Supporting Information

Turroni et al. 10.1073/pnas.1011100107

SI Text

Mobilome of *B. bifidum* PRL2010 Genome. The genome of *Bifidobacterium bifidum* PRL2010 harbors 22 insertion sequence (IS) elements (Table S1 and Fig. S1). Additionally, a 19-kb DNA region, encompassing BBPR_0298–BBPR_0316, encodes proteins similar to recombinases and mobilization proteins (e.g., *par*A and *mob*A) and thus appears to represent an integrated plasmid, whereas a DNA region of 7 kb (BBPR_0078– BBPR_0083) is predicted to encode glycosyltransferases associated with an ATP-binding cassette (ABC) transport system, resembling a genetic composition that is linked to the production of a pellicle and that may provide protection against host phagocytosis by macrophages (2).

Conservation of Mucin-Degrading Genes in the *B. bifidum* Species. The genome variability among different strains of *B. bifidum* was assayed by comparative genomic hybridization (CGH) experiments through *B. bifidum* PRL2010-based microarrays. We highlighted which and how many ORFs from the sequenced *B. bifidum* PRL2010 strain did or did not hybridize with total genomic DNA extracted from seven *B. bifidum* strains originally isolated from fecal samples of different infants. When projected on the genome map of *B. bifidum* PRL2010, the CGH results highlight clustering of these variable ORFs within particular genomic regions. Based on the PRL2010 gene annotation, the types of genomic diversity thus identified can be assigned to two classes: (*i*) mobile DNA that constitutes the *B. bifidum* mobilome and (*ii*) plasticity regions of the *B. bifidum* genome, which may house specific genetic adaptations as a result of laterally acquired DNA or remnants of ancestral DNA that have not (yet) been lost. Various DNA segments that are present in *B. bifidum* PRL2010 but absent in other *B. bifidum* strains [the prophage-like element Bbif-1, as well as genes encoding partitioning functions (Par system) representing a presumptive integrated plasmid] clearly represent mobile DNA. Within the variable regions of the CGH map that are marked as plasticity regions, genes associated with bacterium–environment interaction and metabolic abilities appear to be particularly enriched. These genes include a putative pili/fimbriae-biosynthesis gene cluster (BBPR1707–BBPR1709), two different two-component regulatory systems (BBPR1388–BBPR1389 and BBPR1495–BBPR1496), and the pellicle-associated cluster (BBPR0770–BBPR0785) mentioned above. Furthermore, DNA regions encompassing four of the five restriction modification (R/M) systems detected in the genome of *B. bifidum* PRL2010 were highly variable in the different *B. bifidum* strains tested. The *B. bifidum* PRL2010 repertoire of R/M systems is the most numerous so far detected in bifidobacterial genomes (3–7).

***B. bifidum* PRL2010–Host Interaction** Cell-line gene-expression analyses were performed using Affymetrix gene arrays and MetaCore pathway analyses (GeneGo v. 6.1) and showed differential expression of various immune response and apoptotic/survival pathways. MetaCore pathway analysis revealed that the host immune response to *B. bifidum* PRL2010 exposure is strongly directed toward the IL-17 signaling pathway, confirming the specific immune responses driven by *B. bifidum* species, as previously described (8). Additionally, pathways assigned to mucin expression in cystic fibrosis involving Toll-like receptor, epidermal growth factor receptor, IL-6, and IL-17 showed strong regulation.

Other genes with cytoprotective properties up-regulated by *B. bifidum* included *GJA*1, a gap junction protein involved in barrier function, and epiregulin (*EREG*), which is involved in cell survival and proliferation. Finally, we evaluated the genes of *B. bifidum* PRL2010 whose expression was affected when PRL2010 cells were grown in presence of HT-29 cells by analyzing PRL2010 whole-genome transcriptional profiling. Eighty-two genes showed significantly changed expression when PRL2010 cells were exposed to HT-29 cells. Comparison of up-regulated versus down-regulated *B. bifidum* PRL2010 genes revealed that the only difference in the representation of clusters of orthologous groups (COGs) involved up-regulated genes associated with carbohydrate metabolism and transport and genes encoding for hypothetical proteins. Transcription of a large repertoire of *B. bifidum* PRL2010 genes involved in mucin breakdown (e.g., lacto-*N*-biosidase, endo-α-*N*-acetylgalactosaminidase, and lacto-*N*-biose-phosphorylase), as described in the main text, were shown to be changed upon exposure to HT-29 cells, thus providing additional evidence of the specific role of *B. bifidum* PRL2010 in the degradation of host-produced glycans. Such findings concerning glycan foraging by bifidobacteria in the human gut may have important implications, because this foraging may influence functional relationships between intestinal members of the gut microbiota (9).

Bacterial Strains, Growth Conditions, and Chromosomal DNA Extraction. Cultures were grown anaerobically in de Man–Rogosa–Sharpe (MRS) medium (Sharlau) supplemented with 0.05% L-cysteine-HCl and incubated at 37 °C for 16 h. Anaerobic conditions were achieved by the use of an anaerobic cabinet (Ruskin) in which the atmosphere consisted of 10% CO_2, 80% N_2, and 10% H_2. Bacterial DNA was extracted as described previously (10) and was subjected to further purification using the Qiagen Genomic DNA Purification Kit.

Carbohydrate Growth Assay. Cell growth on semisynthetic MRS medium supplemented with 1% (wt/vol) of a particular sugar was monitored at OD_{600} using a plate reader (Biotek). The plate reader was run in discontinuous mode, with absorbance readings performed in 60-min intervals and preceded by 30-s shaking at medium speed. Cultures were grown in biologically independent triplicates, and the resulting growth data were expressed as the mean of these replicates. Carbohydrates, including porcine mucin, were purchased from Sigma and Carbosynth. Porcine mucin is used routinely as a model for its human equivalent in various studies describing bacterial degradation of this substrate (11, 12).

Genome Sequencing and Assembly. Chromosomal DNA was sheared mechanically via a GeneMachine hydroshear device (Genomic Solutions), and the prepared inserts then were ligated into appropriate vectors. A fosmid library was constructed using the CopyControl Fosmid Production Kit (Epicentre). DNA was sheared, fragment-size selected by agarose pulsed-field gel electrophoresis, excised, and purified before ligation in the pCCqFos vector. The ligated vector was packaged using a MaxPlax Lambda Packaging Extract (Epicentre) kit and was used to transduce *Escherichia coli* (EP300). DNA for sequencing was produced using Templiphi (GE HeathCare) on aliquots of subclones grown in 384-well plates according to product specifications. Standard cycle sequencing from both ends of the subclones using universal primers was performed with BigDye v. 3.1 reaction (Applied Biosystems) and was resolved on ABI Prism 3730XL capillary instruments. Sequence reads were processed using Phred base-calling software and were monitored constantly against quality metrics using Phred Q20. The quality scores for each run were monitored through Agencourt's Galaxy LIMS system. A hybrid approach was used to obtain a single contig, which included a 15-fold sequencing

coverage using pyrosequencing technology on a 454 FLX instrument. The files generated by the 454 FLX instrument (Roche) were assembled with the Newbler software to generate a consensus sequence, which then was used for assembly using data from Sanger sequencing of the fosmid library using the Arachne genome assembly software (http://www.broadinstitute.org/science/programs/genome-biology/crd). Two rounds of additional sequencing walks were performed, resulting in a single contig (2,214,650 bp). Quality improvement of the genome sequence involved sequencing more than 400 PCR products (2,400 sequencing reads) across the entire genome to ensure correct assembly, double stranding, and the resolution of any remaining base conflicts. The genome sequence finally was edited to a Phred confidence value of 30 or more. Based on the final consensus quality scores, we estimate an overall error rate of <1 error per 10^5 nucleotides.

Sequence Annotation. Protein-encoding ORFs were predicted using a combination of the software programs Glimmer (13) and FrameD (14) as well as a comparative analysis involving Orpheus (15), BLASTX (16), and Prodigal (17). Results of the four gene-finder programs were combined manually, and a preliminary identification of ORFs was performed on the basis of BLASTP (18) analysis against a nonredundant protein database provided by the National Center for Biotechnology Information. The results of the combined gene finders and the associated BLASTP results were inspected by Artemis (19), which was used for a manual editing effort to verify and, if necessary, to redefine the start of every predicted coding region or to remove or add coding regions.

Assignment of protein function to predicted coding regions of the *B. bifidum* PRL2010 genome was performed manually. Moreover, the revised gene/protein set was searched against the Swiss-Prot (www.expasy.ch/sprot/)/TrEMBL, PRIAM (http://priam.prabi.fr/), protein family (Pfam, http://pfam.sanger.ac.uk/), TIGRFam (http://www.jcvi.org/cms/research/projects/tigrfams/overview/), Interpro (INTERPROSCAN, http://www.ebi.ac.uk/Tools/InterProScan/), Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/), and COGs (http://www.ncbi.nlm.nih.gov/COG/) databases, in addition to BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi) vs. nr. From all these results, functional assignments were made. Manual corrections to automated functional assignments were completed on an individual gene-by-gene basis as needed.

