains, plasmids, and culture conditions. Bacterial strains and plasmids used in this study are listed in Table S3. *B. breve* UCC2003 was cultured routinely in reinforced clostridial medium (Oxoid Ltd.). Bifidobacterial cultures were incubated at 37 °C under anaerobic conditions which were maintained using an anaerobic hood (Davidson and Hardy). *Escherichia coli* were cultured in LB broth (31) at 37 °C with agitation. Where appropriate, growth medium contained tetracycline (Tet) (15 μg/mL), erythromycin (100 μg/mL for *E. coli*), or kanamycin (50 μg/mL for *E. coli*). Recombinant *E. coli* cells containing pORI19 were selected on LB agar containing erythromycin and supplemented with X-Gal (40 μg/mL) and 1 mM isopropyl-β-D-galactopyranoside.

DNA manipulations. Chromosomal DNA was isolated from *B. breve* UCC2003 as previously described (4). Minipreparation of plasmid DNA from *E. coli* was achieved using the Qiaprep spin plasmid miniprep kit (Qiagen GmBH) as described previously (32). Single-stranded oligonucleotide primers used in this study were synthesized by MWG Biotech AG. Standard PCRs were performed using TaqPCR Master Mix (Qiagen). High-fidelity PCR was achieved using KOD polymerase (Novagen). *B. breve* colony PCRs were performed as described previously (32). PCR fragments were purified using the Qiagen PCR purification kit. Electroporation of plasmid DNA into *E. coli* was performed as described by Sambrook et al. (31) and into *Lactococcus lactis* as described by Wells et al. (33). Electrotransformation of *B. breve* UCC2003 was performed as described by Mazé et al. (34).

Construction of B. breve UCC2003 tadA insertion mutant. An internal fragment of *tad*A was amplified by PCR using *B. breve* UCC2003 chromosomal DNA as a template and primer pair IMtadAF and IMtadAR (Table S4). Insertion mutation of the *tadA* gene was generated essentially as described previously (35) to produce *B. breve* UCC2003-tadA. The expected site-specific recombination in potential Tet-resistant mutants was confirmed by colony PCR using primer combinations tetWFw and tetWRv to verify *tetW* gene integration and primers tadA-Fw (upstream of the *tadA* gene fragments) with pORI19For to confirm integration at the correct chromosomal position. The stability of the pORI19-tadA insertion was tested in the absence of antibiotic over 100 generations and was found to be >99% stable.

Colonization of Balb/c Mice. Seven-week-old male Balb/c mice were housed individually in vented cages (Animal Care Systems) under a strict 12-h light cycle. Mice (n = 5 per group) were fed a standard polysaccharide-rich mouse chow diet and water ad libitum. Mice were inoculated by oral gavage (10^9 cfu of *B. breve* UCC2003PK1 or *B. breve* UCC2003-tadAPK1 in 100 μL of PBS).

Fecal pellets were collected at intervals over 25 d to enumerate bacteria. Mice were killed 25 d after inoculation, and their intestinal tracts were dissected promptly. The small intestine, cecum, and large intestine were harvested for determination of cfu [serial dilution plating on reinforced clostridial agar (RCA) plates with appropriate antibiotics]. Ceca were placed in RNAlater (Ambion) for immediate RNA isolation (see below).

Colonization of Germ-Free Mice. Eight-week-old female germ-free Swiss Webster mice were housed in flexible film gnotobiotic isolators under a strict 12-h light cycle. Mice were fed an autoclaved standard polysaccharide-rich mouse chow diet. Mice (n = 10 per group) were inoculated by with 1×10^9 cfu of *B. breve* UCC2003PK1 or *B. breve* UCC2003-tadAPK1 in 20 μL of PBS by oral pipetting whereby the inoculums are delivered by positioning a micropipette tip immediately behind the incisors. Five mice were maintained as uninoculated controls to monitor the germ-free status of the facility. Fecal pellets were collected twice weekly to determine the number of each strain present. Twenty-eight days after the first inoculation, each group of the inoculated mice was divided randomly into two groups. Five mice from each group were retained in the germ-free facility, and the other five mice from each group were housed individually in vented cages in the non–germ-free animal unit. Fecal pellets from non–germ-free mice were put in the cages to promote establishment with a normal murine gastrointestinal microbiota. Fecal pellets were collected from all groups twice weekly. The mice were killed at day 49, and their intestinal tracts were dissected promptly. Aliquots of the small intestine, cecum, and large intestine were retained in PBS for cfu determination (serial dilution plating on RCA agar supplemented with the appropriate antibiotics). Cecal samples were prepared for immunogold electron microscopy.

Generation of Bbr_0136 (Flp_{2003})-Specific Antibodies. An antibody against a 20-amino acid Bbr_0136-specific peptide (KSDAIKTLLTNLIKKALNVG) was raised in rabbits by Davids Biotechnologie GmbH. An initial immunization with the Bbr_0136 peptide conjugated to a carrier and Freund's complete adjuvant was followed by five subsequent boost injections. A final serum sample was acquired 11 wk after the initial immunization, and the anti-Bbr_0136 antibody was affinity purified.

Immunogold Electron Microscopy. Cecal samples were washed once in 0.1 M phosphate buffer and then were fixed for 4 h at room temperature in 4% paraformaldehyde and 0.1% glutaraldehyde in 0.1 M phosphate buffer. Subsequently, the fixed cecal chyme was collected by centrifugation, suspended into 2% paraformaldehyde in 0.1 M phosphate buffer, and stored at 4 °C. The cecal sample was centrifuged for 1 min at $1,000 \times g$ to pellet the chyme debris, and the resulting supernatant was centrifuged further for 3 min at $16,000 \times g$ to collect bacteria. The cells then were washed once with PBS before they were applied to the Formvar carbon-coated copper grid. The grid was floated on the bacterial suspension for 1 h, washed three times with 0.02 M glycine in PBS, and blocked for 15 min with 1% BSA in PBS. The grid was incubated for 1 h on a droplet of 1:5 diluted anti-Bbr_0136 or anti-Bbr_0136 preimmune serum in blocking solution, washed several times with 0.1% BSA in PBS, and incubated for 20 min with protein A conjugated to 10-nm gold particles diluted 1:55 in blocking solution. The grid was washed four times with PBS and fixed for 5 min with 1% glutaraldehyde, followed by several washes with distilled water before negative staining with a mixture of 1.8% methylcellulose and 0.4% uranyl acetate. The grid was examined, and micrographs were visualized using a JEOL 1200 EX II transmission electron microscope.