Bioinformatic Analyses. Transfer RNA genes were identified using tRNAscan-SE (20). Ribosomal RNA genes were detected on the basis of BLASTN searches and were annotated manually. Insertion sequence elements were identified using Repeat Finder (21) and BLAST (18) 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 (22), and transporter classification was performed according to the Transporter Classification Database scheme (23).

Variances in guanine and cytosine (G+C) content were profiled by the DNA segmentation algorithm hosted at http://tubic.tju.edu.cn/GC-Profile/ (24). Atypical codon usage regions were mapped using the factorial correspondence analysis through the assistance of the GCUA software (25).

CGH Microarray, Description, Labeling, and Hybridizations. CGH analysis was performed with a *B. bifidum* PRL2010 array. A total of 11,064 probes 35 bp in length were designed on 1,844 ORFs using OligoArray 2.1 software (26). Oligos were synthesized in triplicate on a 2 × 40-k CombiMatrix array. Replicates were distributed on the chip in random, nonadjacent positions. A set of 19 negative control probes designed on phage and plant sequences also was included on the chip. Two micrograms of purified genomic DNA was labeled with Cy5-ULS using the Kreatech ULS array CGH Labeling kit (Kreatech Diagnostics) according to the supplier's instructions. Hybridization of labeled DNA to *B. bifidum* PRL2010 arrays was performed according to CombiMatrix protocols (http://www.combimatrix.com/support_docs.htm).

CGH Microarray Data Acquisition and Treatment. Fluorescence scanning was performed on a ScanArray 4000XL confocal laser scanner (Perkin-Elmer). Signal intensities for each spot were determined using Microarray Imager 5.8 software (CombiMatrix). Signal background was calculated as the mean of negative controls plus two times the SD (27). A global quantile normalization was performed (28), and log2 ratios among the reference sample (*B. bifidum* PRL2010) and the other samples analyzed were calculated. The distribution of the log2-transformed ratios for each hybridization reaction was calculated separately. Log2-transformed ratios of each probe were visualized and ranked by position on the *B. bifidum* PRL2010 genome by a heatmap using TMev 4.0 software (http://www.tm4.org/mev.html). Hierarchical clustering was performed with average linkage and Euclidean distance (29) using TMev 4.0 software.

Proteomic Sample Preparation. *B. bifidum* PRL2010 was taken at the exponential phase of growth (at an OD_{600} of 0.5) and normalized at an OD_{600} of 1.0 by concentration. After centrifugation to remove medium, 15 mL of cells were washed three times by PBS. The cell pellet was resuspended in 600 μL of lysis buffer containing 100 mM Tris and 8.0 M urea and then was lysed mechanically using silica beads and a bead-beater (FastPrep; QBiogene,) for eight cycles of 30-s pulses, each with a 30-s interval on ice. Beads and cell debris were removed by centrifugation, and the soluble fraction was stored at −80 °C for further analysis. Protein concentration was measured by the BioRad Protein Assay Kit. A volume of 200 mg/mL of protein was transferred to a new cap tube and precipitated with ethanol [75% (vol/vol)] at −20 °C. After centrifugation, the protein pellet was resuspended in 100 μL of 0.1 M Tris/1M urea buffer (pH 8.0). Proteins then were digested overnight by 5 μg of MS-grade trypsin (Promega) at 37 °C. The tryptic peptides were purified by Macro Trap with peptide concentration and desalting cartridge (Michrom) according to the manufacturer's manual. The peptides were eluted in 98% acetonitrile in water and then were dried before mass spectrometry analysis.

Protein Identification. The digested protein samples were submitted to the Genome Center Proteomics Core at the University of California, Davis. Protein identification was performed using an Eksigent Nano LC 2D system coupled to an LTQ ion-trap mass spectrometer (Thermo-Fisher) using a Picoview nano-spray source. Peptides were loaded onto a nanotrap (Zorbax 300SB-C18; Agilent Technologies) at a loading flow rate of 5.0 μL/min. Peptides then were eluted from the trap and separated using a nano-scale 75 μm × 15 cm New Objectives picofrit column that was packed in house. The top 10 ions in each survey scan were subjected to automatic low-energy collision-induced dissociation.

Tandem mass spectra were extracted, and the charge states were deconvoluted by BioWorks v. 3.3. All MS/MS samples were analyzed using X! Tandem (GPM-XE manager, v. 2.2.1). X! Tandem was set up to search against the *B. bifidum* PRL2010 whole proteome. X! Tandem was searched with a fragment ion mass tolerance of 0.60 Da. Oxidation of methionine was specified as a variable modification in X! Tandem. The cutoff of log(e) for the peptide [log(e)] was set at less than −2, and a protein with a log(e) score lower than −6 was considered to represent an identity hit.

RNA Isolation. RNA was isolated according to the protocol described previously (30). The quality of the RNA was checked by analyzing the integrity of rRNA molecules by Experion (BioRad).

Expression Microarray. cDNA was synthesized using the cDNA synthesis and labeling kit (Kreatech) according to the manufacture's instructions. DNA microarrays containing oligonucleotide primers representing each of the 1,843 annotated genes of *B. bifidum* PRL2010 were obtained from Agilent Technologies. Labeled cDNA was hybridized using the Agilent Gene Expression hybridization kit (#5188–5242) as described in the manual for Agilent Two-Color Microarray-Based Gene Expression Analysis v. 4.0 (publication G4140-90050). Following hybridization, microarrays were washed as described in the manual and scanned using an Agilent G2565A DNA microarray scanner. The scanning results were converted to data files with Agilent's Feature Extraction software (v. 9.5). Differential expression tests were performed with the Cyber-T implementation of a variant of the *t* test (31). A gene was considered differentially expressed between a test condition and a control when an expression ratio >5 or <0.2 relative to the result for the control was obtained with a corresponding *P* value that was <0.001. Final data presented are the averages from at least two independent array experiments.

Tissue Culture Experiments. All cell-culture reagents, unless otherwise specified, were from Sigma Aldrich. We seeded 2×10^5 Caco-2 or HT-29 cells in 1.5 mL of DMEM (high-glucose Hepes) supplemented with 10% heat-inactivated FBS (FSC; Gibco), penicillin (100 U/mL), streptomycin (0.1 mg/mL), amphotericin B (0.25 μg/mL), and 4 mM L-glutamine into the upper compartments of a six-well transwell plate (Corning). The lower compartments contained 3.0 mL of the same medium. Cells were incubated at 37 °C in a 5% CO_2 atmosphere until 3 d postconfluence and then were washed with Hanks' solution to remove antibiotics and FCS and were stepped down in DMEM supplemented with L-glutamine (4 mM), sodium selenite (0.2 μg/mL), and transferrin (5 μg/mL) for 24 h without antibiotics. These transwell inserts were transferred to an anaerobic culture box within an MACS-MG-1000 anaerobic workstation (Don Whitley) at 37 °C. The upper compartment of each insert (apical cell surface) was filled with anaerobic DMEM cell medium to maintain the anaerobic conditions necessary to support bacterial viability and growth. The lower compartment of each insert, which was completely sealed from the upper compartment, was filled with oxygenated DMEM previously shown to support fully intestinal epithelial cell viability. Furthermore, in previous optimization experiments, resazurin was used to demonstrate lack of gaseous exchange between the apical and basal cell compartments. The *B. bifidum* PRL2010 culture was harvested at exponential phase by centrifugation at 3,500 × *g* for 5 min. The pellet was washed with 10 mL of anaerobic DMEM and was resuspended in 0.8 mL of the same medium. One hundred microliters of bacterial suspension (10^8 cfu/mL) was added to experimental wells. The control wells received the same amount of medium without bacterial cells. Additional control included bacterial cells incubated without Caco-2 and HT-29 cells.

Bacterial and tissue culture cells were harvested for analysis after 2 h and 4 h of incubation. Nonadherent bacteria were aspirated carefully from the wells and pooled. The adherent fraction was collected after washing of the inserts with anaerobic medium and also was pooled. Each fraction was collected into 1.5-mL tubes and centrifuged at 3,500 × *g* for 5 min, and the resulting pellet was resuspended in 400 μL of RNAlater (Ambion) and submitted to RNA extraction following the protocol described above. Caco-2 cells or HT-29 cells were harvested from the wells, pooled, and stored in RNAlater at 4 °C.

Eukaryotic RNA Isolation. Caco-2 and HT-29 cells stored in RNAlater were diluted 1:1 in an equal volume of sterile PBS, followed by centrifugation at 5,000 × *g* for 10 min at 4 °C. Total RNA from the pellet was isolated using the RNeasy mini kit (Qiagen.) according to the manufacturer's instructions, including an RNase-free DNase I (Qiagen) digestion step. Eukaryotic RNA integrity was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies.).