Microarray Procedures. DNA microarrays containing oligonucleotide primers representing each of the 1,926 annotated genes of *B. breve* UCC2003 were obtained from Agilent Technologies. An overnight culture of *B. breve* was used to inoculate (1% inoculum) 50 mL of de Man, Rogosa, Sharpe (MRS) broth. Cells were incubated at 37 °C until an OD_{600} of 0.5 was reached. Cells then were harvested by centrifugation at $8,000 \times g$ for 1 min at room temperature and were frozen immediately at −80 °C. Cell disruption, RNA isolation, DNA isolation, and RNA quality control were performed as described previously (36, 4). Genomic DNA (gDNA) labeling for comparative genome hybridizations (CGH) was undertaken according to the standard protocols of the Bacterial Microarray Group at St George's Hospital Medical School (37). In brief 5 μg of bifidobacterial gDNA was labeled with dCTP Cy3 (UCC2003) or dCTP Cy5 (test strain) using random primers (Promega) and a DNA polymerase I large Klenow fragment (Invitrogen). Labeled gDNA was hybridized using the Agilent CGH kit (5188–5220) as described in the Agilent manual, *Oligonucleotide Array-Based CGH for Genomic DNA Analysis* (v6.0) (publication no. G4410-90010). Ceca isolated from mice were placed in RNAlater (Ambion) and were processed immediately. Bacterial mRNA was extracted from cecal RNA preparations using the MicrobEnrich and MicrobExpress kits (Ambion) according to the manufacturer's instructions. cDNA from bacterial mRNA was synthesized using the cDNA synthesis and labeling kit DSK-001 (Kreatech) according to the manufacturer's instructions. One microliter of each cDNA solution (10 ng) was amplified in triplicate using the GenomiPhi V2 DNA Amplification Kit (Amersham Biosciences) according to the manufacturer's protocol and was labeled with Cy3 or Cy5 using Cy3-ULS and Cy5-ULS from the cDNA synthesis and labeling kit DSK-001 (Kreatech). Labeled and amplified cDNA was hybridized using the Agilent Gene Expression hybridization kit (5188–5242) as described in the Agilent manual, *Two-Color Microarray-Based Gene Expression Analysis* (v4.0) (publication no. G4140-90050). Following hybridization, microarrays were washed as described in the manuals and were scanned using Agilent's G2565A DNA microarray scanner. The scanning results were converted to data files with Agilent's Feature Extraction software (version 9.5). DNA microarray data were processed as described previously (38). Differential expression tests were performed using a t test. A gene was considered differentially expressed between a test strain and control when an expression ratio of >5 or <0.2 relative to the result for the control was obtained with a corresponding P value <0.0001. Validation of the CGH was performed by comparing the hybridization efficiency of DNA of the *B. breve* type strain DSM 20213 and the sequence identity of this strain with the probes on the microarray. A clear correlation was observed between sequence identity and hybridization efficiency, suggesting that a natural logarithm (ln) twofold change in signal intensity between target and control is sufficient to determine whether a gene is present or absent on a given genome. Therefore, a gene was considered absent when the ratio between a test strain and UCC2003 was <ln 2 with a corresponding P value <0.0001.

Differential expression of genes was confirmed by real-time quantitative RT-PCR (qRT-PCR). De novo cDNAs were prepared as described above. PCR primers and probes were designed using the Universal ProbeLibrary Assay Design Center (https://www.roche-applied-science.com/sis/rtpcr/upl/adc.jsp). Primer sequences are provided in Table S5. The *rnpA* gene (encoding ribonuclease P protein component) was used as a housekeeping gene with an assumed constant level of transcription to correct for variability in the initial amount of total RNA. Amplification reactions contained 1 μL of 6.7-fold diluted cDNA, 10 μL of the 2× FastStart TaqMan Probe Master (Roche), 900 nM of each primer, and 250 nM of probe mix and were brought to a total volume of 20 μL by the addition of RNase-free water. All qRT-

PCR reactions were performed in triplicate by means of a LightCycler 480 System (Roche) using 384-well plates. Thermal cycling conditions were as recommended by the manufacturer (Roche). The $2^{-\Delta\Delta Ct}$ method (39) was used to calculate relative changes in gene expression determined from real-time qPCR experiments. Pooled cDNA from all groups was used as a calibrator control for analysis of differential gene expression. Results were calculated from at least two independent RNA extractions. After $2^{-\Delta\Delta Ct}$ analysis, the qRT-PCR expression data, were subjected to a Mann–Whitney *t* test to compare all groups using GraphPad Prism 4 software (GraphPad Software). Data are shown as mean ± SEM. *P* values <0.05 were considered significant. The statistical analysis was performed blind to the origin of the data.