Human Microarray Hybridizations and Data Analysis. Eight micrograms of eukaryotic total RNA was reverse-transcribed to cDNA and then transcribed into biotin-labeled cRNA using the One-Cycle Target Labeling Kit (Affymetrix) according to the manufacturer's instructions. cRNA quality was determined using the Agilent 2100 Bioanalyzer. Hybridization to the custom-designed human NuGO GeneChip array (Affymetrix) on a GeneChip Fluidics Station 450 (Affymetrix) was performed at the Institute of Medical Sciences Microarray Core Facility (University of Aberdeen, Aberdeen, United Kingdom). Chips were scanned with an Affymetrix GeneChip Scanner 3000. Image-quality analysis was performed using GeneChip Operating Software (GCOS) (Affymetrix).

Further quality analysis, normalization [GeneChip Robust Multiarray Average (GC-RMA)], statistical analysis, and heatmap generation was performed with the freely available software packages R (http://www.r-project.org) and Bioconductor (http://www.bioconductor.org) (32). The moderated *t* test provided by the Bioconductor package *limma* was used to test for differential expression. Data were considered significant when $P < 0.05$ using the Benjamini–Hochberg false-discovery method (33).

Functional Analysis of Microarray Data. All differentially expressed genes ($P < 0.05$) were imported into MetaCore analytical software (GeneGo) to generate pathway maps. Ranking of relevant integrated pathways was based on *P* values calculated using hypergeometric distribution.

Real-time PCR Analysis of Differentially Expressed Genes. ifferentially expressed genes as observed by microarray analyses using mRNA obtained from various test groups were validated further using real-time PCR. One microgram of total eukaryotic RNA isolated from the Caco-2 and HT-29 cells was reverse-transcribed into cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) with random primers. Real-time PCR analysis was performed using a 7500 Fast Real-Time PCR System (Applied Biosystems) with the QuantiFast SYBR Green PCR Kit (Qiagen) and QuantiTect Primer Assays (Qiagen) according to the manufacturer's recommendations. PCR cycling conditions were as follows: one cycle at 95 °C for 5 min, followed by 40 cycles at 95 °C for 10 s and at 60 °C for 30 s, ending with a dissociation step. All samples were run in triplicate. Hypoxanthine guanine phosphoribosyl transferase (*HPRT*) was selected as a reference gene for normalization because of its low variation between samples in the microarray analysis.

Data were analyzed on a logarithmic scale with base 2 by Student's *t* test allowing unequal variances with $P < 0.05$ to be considered statistically significant. The SEs of differences also were calculated on this scale. Differences were back-transformed to calculate fold changes.

1. Ventura M, et al. (2009) Comparative analyses of prophage-like elements present in bifidobacterial genomes. *Appl Environ Microbiol* 75:6929–6936.
2. Chapot-Chartier MP, et al. (2010) Cell surface of Lactococcus lactis is covered by a protective polysaccharide pellicle. *J Biol Chem* 285:10464–10471.
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. Lee JH, et al. (2008) Comparative genomic analysis of the gut bacterium Bifidobacterium longum reveals loci susceptible to deletion during pure culture growth. *BMC Genomics* 9:247.
5. 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.
6. 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.

7. Ventura M, et al. (2009) The Bifidobacterium dentium Bd1 genome sequence reflects its genetic adaptation to the human oral cavity. *PLoS Genet* 5:e1000785.
8. López P, Gueimonde M, Margolles A, Suárez A (2010) Distinct Bifidobacterium strains drive different immune responses in vitro. *Int J Food Microbiol* 138: 157–165.
9. Sonnenburg JL, Chen CT, Gordon JI (2006) Genomic and metabolic studies of the impact of probiotics on a model gut symbiont and host. *PLoS Biol* 4:e413.
10. Ventura M, Elli M, Reniero R, Zink R (2001) Molecular microbial analysis of Bifidobacterium isolates from different environments by the species-specific amplified ribosomal DNA restriction analysis (ARDRA). *FEMS Microbiol Ecol* 36:113–121.
11. Ruas-Madiedo P, Gueimonde M, Fernández-García M, de los Reyes-Gavilán CG, Margolles A (2008) Mucin degradation by Bifidobacterium strains isolated from the human intestinal microbiota. *Appl Environ Microbiol* 74:1936–1940.
12. Derrien M, Vaughan EE, Plugge CM, de Vos WM (2004) Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol* 54:1469–1476.
13. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 27:4636–4641.
14. 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.
15. 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.
16. Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. *Nat Genet* 3:266–272.
17. Hyatt D, et al. (2010) Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.
18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
19. Rutherford K, et al. (2000) Artemis: Sequence visualization and annotation. *Bioinformatics* 16:944–945.
20. 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.
21. Volfovsky N, Haas BJ, Salzberg SL (2001) A clustering method for repeat analysis in DNA sequences. *Genome Biol* 2(8):RESEARCH0027.
22. Coutinho PM, Henrissat B (1999) Life with no sugars? *J Mol Microbiol Biotechnol* 1: 307–308.
23. Busch W, Saier MH, Jr (2002) The transporter classification (TC) system, 2002. *Crit Rev Biochem Mol Biol* 37:287–337.
24. Gao F, Zhang CT (2006) GC-Profile: A web-based tool for visualizing and analyzing the variation of GC content in genomic sequences. *Nucleic Acids Res* 34(Web Server issue): W686–691.
25. McInerney J, Kooij D (1997) Economic analysis of alternative AD control programmes. *Vet Microbiol* 55:113–121.
26. Rouillard JM, Zuker M, Gulari E (2003) OligoArray 2.0: Design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res* 31:3057–3062.
27. Bilban M, Buehler LK, Head S, Desoye G, Quaranta V (2002) Defining signal thresholds in DNA microarrays: Exemplary application for invasive cancer. *BMC Genomics* 3:19.
28. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19:185–193.
29. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863–14868.
30. Ventura M, et al. (2005) The ClgR protein regulates transcription of the clpP operon in Bifidobacterium breve UCC 2003. *J Bacteriol* 187:8411–8426.
31. Long AD, et al. (2001) Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in Escherichia coli K12. *J Biol Chem* 276:19937–19944.
32. Gentleman RC, et al. (2004) Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.
33. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:1–10.
34. Gill SR, et al. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312:1355–1359.
35. Eckburg PB, et al. (2005) Diversity of the human intestinal microbial flora. *Science* 308: 1635–1638.