1. O'Riordan K (1998) *Studies on Antimicrobial Activity and Genetic Diversity of Bifidobacterium Species: Molecular Characterization of a 5.75 kb Plasmid and a Chromosomally Encoded recA Gene Homologue from Bifidobacterium breve.* PhD thesis (National Univer of Ireland, Cork, Ireland).
2. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using *phred*. I. Accuracy assessment. *Genome Res* 8:175–185.
3. Ewing B, Green P (1998) Base-calling of automated sequencer traces using *phred*. II. Error probabilities. *Genome Res* 8:186–194.
4. Staden R, Beal KF, Bonfield JK (2000) The Staden package, 1998. *Methods Mol Biol* 132:115–130.
5. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 27:4636–4641.
6. Salzberg SL, Delcher AL, Kasif S, White O (1998) Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* 26:544–548.
7. Schiex T, Gouzy J, Moisan A, de Oliveira Y (2003) FrameD: A flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. *Nucleic Acids Res* 31:3738–3741.
8. Hyatt D, et al. (2010) Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.
9. Frishman D, Mironov A, Mewes HW, Gelfand M (1998) Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucleic Acids Res* 26:2941–2947.
10. Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. *Nat Genet* 3:266–272.
11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
12. Wheeler DL, et al. (2005) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 33(Database issue):D39–D45.
13. Rutherford K, et al. (2000) Artemis: Sequence visualization and annotation. *Bioinformatics* 16:944–945.
14. Claesson MJ, van Sinderen D (2005) BlastXtract—a new way of exploring translated searches. *Bioinformatics* 21:3667–3668.
15. Suzek BE, Ermolaeva MD, Schreiber M, Salzberg SL (2001) A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics* 17:1123–1130.
16. Overbeek R, et al. (2003) The ERGO genome analysis and discovery system. *Nucleic Acids Res* 31:164–171.
17. Riley M (1993) Functions of the gene products of *Escherichia coli*. *Microbiol Rev* 57: 862–952.
18. Riley M (1998) Systems for categorizing functions of gene products. *Curr Opin Struct Biol* 8:388–392.
19. Tusnády GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17:849–850.
20. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 305:567–580.
21. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795.
22. Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964.
23. Griffiths-Jones S, et al. (2005) Rfam: Annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* 33(Database issue):D121–D124.
24. Eddy SR (2002) A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure. *BMC Bioinformatics* 3:18.
25. Volfovsky N, Haas BJ, Salzberg SL (2001) A clustering method for repeat analysis in DNA sequences. *Genome Biol*, 2:RESEARCH0027.
26. Kurtz S, Schleiermacher C (1999) REPuter: Fast computation of maximal repeats in complete genomes. *Bioinformatics* 15:426–427.
27. Coutinho PM, Henrissat B (1999) In *Recent Advances in Carbohydrate Bioengineering*, eds Gilbert HJ, Davies G, Henrissat B, Svensson B (Royal Society of Chemistry, Cambridge), pp 3–12.
28. Busch W, Saier MH Jr; International Union of Biochemistry and Molecular Biology (IUBMB) (2004) The IUBMB-endorsed transporter classification system. *Mol Biotechnol* 27:253–262.
29. Sonnhammer ELL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 167:GC1–GC10.
30. Delcher AL, et al. (1999) Alignment of whole genomes. *Nucleic Acids Res* 27: 2369–2376.
31. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), 2nd Ed.
32. O'Connell Motherway M, et al. (2008) Characterization of ApuB, an extracellular type II amylopullulanase from *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* 74: 6271–6279.
33. Wells JM, Wilson PW, Le Page RW (1993) Improved cloning vectors and transformation procedure for *Lactococcus lactis*. *J Appl Bacteriol* 74:629–636.
34. Mazé A, O'Connell-Motherway M, Fitzgerald GF, Deutscher J, van Sinderen D (2007) Identification and characterization of a fructose phosphotransferase system in *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* 73:545–553.
35. O'Connell Motherway M, O'Driscoll J, Fitzgerald GF, Van Sinderen D (2009) Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in *Bifidobacterium breve* UCC2003. *Microb Biotechnol* 2:321–332.
36. van Hijum SA, et al. (2005) A generally applicable validation scheme for the assessment of factors involved in reproducibility and quality of DNA-microarray data. *BMC Genomics* 6:77.
37. Hinds J, Laing KG, Mangan JA, Butcher PD (2002) Microarrays for microbes: The BmuG@S approach. *Comp Funct Genomics* 3:333–337.
38. Zomer A, et al. (2009) An interactive regulatory network controls stress response in *Bifidobacterium breve* UCC2003. *J Bacteriol* 191:7039–7049.
39. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25:402–408.

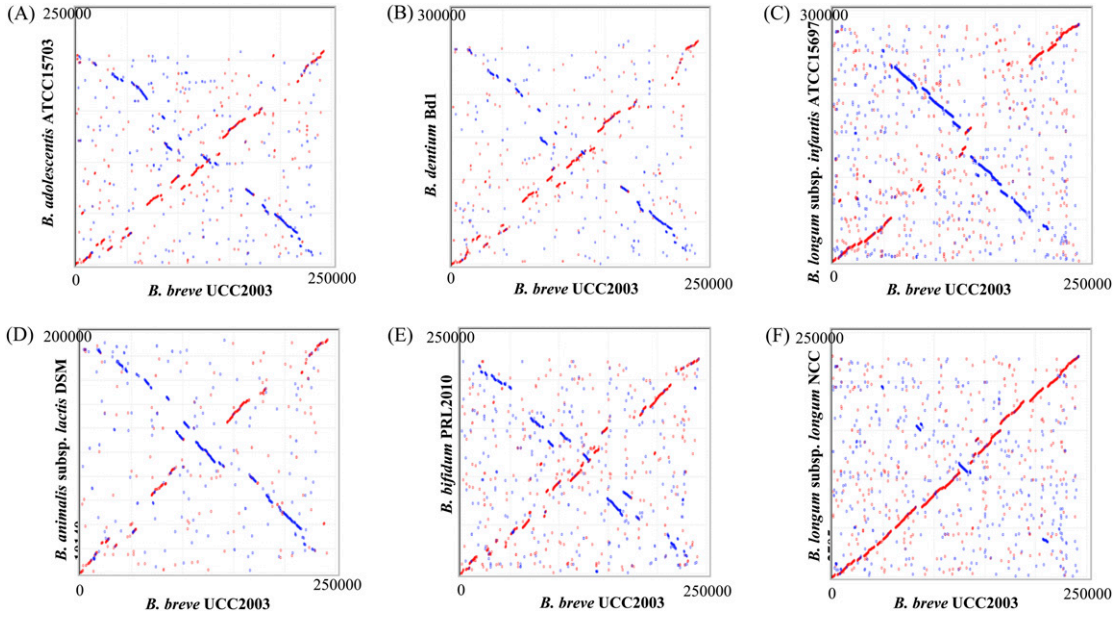

Fig. S1. Genome colinearity profiles of *B. breve* UCC2003 with (*A*) *B. adolescentis* ATCC15703, (*B*) *B. dentium* Bd1, (C) *B. longum* subsp. *infantis* ATCC15697, (*D*) *B. animalis* subsp. *lactis* DSM 10140, (*E*) *B. bifidum* PRL2010, and (*F*) *B. longum* subsp. *longum* NCC 2705 (1–6). Comparisons were performed using the MUMmer program, and plots were generated using MUMmerplot (7). Where two sequences agree, a dot is plotted. The forward matches are plotted in red, and the reverse matches are plotted in blue. If two sequences were perfectly identical, a single red line would go from the bottom left corner, the start position for each genome, to the top right corner.