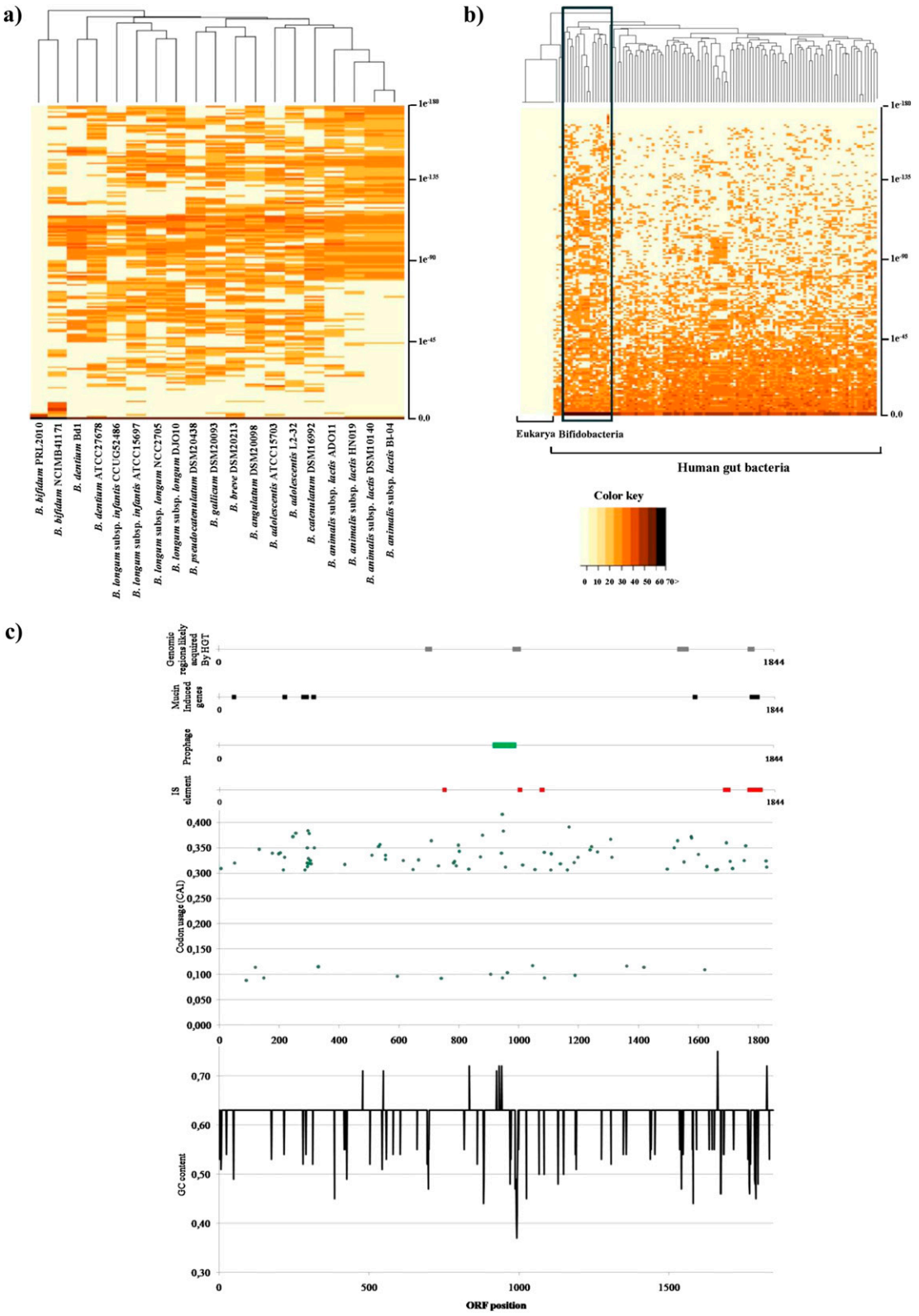

Fig. S1. Comparative genomic clustering analyses of *B. bifidum* PRL2010 and other bifidobacterial species (*A*) as well as with bacterial and eukaryotic genomes from the human gut microbial genomes (*B*), and identification of mobile genetic elements of the *B. bifidum* PRL2010 genome (*C*). In *A*, the *x* axis shows all available bifidobacterial genomes. The *y* axis indicates all E-value frequencies between 0 and $1e^{-180}$ resulting from BLAST comparisons of available bifidobacterial strains and *B. bifidum* PRL2010 as the reference genome. In *B*, the *x* axis shows all the genomes of bacteria cultured from the human gut representing the previously identified colonic dataset (34, 35) and Eukarya genomes, including that of *Mus musculus* as outgroup. The *y* axis shows the similarity level (all E-value frequencies of BLAST hits between 0 and $1e^{-180}$) of each organism compared with the *B. bifidum* PRL2010 genome as the reference genome. In *B*, the gut bifidobacterial genomes are marked. In *A* and *B*, the white-to-orange color code represents the level of similarity as shown in the color key provided. In *C*, IS elements and predicted prophage-like elements are labeled in red and green, respectively. The lower plot indicates the deviation of the G+C content of each ORF of the *B. bifidum* PRL2010 genome from the mean average (62.67%). In the upper plot, each dot represents an ORF displaying a biased codon usage determined by factorial correspondence analysis of codon usage.

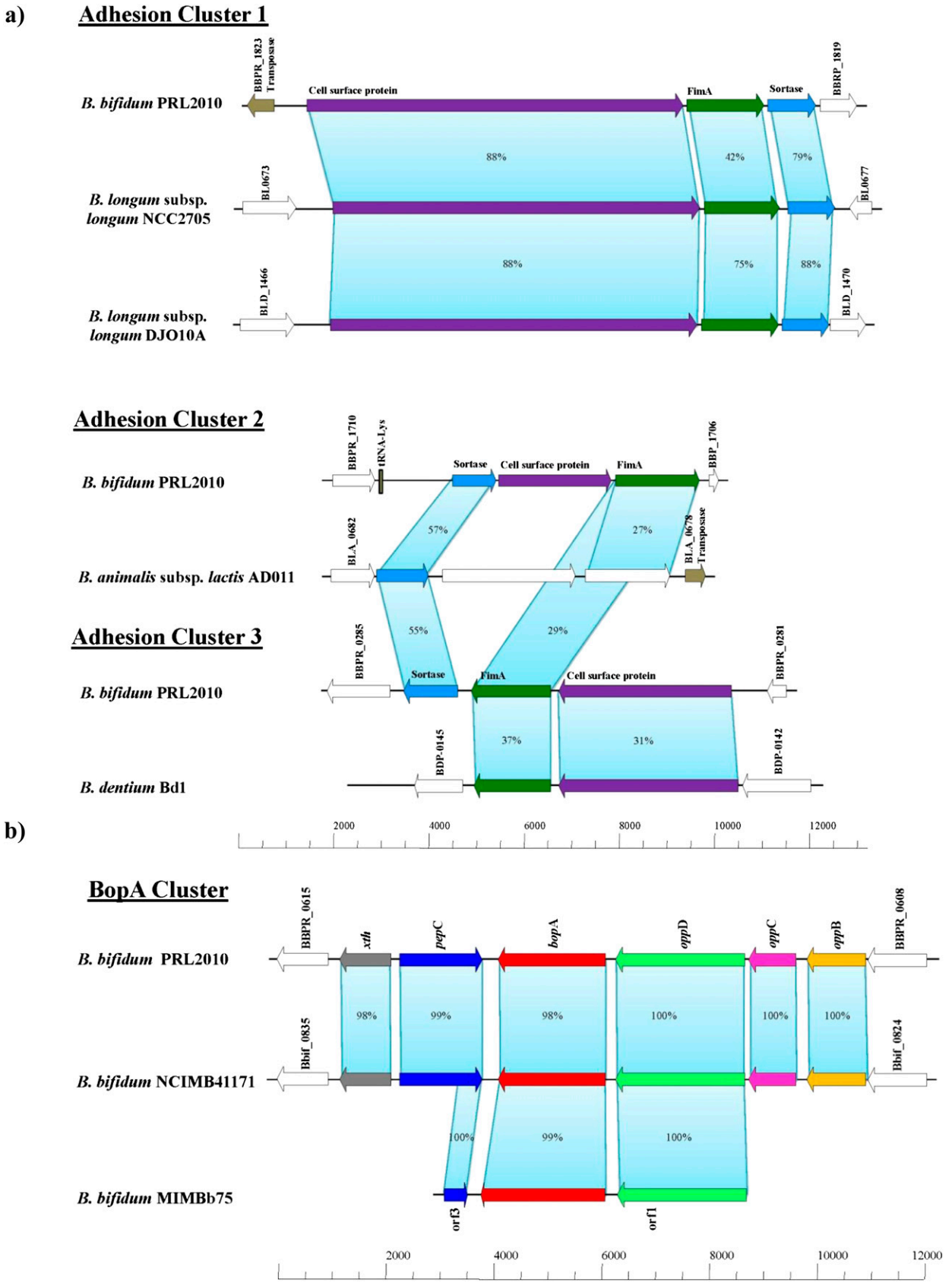

Fig. S2. Schematic comparative representation of potential adhesion factors encoded by *B. bifidum* PRL2010 and by various other bifidobacterial strains. Each arrow indicates an ORF, the size of which is proportional to the length of the arrow. The predicted protein function is indicated above each arrow. The amino acid identity in percentages is indicated.

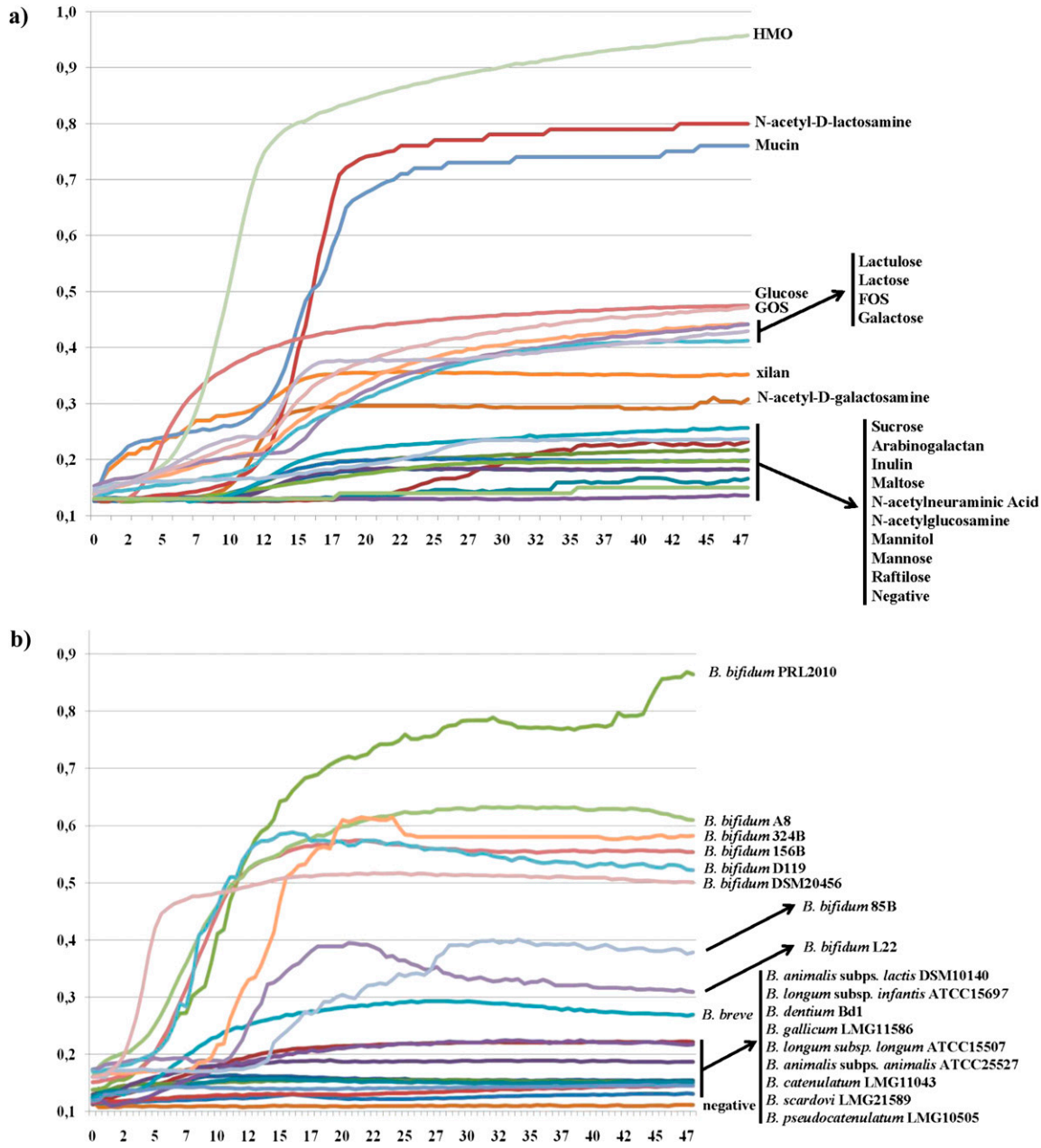

Fig. S3. Growth curves of bifidobacteria in a growth medium containing mucin as the sole carbon source. Growth was measured at OD_{600}. (*A*) Growth curves of *B. bifidum* PRL2010 with different carbohydrates as the sole carbon source. HMO, human milk oligosaccharides; FOS, fructo oligosaccharides; GOS, galacto oligosaccharides. (*B*) Growth curves of various *B. bifidum* strains as well as various human intestinal bifidobacteria species on mucin-based medium.

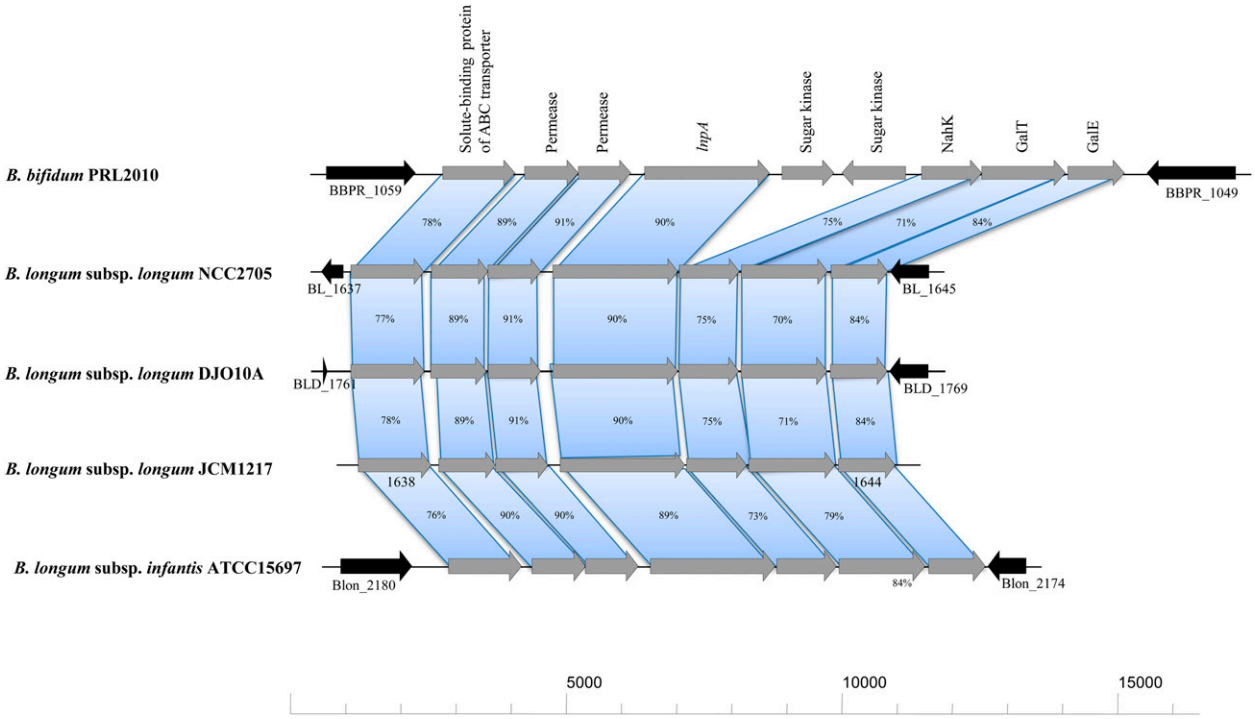

Fig. S4. Comparison of the galacto-*N*-biose/lacto-*N*-biose (GNB/LNB) locus (BBPR_1050–BBPR_1058) in *B. bifidum* PRL2010 with the corresponding loci in different bifidobacteria. Each arrow indicates an ORF. The length of the arrow is proportional to the length of the predicted ORF. Corresponding genes are shown in the same color. The putative function of the protein is indicated above each arrow.