1. Yasui K, et al. (2009) Intra-species diversity between seven *Bifidobacterium adolescentis* strains identified by genome-wide tiling array analysis. *Biosci Biotechnol Biochem* 73: 1422–1424.
2. Ventura M, et al. (2009) The *Bifidobacterium dentium* Bd1 genome sequence reflects its genetic adaptation to the human oral cavity. *PLoS Genet* 5:e1000785.
3. Sela DA, et al. (2008) The genome sequence of *Bifidobacterium longum* subsp. infantis reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci USA* 105:18964–18969.
4. Barrangou R, et al. (2009) Comparison of the complete genome sequences of *Bifidobacterium animalis* subsp. lactis DSM 10140 and Bl-04. *J Bacteriol* 191:4144–4151.
5. Turroni F, et al. (2010) Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proc Natl Acad Sci USA* 107:19514–19519.
6. Schell MA, et al. (2002) The genome sequence of *Bifidobacterium longum* reflects its adaptation to the human gastrointestinal tract. *Proc Natl Acad Sci USA* 99:14422–14427.
7. Delcher AL, et al. (1999) Alignment of whole genomes. *Nucleic Acids Res* 27:2369–2376.

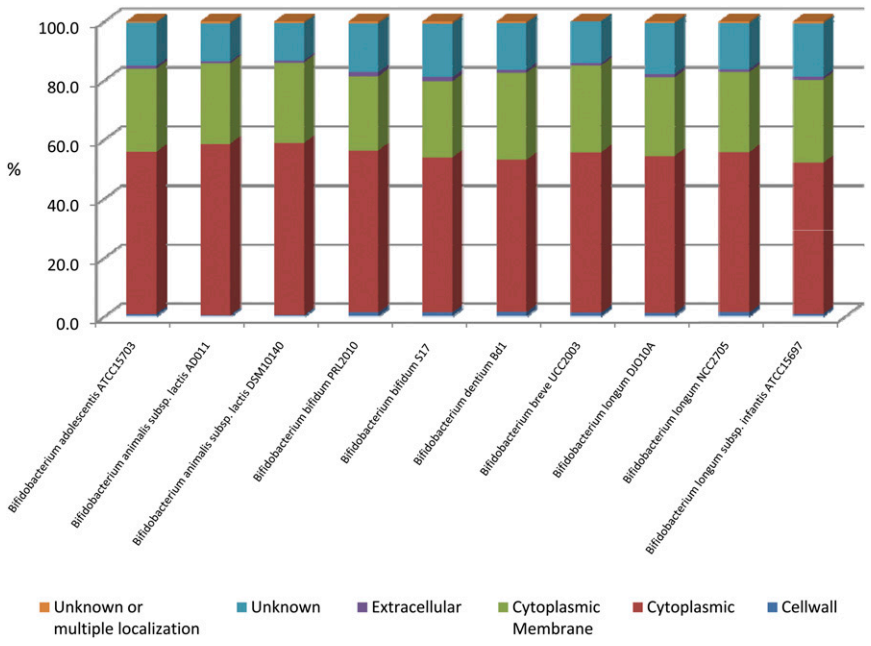

Fig. S2. Predicted subcellular location of the proteome of *B. breve* UCC2003 and representative bifidobacterial strains.

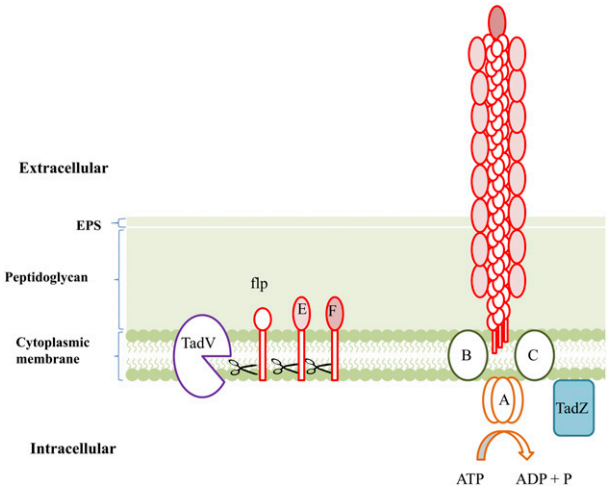

Fig. S3. Hypothetical structure and function of the *B. breve* UCC2003 Tad secretion system. The Tad proteins are labeled with the fourth letter of their designations. EPS, extracellular polymeric substances. [Adapted from Tomich et al. (1).]

1. Tomich M, Planet PJ, Figurski DH (2007) The Tad locus: Postcards from the widespread colonization island. *Nat Rev Microbiol* 5:363–375.

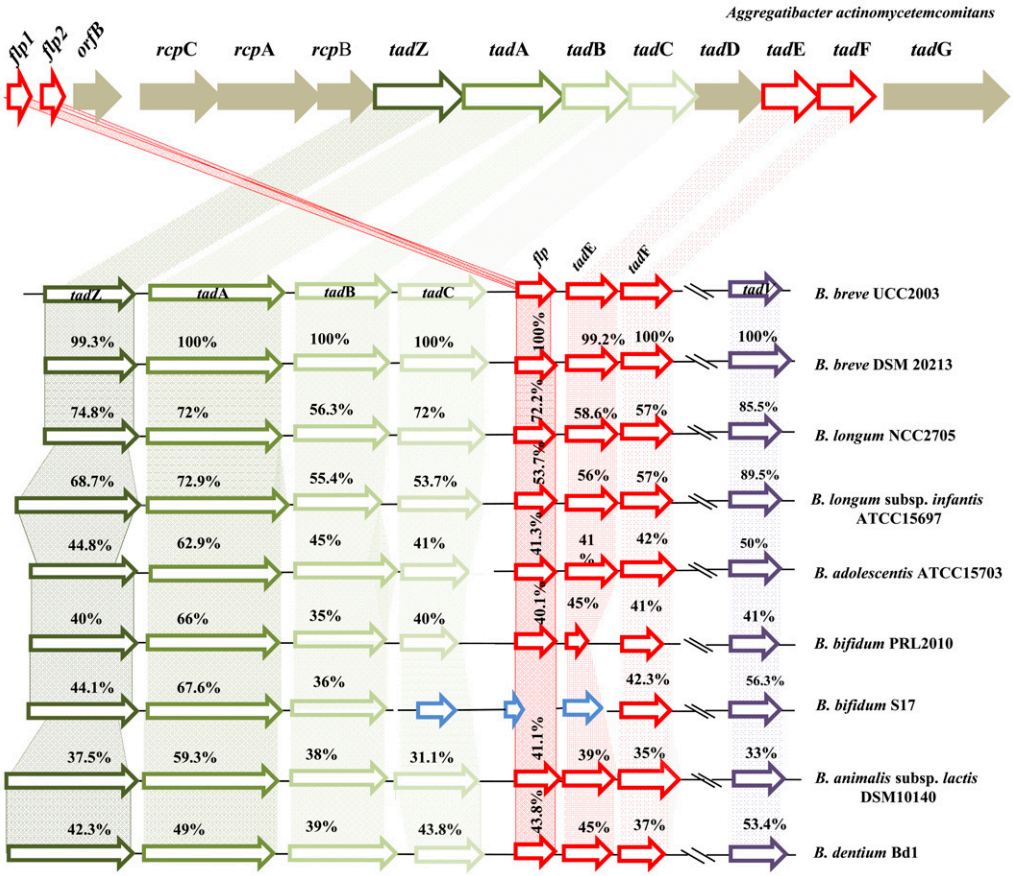

Fig. S4. Schematic diagram of Tad loci from *Aggregatibacter actinomycetemcomitans*, *B. breve* UCC2003, and putative tad genes from other representative bifidobacterial strains. Each arrow indicates an ORF. The predicted protein functions are indicated above the arrows. The levels of amino acid identity compared with *B. breve* UCC2003 (expressed as percentages) are indicated.

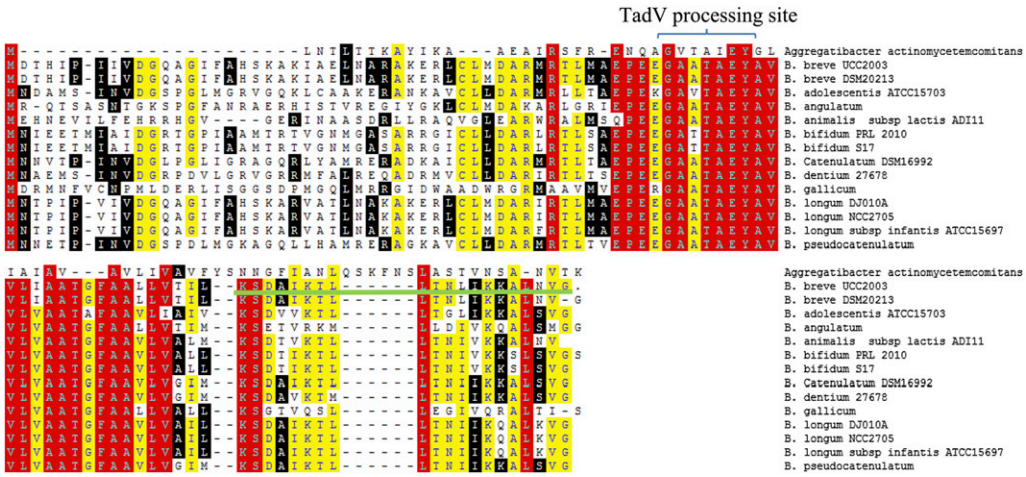

Fig. S5. Multiple sequence alignment of Flp from *A. actinomycetemcomitans* and predicted flp proteins from selected bifidobacterial strains. Amino acids are shaded in black if six of the aligned sequences contain identical residues at corresponding positions, in yellow if 10 of the aligned sequences contain identical residues at corresponding positions, and in red if 14 of the aligned sequences contain identical residues at corresponding positions. The position of the TadV processing site on *A. actinomycetemcomitans* Flp is indicated. The 20-amino acid sequence against which antibodies were raised is underlined in green.