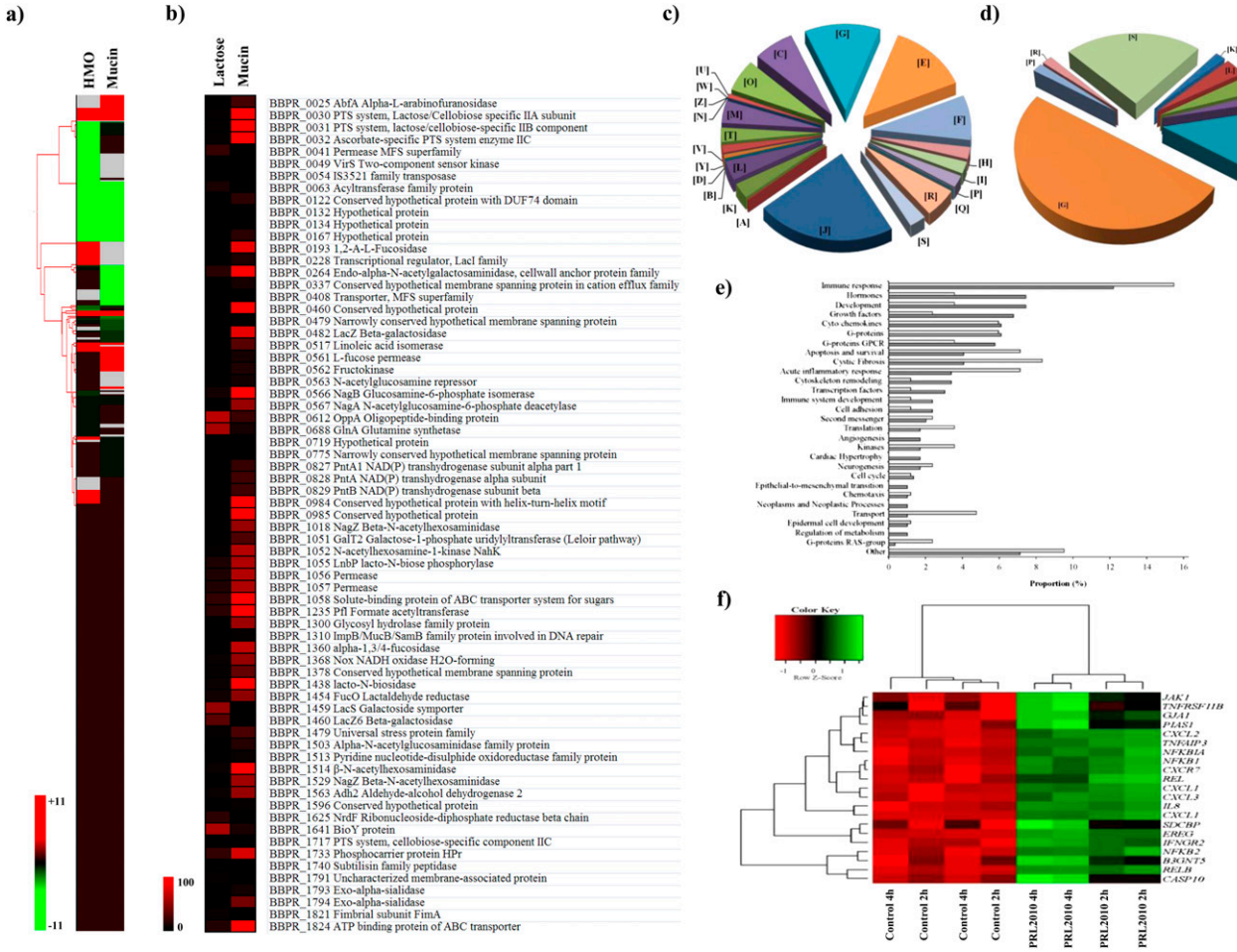

Fig. S5. Identification of proteins differentially expressed in *B. bifidum* PRL2010 by proteomic analysis (*A* and *C*) or by transcriptomics analysis (*B* and *D*) of host-induced genes after exposure of human intestinal cells to PRL2010 cells (*E* and *F*). *A* indicates the change in the expression upon cultivation of PRL2010 cells in HMO- and mucin (MUC)-based medium. Red indicates increased protein expression levels; green indicates decreased protein expression level as compared with the glucose-grown samples. In *B* the heatmap indicates the change in the expression upon cultivation of PRL2010 cells in mucin-based medium. Each row represents a separate transcript, and each column represents a separate sample. Red indicates increased transcription levels as compared with the reference (lactose-grown) samples. *C* and *D* show COG functional categories of the *B. bifidum* PRL2010 proteins overexpressed in the presence mucin relative to lactose in proteomics experiments (*C*) and in transcriptomics analysis (*D*). Colors indicate COG families, and each COG family is identified by a one-letter abbreviation: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell-cycle control and mitosis; E, amino acid metabolism and transport; F, nucleotide metabolism and transport; G, carbohydrate metabolism and transport; H, coenzyme metabolism; I, lipid metabolism; J, translation; K, transcription; L, replication and repair; M, cell wall/membrane/envelop biogenesis; N, cell motility; O, posttranslational modification, protein turnover, chaperone functions; P, inorganic ion transport and metabolism; Q, secondary structure; R, general functional prediction only; S, function unknown; T, signal transduction; U, intracellular trafficking and secretion; Y, nuclear structure; Z, cytoskeleton. *E* shows the classification of host-induced genes according to cell-process categories. White bars indicate host genes induced upon exposure of Caco-2 cells for 2 h; gray bars represent host genes induced upon exposure of Caco-2 cells for 4 h. In *F*, the heatmap indicates the change in host gene expression related to immune response and mucin production upon exposure of host intestinal Caco-2 cells to PRL2010 cells for 2 and 4 h, respectively. Each row represents the transcript of a separate host, and each column represents a separate PRL2010 sample. The color legend is above the array plot; green indicates increased transcription levels as compared with the reference samples.

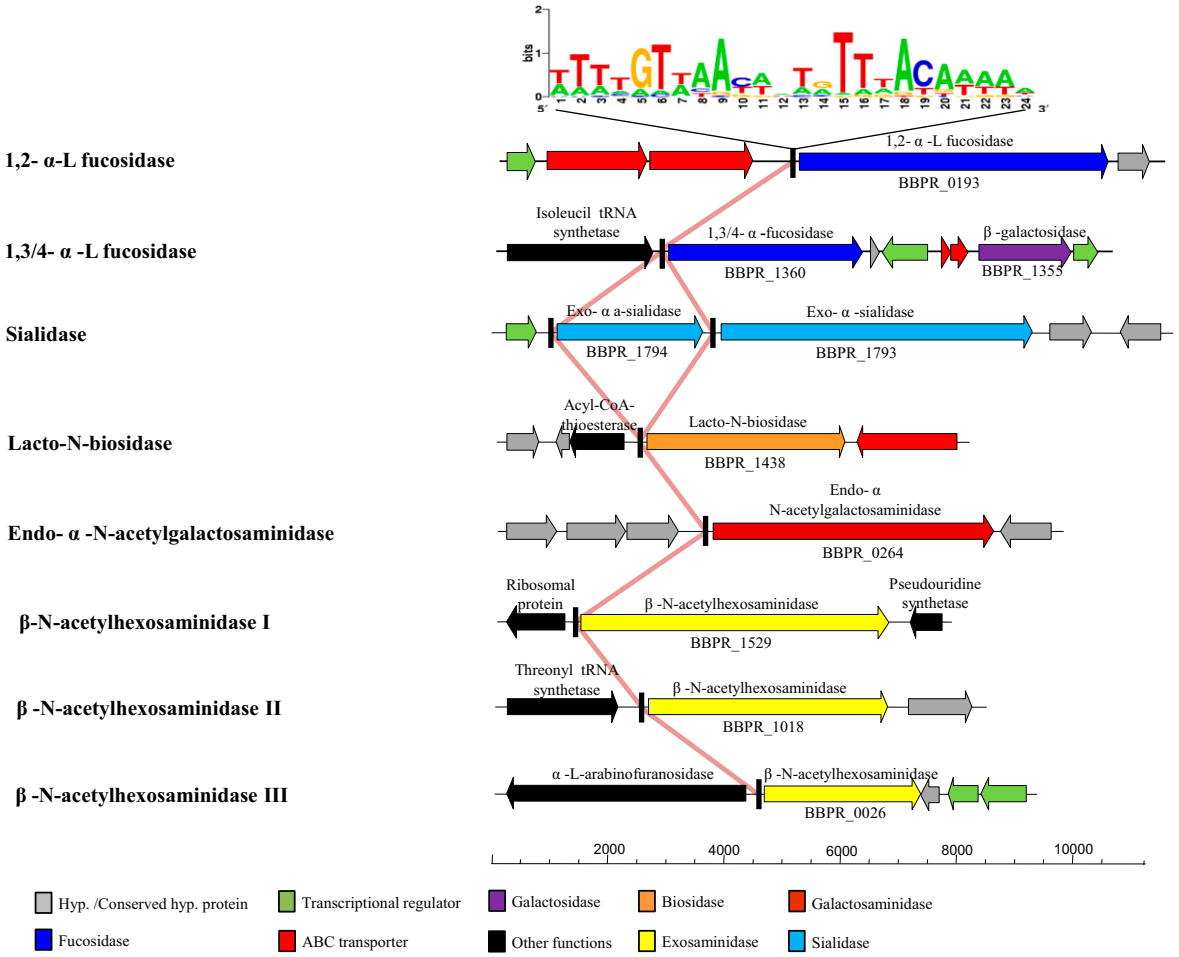

Fig. S6. Additional mucin-related gene clusters. Genes are represented by arrows; colors indicate predicted function. The vertical bar indicates the inverted repeat (putative operator binding site) involved in the regulation of the expression of mucin-induced genes. The deduced regulatory binding sites are depicted in WebLOGO format based on the comparative sequence analysis of actual target sequences in the *B. bifidum* PRL2010 genome.

Table S1. Genome features of *B. bifidum* PRL2010

Trait	No./value
Size (Mb)	2,214,650
Guanine + cytosine content	62.67%
Number of identified ORF	1844
Assigned function	1248
Amino acid transport and metabolism	114
Carbohydrate transport and metabolism	111
Transcription	77
Translation	135
Replication, recombination and repair	148
Defense mechanisms	56
Signal transduction	29
Cell wall/membrane biogenesis	66
Posttranslational modification, protein turnover, chaperones	46
Energy production and conversion	49
Nucleotide transport and metabolism	48
Coenzyme transport and metabolism	37
Lipid transport and metabolism	42
Inorganic ion transport and metabolism	49
RNA processing and modification	2
Cell cycle control, cell division, chromosome partitioning	20
Cell motility	2
Intracellular trafficking, secretion, and vesicular transport	9
Secondary metabolite biosynthesis, transport and catabolism	3
Phage region	1
IS transposase families	9
ISL3	6
IS3/IS911	2
ISL30	2
IS1557	4
ISSdy1	1
IS204/IS1001/IS1096/IS1165	1
IS3520	1
IS21	4
IS3521	1

Table S2. Proteins encoded by *B. bifidum* PRL2010 involved in the metabolism of the major core types of mucin O-glycans