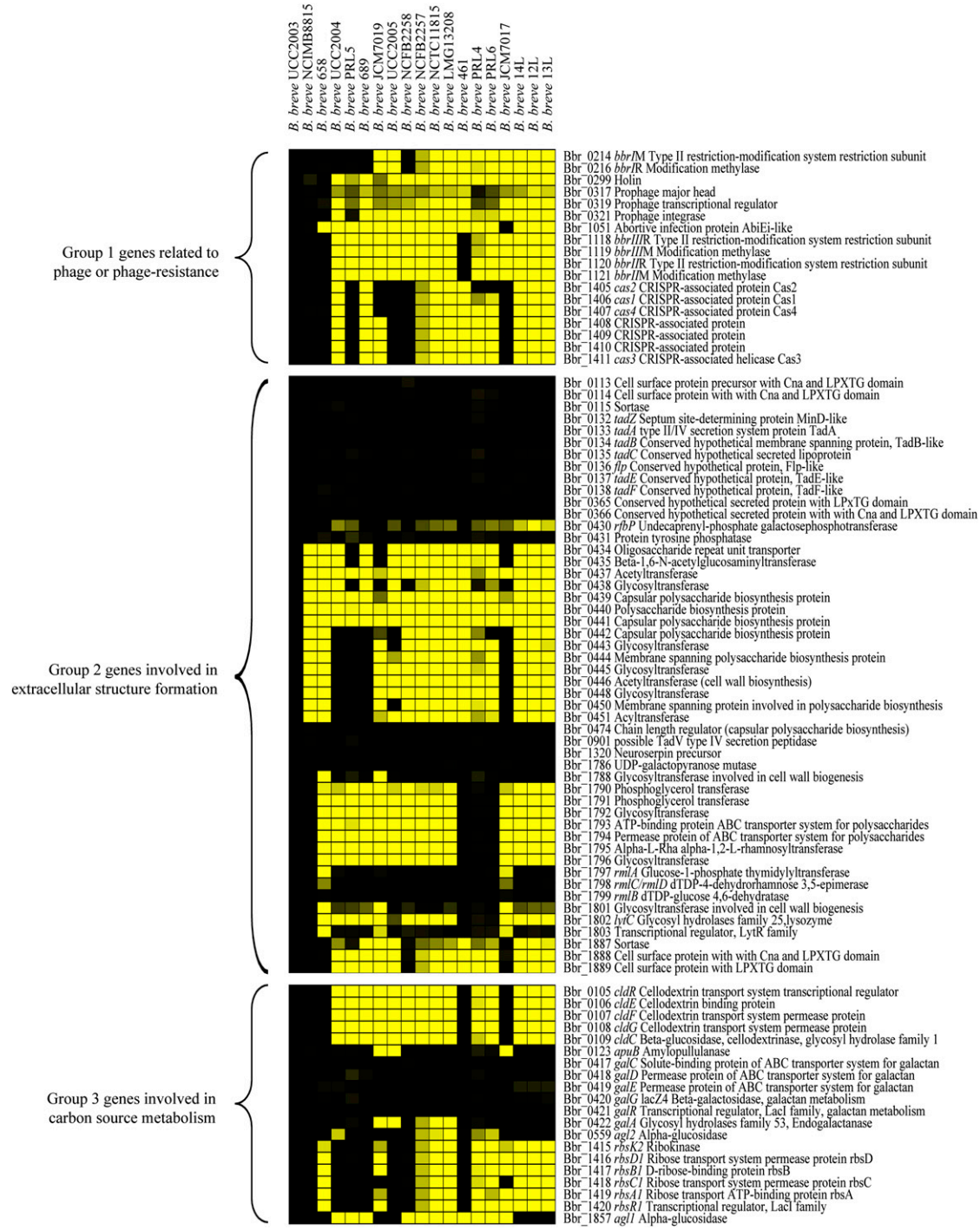

Fig. S6. Heat map of selected CGH data of genes related to phage or phage resistance, involved in extracellular structure formation, or involved in carbon source metabolism. The predicted or proven functions of the selected genes are given in the right-hand margin.

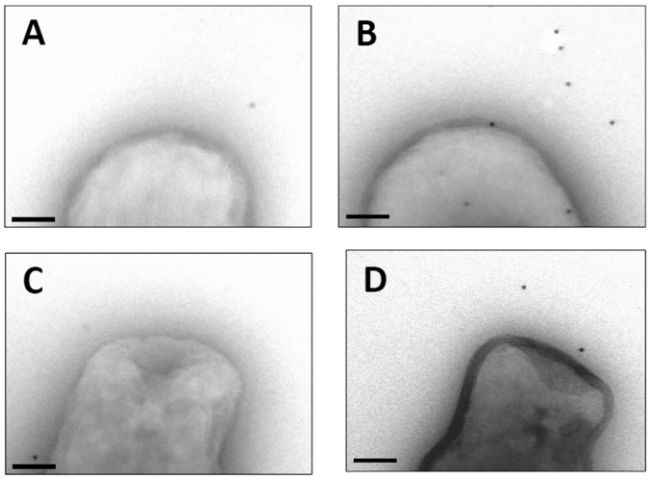

Fig. S7. Immunogold transmission electron microscopy analysis. *B. breve* UCC2003PK1 (*A* and *B*) and *B. breve* UCC2003-tadAPK1 mutant strain (*C* and *D*) following in vitro growth in MRS broth. No immunostaining of pili could be observed when both strains were labeled with Bbr_0136 antiserum (*A* and *C*) or Bbr_0136 preimmune serum (*B* and *D*). (Scale bars: 100 nm.)

Table S1. General features of bifidobacterial genomes

Feature	*B. breve* UCC2003	*B. adolescentis* ATCC15703	*B. longum subsp longum* NCC2705	*B. longum subsp longum* DJO10A	*B. longum subsp infantis* ATCC15697	*B. dentium* Bd1	*B. animalis* subsp. *lactis* DSM 10140	*B. bifidum* PRL2010
Genome size (bp)								
Chromosome	2,422,669	2,084,445	2,256,640	2,375,792	2,832,74	2,636,367	1,938438	2,214,650
Plasmid 1	—	—	3,626	3,661	—	—	—	—
Plasmid 2	—	—	—	10,073	—	—	—	—
G+C content (%)	58.73	59.18	60.12	60.15	59.86	59.32	60.48	62.67
No. of annotated genes	1,985	1,564	1,727	1,811	2,416	2,129	1,566	1,731
Avg. gene length (bp)	1,099	1,109	1,115	1,031	997	1,066	1,062	1,076
Coding diversity (%)	86	86.5	85.3	86.4	85.1	86.1	85.8	84.1
rRNA operons	2	5	4	4	4	4 (+1 5S rRNA gene)	4	3
tRNAs	54	54	57	58	79	55	52	52

Table S2. qRT-PCR analysis of selected in vivo microarray results

Locus tag	Function	Fold change qRT	*P* value qRT
Up-regulated genes			
Bbr_0044	Glycosyl transferase	132.4	8.68E-10
Bbr_0084	Glycosyltransferase	57.7	9.70E-10
Bbr_0116	malQ1 4-alpha-glucanotransferase	2.0	5.74E-02
Bbr_0117	agl3 Alpha-glucosidase	3.6	3.08E-05
Bbr_0118	malE Maltose/maltodextrin-binding protein	1.5	2.09E-02
Bbr_0119	malC Maltodextrin transport system permease protein malC	3.0	6.26E-05
Bbr_0120	malG Maltose transport system permease protein malG	1.4	7.37E-02
Bbr_0121	Conserved hypothetical membrane spanning protein	13.6	5.71E-04
Bbr_0122	Transcriptional regulator, LacI family	14.0	2.31E-04
Bbr_0124	dnaK Chaperone protein dnaK	1.8	1.73E-01
Bbr_0132	**tadZ Septum site-determining protein minD**	**25.1**	**7.19E-07**
Bbr_0133	**tadA type II/IV secretion system protein TadA**	**56.2**	**1.08E-07**
Bbr_0134	**tadB Conserved hypothetical membrane spanning protein, TadB-like protein**	**48.6**	**6.89E-07**
Bbr_0135	**tadC Conserved hypothetical secreted protein with prokaryotic membrane lipoprotein lipid attachment site**	**52.8**	**6.88E-07**
Bbr_0137	**tadE Conserved hypothetical protein, TadEike**	**44.1**	**1.41E-06**
Bbr_0138	**tadF Conserved hypothetical protein, TadF-like**	**62.1**	**4.18E-06**
Bbr_0352	Sulfatase family protein	6.2	1.63E-05
Bbr_0502	moxR1 MoxR protein	28.4	5.24E-08
Bbr_0868	Histidine kinase sensor of two-component system	2.1	5.49E-04
Bbr_0961	Lipase	58.3	6.85E-11
Bbr_1004	hrcA Heat-inducible transcription repressor hrcA	1.9	8.16E-03
Bbr_1060	Hypothetical membrane spanning protein	4.3	3.23E-05
Bbr_1068	RNA polymerase sigma-E factor	3.4	5.98E-03
Bbr_1080	Two component system histidine kinase	2.4	1.36E-04
Bbr_1318	Two-component response regulator	22.4	1.32E-05
Bbr_1320	Neuroserpin precursor	10.8	3.24E-04
Bbr_1356	clpC Negative regulator of genetic competence clpC/mecB	3.2	1.80E-03
Bbr_1696	Sugar-binding protein of ABC transporter system	28.2	3.84E-08
Bbr_1742	L-fucose permease	9.1	1.70E-08
Bbr_1925	rnpA Ribonuclease P protein component	1.0	2.06E-01
Down-regulated genes			
Bbr_0123	apuB Amylopullulanase	−1.6	6.65E-04
Bbr_0162	*N*-acetylmannosamine-6-phosphate 2-epimerase	−2.6	1.53E-03
Bbr_0168	*N*-acetylneuraminate lyase	−10.1	6.18E-06
Bbr_0171	Sialidase A	−2.2	6.44E-03
Bbr_0323	atpB ATP synthase A chain	−10.5	6.11E-06
Bbr_0795	clpP2 ATP-dependent Clp protease proteolytic subunit 2	−2.1	5.62E-03
Bbr_0846	nagA1 *N*-acetylglucosamine-6-phosphate deacetylase	−1.6	8.21E-02
Bbr_0847	nagB2 Glucosamine-6-phosphate isomerase	−2.7	9.62E-04
Bbr_0848	Sugar kinase, ROK family	−8.8	8.54E-06
Bbr_1668	Hsp10 10 kDa chaperonin GROES	−1.1	4.25E-01

Genes of the *B. breve* UCC2003 tad locus are highlighted in bold.

Table S3. 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$^+$ km^r	1
E. coli EC101-pNZ-M.BbrII + M.BbrIII	EC101 harboring pNZ8048 derivative containing *bbrIIM* and *bbrIIIM*	2
Bifidobacterium strains		
B. breve UCC2003	Isolate from nursling stool	3
B. breve UCC2003-tadA	pORI19-tet-tadA insertion mutant of UCC2003	This study
B. breve UCC2003PK1	*B. breve* UCC 2003 harboring pPKCM1 (cm^r)	This study
B. breve UCC2003-tadAPK1	*B. breve* UCC2003-tadA harboring pPKCM1 (tet^r cm^r)	This study
B. breve UCC2004	Isolate from human feces	UCC
B. breve UCC2005	Isolate from human feces	UCC
B. breve NIZOB658	Isolate from human feces	4
B. breve JCM7017	Isolate from human feces	JCM
B. breve JCM7019	Isolate from infant feces	JCM
B. breve NCFB2257	Isolate from infant intestine	NCFB
B. breve NCFB2258	Isolate from infant intestine	NCFB
B. breve NCTC11815	Isolate from infant intestine	NCTC
B. breve NCIMB8815	Isolate from human feces	NCIMB
B. breve LMG13208	Isolate from human feces	LMG
B. breve PRL4	Isolate from infant feces	PRL
B. breve PRL5	Isolate from infant feces	PRL
B. breve PRL6	Isolate from infant feces	PRL
B. breve 689	Isolate from infant feces	PRL
B. breve 461	Isolate from infant feces	PRL
B. breve 12L	Isolate from human milk	PRL
B. breve 13L	Isolate from human milk	PRL
B. breve 14L	Isolate from human milk	PRL
Plasmids		
pORI19	Emr, repA-, ori+, cloning vector	1
pORI19-tet-tadA	Internal 479 bp fragment of tadlA and tetW cloned in pORI19	This study

JCM, Japan Collection of Microorganisms; LMG, Belgian Co-ordinated Collection of Microorganisms; NCFB, National Collection of Food Bacteria; NCIMB, National Collection of Industrial and Marine Bacteria; NCTC, National Collection of Type Cultures; UCC, University College Cork Culture Collection; PRL Culture collection at Laboratory of Probiogenomics, University of Parma.

1. Law J, et al. (1995) A system to generate chromosomal mutations in *Lactococcus lactis* which allows fast analysis of targeted genes. *J Bacteriol* 177:7011–7018.
2. O'Connell Motherway M, O'Driscoll J, Fitzgerald GF, Van Sinderen D (2009) Overcoming the restriction barrier to plasmid transformation and targeted mutagenesis in *Bifidobacterium breve* UCC2003. *Microb Biotechnol* 2:321–332.
3. Mazé A, O'Connell-Motherway M, Fitzgerald GF, Deutscher J, van Sinderen D (2007) Identification and characterization of a fructose phosphotransferase system in *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* 73:545–553.
4. Gueimonde M, Noriega L, Margolles A, de los Reyes-Gavilan CG, Salminen S (2005) Ability of Bifidobacterium strains with acquired resistance to bile to adhere to human intestinal mucus. *Int J Food Microbiol* 101:341–346.