ORF	Homology	Domain	Similarity (%)	E-value	Accession no.
BBPR_0064	Nucleoside-diphosphate-sugar epimerase, *Bifidobacterium adolescentis* ATCC 15703	GalE	94	4e-135	YP_910369
BBPR_0193	1,2-A-L-fucosidase, *Bifidobacterium bifidum*	—	100	0	2EAB_A
BBPR_0217	UDP-*N*-acetylglucosamine transferase, *Streptomyces lividans*	MurA	98	5e-161	BAA85335.1
BBPR_0233	Lacto-*N*-biose phosphorylase, *Bifidobacterium bifidum* JCM 1254	Lact_bio_phlase (pfam09508)	100	0	BAE95374.1
BBPR_0239	PTS system, glucose subfamily, IIA subunit, *Bifidobacterium longum* subsp. *infantis* ATCC 15697	PTS_EIIA_1 (pfam00358)	95	8e-49	YP_002323902
BBPR_0240	PTS system, *N*-acetylglucosamine-specific IIBC subunit, *Bifidobacterium longum* subsp. *infantis* ATCC 15697	PTS_EIIA_1 (pfam00358)	97	0	YP_002323903
BBPR_0264	Endo-alpha-*N*-acetylgalactosaminidase, *Bifidobacterium longum*	—	97	0	2ZXQ_A
BBPR_0441	Phosphoglucosamine mutase, *Bifidobacterium longum* subsp. *infantis* ATCC 15697	GlmM	100	0	YP_002323476
BBPR_0482	β-galactosidase, *Bifidobacterium bifidum* JCM 1254	LacZ	99	0	ABE27152.1
BBPR_0567	*N*-acetylglucosamine-6-phosphate deacetylase, *Bifidobacterium longum* subsp. *infantis* ATCC 15697	NagA	99	0	YP_002322360
BBPR_0976	Udp-glucose pyrophosphorylase, *Leishmania*	UDPGP (pfam01704)	95	3e-72	2OEF_A
BBPR_1018	β-*N*-acetylhexosaminidase, *Bifidobacterium bifidum* JCM 1254	GH20_hexosaminidase	100	0	BAI94823.1
BBPR_1050	UDP-glucose 4-epimerase, *Bifidobacterium longum* subsp. *longum* JCM1217	GalE	84	8e-169	AB303839
BBPR_1051	Galactose-1-phosphate uridylyltransferase, *Bifidobacterium longum* subsp. *longum* JCM1217	GalT	71	9e-78	AB303839
BBPR_1055	Lacto-*N*-biose phosphorylase, *Bifidobacterium longum* subsp. *longum* JCM1217	Lact_bio_phlase (pfam09508)	75	5e-86	AB303839
BBPR_1056	Permease of ABC transporter for sugars, *Bifidobacterium longum* subsp. *longum* NCC2705	TM_PBP2	91	7e-162	NP_696791
BBPR_1057	Binding-protein–dependent transport systems inner membrane component, *Bifidobacterium longum* subsp. *infantis* ATCC 15697	TM_PBP2	89	6e-164	YP_002323615
BBPR_1058	Solute binding protein of ABC transporter for sugars, *Bifidobacterium longum* subsp. *longum* NCC2705	TM_PBP2	78	7e-108	NP_696789
BBPR_1125	Bifunctional *N*-acetylglucosamine-1phosphate uridyltransferase/glucosamine-1-phosphate acetyltransferase, *Bifidobacterium longum* subsp. *longum* NCC2705	GT2_GlmU_N_bac	100	0	NP_696139
BBPR_1360	α-L-fucosidase, *Bifidobacterium bifidum* JCM1254	Alpha_L_fucos (pfam01120)	100	0	BAH80310.1
BBPR_1456	UDP-glucose 4-epimerase, *Bifidobacterium longum* subsp. *longum* DJO1A	GalE	100	0	YP_001955779
BBPR_1459	Galactoside symporter, *Bifidobacterium animalis* subsp. *lactis* HN019	Gph	97	0	ZP_02963405
BBPR_1461	*N*-acetylglucosamine-6-phosphate deacetylase, *Bifidobacterium longum* subsp. *longum* NCC2705	NagA	64	1e-39	NP_696508
BBPR_1503	alpha-*N*-acetylglucosaminidase, *Bacteroides*	NAGLU (pfam05089)	33	4e-120	ZP_06093587
BBPR_1514	β-*N*-acetylglucosaminidase, *Clostridium paraputrificum*	NAGidase (pfam07555)	75	0	BAC99989.1
BBPR_1529	β-*N*-acetylhexosaminidase, *Clostridium perfringens*	GH20_hexosaminidase	68	1e-92	ZP_02638134
BBPR_1578	UDP-*N*-acetylmuramate dehydrogenase, *Bifidobacterium longum* subsp. *infantis* ATCC 15697	MurB_C (pfam02873)	99	3e-157	YP_002323691
BBPR_1793	Sialidase, *Arthrobacter ureafaciens*	Sialidase (pfam02973)	46	2e-108	BAD66680.2
BBPR_1794	Sialidase, *Actinomyces viscosus*	Sialidase	66	7e-98	AAA21932.1

ATCC, American Type Culture Collection.

Table S3. Selected genes of *B. bifidum* PRL2010 exhibiting increased expression as determined by proteome analysis upon cultivation in different sugar-based media relative to growth in lactose