Table S4. Oligonucleotide primers used in this study

Purpose	Primer	Sequence
Construction of *tad*A insertion mutant	IMtadAF	tgcggaaagcttctggatgatgcgtgccctattg
	IMtadAR	ctatgctctagatccacaatctcctcgccacg
Conformation of tadA insertion mutant	tadA-Fw	atgtttttcgggttattggatg
Amplification of tetW	TetwF	tcagctgtcgacatgctcatgtacggtaag
	TetwR	gcgacggtcgaccattaccttctgaaacata

Table S5. qRT-PCR primers

Locus tag	Primer name forward	Primer sequence forward	Primer name reverse	Primer sequence reverse
Bbr_0044	fw_0044	tatgccgacgctttccatac	rv_0044	cgctgcccatcaggtataa
Bbr_0084	fw_0084	accatgtgggaatcgttgat	rv_0084	aggtggggtgagtcatgaag
Bbr_0116	fw_0116	catgatctgttcgagttcaacc	rv_0116	gagcgtcgttcacctgct
Bbr_0117	fw_0117	atgccgataagccgaatctt	rv_0117	cgaccggcagatataacgac
Bbr_0118	fw_0118	agatcgattgggagaaggcta	rv_0118	cgaggggcagcaacatag
Bbr_0119	fw_0119	ctcaccaaactatccgcagtc	rv_0119	ttgacgtattgcttgcgaac
Bbr_0120	fw_0120	ttcgtgtgcctcatctcg	rv_0120	catgcggctcatgcagta
Bbr_0121	fw_0121	attacgcccggtattggtg	rv_0121	cttgaacccgacttcgtcat
Bbr_0122	fw_0122	gctggcgatggatgagtt	rv_0122	agcggaatcgccatcata
Bbr_0123	fw_0123	catgaaggcatgggacatc	rv_0123	gtgaacgggtaggcagca
Bbr_0124	Fw_0124	cgtgtccgccaaggataa	Rv_0124	gaaccaccggtgatggtc
Bbr_0132	fw_0132	gcgctactacacccctcct	rv_0132	cctgttcgaccgttttctct
Bbr_0133	fw_0133	gcgtcgatttccatacgatt	rv_0133	agcccgcttagtaccgaaa
Bbr_0134	fw_0134	ctcaacgcggcctagtgt	rv_0134	caaaccagaccgaccacat
Bbr_0135	fw_0135	tggctgagcatggtgaag	rv_0135	cgtcatgattcgcccatt
Bbr_0137	fw_0137	gactgtgcatcgggaagg	rv_0137	cccgagtaggagtcacattca
Bbr_0138	fw_0138	agtcggatgatgtgcaggt	rv_0138	gcacctcccttatcgcatt
Bbr_0139	fw_0139	gcggtaccaaggtggatg	rv_0139	aaaattatgggagcgacctg
Bbr_0140	fw_0140	gcgatcgctgttctaccact	rv_0140	ctcgagcttggtgaggatct
Bbr_0161	fw_0161	ggacctgaagaaccgcatc	rv_0161	atctcgccagtttcgctgt
Bbr_0162	fw_0162	accagctgcatcaggtcat	rv_0162	cacagcttcgcctcgaac
Bbr_0168	fw_0168	cctacgacatccccgtgt	rv_0168	gccaagcttgacgagcag
Bbr_0171	fw_0171	aagggcagattcgcaaaat	rv_0171	ttgtcgaacaacgcggta
Bbr_0204	fw_0204	gggcttttcagcgacgta	rv_0204	ccgacgttatcgatcagca
Bbr_0323	fw_0323	cttcgccatgacgtgctt	rv_0323	agaatcgcgaagacataggc
Bbr_0352	Fw_0352	caagggaaacggcacctat	rv_0352	gatgaagggaaccctcacg
Bbr_0502	fw_0502	ttcccattcgttacaacgtg	rv_0502	gcaacacaatggtgcctgt
Bbr_0794	fw_0794	tctacgacaccatgcagctc	rv_0794	taacgcttgccctggtg
Bbr_0795	fw_0795	gaaatccaggccaaggaaat	rv_0795	gtcctgaccggtgtgattg
Bbr_0846	fw_0846	aacgcgatgaacggtctg	rv_0846	gatcagctcgatggtgacg
Bbr_0847	fw_0847	aaccaagtcccgatgcact	rv_0847	aagcgagcagcaccagat
Bbr_0848	Fw_0848	cactcgataccatcgtcagc	rv_0848	cgtgatgcagtcgagcag
Bbr_0868	fw_0868	ttcgatggcttcttcatcg	rv_0868	gcagccattccaccaatc
Bbr_0961	fw_0961	cgccgaacatgagtacga	rv_0961	cgaatgaaatcgaagatcacg
Bbr_1004	fw_1004	cgtcgcatgctggtactg	rv_1004	ctacgggctcttgcgaac
Bbr_1060	fw_1060	ttgtcgctggggttcagt	rv_1060	agtcgttgcctgagtgtcg
Bbr_1068	fw_1068	acgtgtggttgggttgct	rv_1068	cagccaggacagcacctt
Bbr_1080	fw_1080	ccaataaggtcacggtcctg	rv_1080	catcgctgcacattcgac
Bbr_1318	fw_1318	aggattgtcgaccaatgagc	rv_1318	catgaaggacagcgaattacc
Bbr_1320	fw_1320	aagtcaccacgctaccgttc	rv_1320	tatgtttggcaatccaatcg
Bbr_1356	fw_1356	ctggctctgacggtgagg	rv_1356	aggcttcggtctgttcagc
Bbr_1364	fw_1364	gtctccaaggacgagaccac	rv_1364	gcgagcgtcgatgtcttc
Bbr_1668	fw_1668	gttcactacgagggtgaggact	rv_1668	ccgagaatggcaagaatgtc
Bbr_1696	fw_1696	aggctgccaagaagatttcc	rv_1696	gtcatacggatcgggcttc
Bbr_1742	fw_1742	tataagggcggcgtcatc	rv_1742	ggcagggatcaccatgaa
Bbr_1925	fw_1925	gcatcgttctcatcgttgg	rv_1925	cgccttacgagccacttt