		Lactose			HMO				Mucin			
ORF	Product	*P* (%)	SpC	NSAF	*P* (%)	SpC	NSAF	[e]Δ(fold)	*P* (%)	SpC	NSAF	[ε]Δ(fold)
BBPR_1733	Phosphocarrier protein HPr				100	12	4.26	IND	100	19	6.58	IND
BBPR_1693	*amt*P ammonium transporter	100	5	0.34				REP	100	3	0.21	−0.9 ± 0.5
BBPR_1692	*gln*B nitrogen regulatory protein P-II	100	4	1.05				REP	100	4	1.06	−0.2 ± 0.3
BBPR_1681	*pyr*E1 orotate phosphoribosyltransferase	100	3	0.37				REP	100	9	1.12	1.7 ± 0.5
BBPR_1514	β-*N*-acetylhexosaminidase	100	17	0.26	100	20	0.31	0.1 ± 0.3	100	133	2.04	6.0 ± 1.3
BBPR_1508	*pts*G PTS system, glucose-specific IIABC component	100	6	0.26	100	6	0.26	−0.1 ± 0.3	100	11	0.47	0.7 ± 0.3
BBPR_1454	*fuc*O lactaldehyde reductase	100	17	1.31	100	22	1.75	0.3 ± 0.3	100	39	3.03	1.1 ± 0.4
BBPR_1438	Lacto-*N*-biosidase								100	25	0.67	IND
BBPR_1437	ATP-binding protein of ABC transporter system	100	17	0.90	100	21	1.15	0.2 ± 0.3	100	15	0.80	−0.3 ± 0.3
BBPR_1403	Aldose 1-epimerase family protein	100	6	0.61	100	12	1.27	1,0 ± 0.5	100	14	1.45	1.1 ± 0.4
BBPR_1377	map methionine aminopeptidase	100	12	1.37	100	7	0.83	−0.8 ± 0.5	100	20	2.31	0.5 ± 0.3
BBPR_1161	*arg*C *N*-acetyl-gamma-glutamyl-phosphate reductase	100	5	0.41	100	11	0.94	1.2 ± 0.5	100	5	0.42	−0.2 ± 0.3
BBPR_1160	*arg*J glutamate *N*-acetyltransferase	100	3	0.23	100	11	0.86	2.6 ± 0.8	100	9	0.69	1.7 ± 0.5
BBPR_1159	*arg*B acetylglutamate kinase	100	11	1.02	100	12	1.15	0.1 ± 0.3	100	9	0.84	−0.4 ± 0.3
BBPR_1058	Solute-binding protein	100	107	7.25	100	247	17.32	1.3 ± 0.5	100	381	26.08	2.2 ± 0.6
BBPR_1057	Solute-binding protein								100	4	0.35	IND
BBPR_1056	Solute-binding protein	99	2	0.19	100	8	0.77	3.0 ± 0.9	100	6	0.57	1.7 ± 0.5
BBPR_1055	*lnb*P lacto-*N*-biose phorylase	100	42	1.65	100	46	1.87	0.1 ± 0.3	100	74	2.93	0.6 ± 0.3
BBPR_1054	Sugar kinase, ROK family	100	12	1.13	100	8	0.78	−0.6 ± 0.5	100	18	1.71	0.3 ± 0.3
BBPR_1052	*N*-acetylhesosammine-1-kinase	99	2	0.16	100	10	0.84	4.0 ± 1.1	100	64	5.26	27.8 ± 5.5
BBPR_1051	*gal*T2 galactose-1-phosphate uridylyltransferase				99	2	0.12	IND	100	19	1.12	IND
BBPR_1050	*gal*E UDP-glucose 4-epimerase	100	36	3.12	100	51	4.58	0.4 ± 0.3	100	60	5.25	0.5 ± 0.3
BBPR_1025	Fructosamine kinase family protein								99	2	0.23	IND
BBPR_1011	YajC protein translocase subunit								100	3	0.63	IND
BBPR_1010	apt adenine phosphoribosyltransferase								100	3	0.45	IND
BBPR_1009	*suc*C succinyl-CoA synthetase beta chain	100	15	1.10	100	8	0.61	−1.0 ± 0.6	100	5	0.37	−2.5 ± 0.8
BBPR_1008	*suc*D succinyl-CoA synthetase alpha chain	100	11	1.05	100	7	0.69	−0.7 ± 0.5	99	2	0.19	−5.4 ± 1.5
BBPR_0940	*pyr*F1 orotidine 5′-phosphate decarboxylase				100	4	0.38	IND	100	6	0.56	IND
BBPR_0939	*pyr*C Dihydroorotase								100	4	0.25	IND
BBPR_0806	*aro*B Shikimate kinase/3-dehydroquinate synthase	100	5	0.28	100	4	0.23	−0.3 ± 0.4	100	6	0.34	0.1 ± 0.3
BBPR_0736	CarD-like transcriptional regulator	99	2	0.30				REP	100	5	0.75	1.3 ± 0.4
BBPR_0735	*glg*B 1,4-alpha-glucan branching enzyme	99	4	0.16	100	9	0.37	1.2 ± 0.5	100	13	0.52	1.9 ± 0.6
BBPR_0734	Two-component response regulator	100	20	2.63	100	24	3.27	0.2 ± 0.3	100	22	2.92	0.0 ± 0.3
BBPR_0714	Exopolyphosphatase	99	2	0.18	99	2	0.18	−0.1 ± 0.3	100	6	0.54	1.7 ± 0.5
BBPR_0579	DegT/DnrJ/EryC1/StrS aminotransferase								100	4	0.30	IND
BBPR_0578	Conserved hypothetical protein	100	4	0.29	100	6	0.45	0.5 ± 0.4	99	2	0.14	−1.3 ± 0.6
BBPR_0567	*nag*A2 *N*-acetylglucosamine-6-phosphate deacetylase	100	8	0.56	100	19	1.39	1.4 ± 0.5	100	29	2.07	2.3 ± 0.6
BBPR_0566	*nag*B3 glucosamine-6-phosphate isomerase	100	17	1.80	100	48	5.27	1.8 ± 0.6	100	69	7.39	2.7 ± 0.7
BBPR_0460	ABC-type xylose transport system	99	3	0.07				REP	100	21	0.51	5.3 ± 1.2
BBPR_0291	Aminotransferase				100	6	0.46	IND	100	3	0.23	IND
BBPR_0265	prsA ribose-phosphate pyrophosphokinase	100	11	0.96	100	5	0.45	−1.4 ± 0.7	100	5	0.44	−1.5 ± 0.6
BBPR_0264	endo-alpha-*N*-acetylgalactosaminidase	100	78	1.23	100	63	1.03	−0.3 ± 0.4	100	193	3.07	1.2 ± 0.4
BBPR_0261	Glycerol-3-phosphate dehydrogenase [NAD(P)+]	100	4	0.36	100	18	1.66	3.5 ± 1.0	100	9	0.81	1.0 ± 0.4
BBPR_0241	*thr*A homoserine dehydrogenase	100	25	1.69	100	24	1.68	−0.1 ± 0.3	100	25	1.70	0.0 ± 0.3
BBPR_0240	PTS system, *N*-acetylglucosamine	100	78	4.46	100	133	7.86	0.7 ± 0.4	100	180	10.39	1.1 ± 0.4
BBPR_0239	PTS system, glucose	100	14	2.66	100	26	5.12	0.8 ± 0.4	100	40	7.68	1.6 ± 0.5
BBPR_0234	*psp*1 Translation initiation inhibitor								100	3	0.69	IND
BBPR_0233	*lnb*P lacto-*N*-biose phorylase	100	43	1.69	100	68	2.77	0.6 ± 0.4	100	63	2.50	0.3 ± 0.3
BBPR_0194	Phospholipids-binding protein	100	4	0.64	100	5	0.83	0.2 ± 0.3	100	7	1.14	0.6 ± 0.3
BBPR_0193	1,2-A-L-fucosidase								99	3	0.05	IND
BBPR_0092	*glg*P1 glycogen phosphorylase	100	88	3.18	100	76	2.84	−0.2 ± 0.4	100	73	2.66	−0.4 ± 0.3
BBPR_0032	Ascorbate-specific PTS system enzyme IIC	99	2	0.12				REP	100	7	0.41	2.2 ± 0.6
BBPR_0030	PTS system, lactose/cellobiose specific IIB subunit								100	6	1.20	IND
BBPR_0026	*nag*Z beta-*N*-acetylhexosaminidase	100	11	0.41	100	19	0.73	0.7 ± 0.4	100	14	0.53	0.1 ± 0.3

The gray-shaded gene names represent genes whose corresponding proteins elicited increased expression in *B. bifidum* PRL2010 when cultivated in mucin vs. lactose and which were not identified by the transcriptomics approach (Table S4). P (%), protein identification probability calculated by Prophate; SpC, number of spectra used for the protein identification; NSAF, normalized spectral abundance factor; IND, induction; REP, repression.

Table S4. Selected genes differentially transcribed upon *B. bifidum* PRL2010 growth in mucin-based media relative to growth in lactose

ORF	Product	Change (fold)[†]	Motif[‡]
BBPR_1824	ATP binding protein of ABC transporter	10.7 (4,34E-08)	–
BBPR_1794	Exo-α-sialidase	110.6 (4,77E-15)	+
BBPR_1793	Exo-α-sialidase	23.8 (7,99E-14)	+
BBPR_1733	Phosphocarrier protein HPr	6.1 (2,40E-05)	–
BBPR_1717	PTS system, cellobiose-specific component IIC	7.3 (6,17E-12)	–
BBPR_1596	Conserved hypothetical protein	36.1 (0)	+
BBPR_1563	Adh2; aldehyde-alcohol dehydrogenase 2	8.3 (6,56E-11)	+
BBPR_1529	NagZ; β-*N*-acetylhexosaminidase	67 (1,45E-10)	+
BBPR_1514	β-*N*-acetylhexosaminidase	48.7 (1,36E-07)	+
BBPR_1513	Pyridine nucleotide-disulphide oxidoreductase family protein	5.8 (3,01E-09)	+
BBPR_1503	α-*N*-acetylglucosaminidase family protein	29.7 (2,73E-12)	+
BBPR_1479	Universal stress protein family	6 (1,13E-13)	–
BBPR_1454	FucO; lactaldehyde reductase	5.3 (2,45E-10)	+
BBPR_1438	lacto-N-biosidase	68.8 (6,22E-08)	+
BBPR_1378	Conserved hypothetical membrane spanning protein	6.4 (2,88E-15)	–
BBPR_1368	Nox; NADH oxidase H_2O-forming	28.3 (1,44E-07)	–
BBPR_1360	1,3/4-α-L-fucosidase	99.9 (5,76E-11)	+
BBPR_1310	ImpB/MucB/SamB family protein involved in DNA repair	9.3 (2,58E-10)	–
BBPR_1300	Glycosyl hydrolase family protein	68.2 (7,58E-11)	+
BBPR_1235	Pfl; Formate acetyltransferase	11.3 (6,18E-07)	+
BBPR_1058	Solute-binding protein of ABC transporter system for sugars	52.2 (3,12E-06)	+
BBPR_1057	Permease	6.5 (4,45E-06)	–
BBPR_1056	Permease	5.9 (7,15E-07)	–
BBPR_1055	LnbP; lacto-*N*-biose phosphorylase	6 (2,42E-09)	–
BBPR_1052	NahK; *N*-acetylhexosamine-1-kinase	26 (4,12E-10)	+
BBPR_1051	GalT2; galactose-1-phosphate uridylyltransferase (Leloir pathway)	25.8 (9,70E-11)	+
BBPR_1018	NagZ; β-*N*-acetylhexosaminidase	19 (1,39E-12)	+
BBPR_0985	Conserved hypothetical protein	163.3 (3,13E-12)	–
BBPR_0984	Conserved hypothetical protein with helix–turn–helix motif	169.8 (1,75E-13)	–
BBPR_0829	PntB NAD(P) transhydrogenase subunit beta	7.8 (4,88E-15)	–
BBPR_0828	PntA NAD(P) transhydrogenase alpha subunit	8 (1,48E-11)	–
BBPR_0827	PntA1 NAD(P) transhydrogenase subunit alpha part 1	8.3 (3,55E-15)	–
BBPR_0719	Hypothetical protein	8.5 (2,55E-09)	–
BBPR_0567	NagA; *N*-acetylglucosamine-6-phosphate deacetylase	26 (2,63E-09)	–
BBPR_0566	NagB; glucosamine-6-phosphate isomerase	14.9 (2,36E-08)	+
BBPR_0563	*N*-acetylglucosamine repressor	10 (1,87E-12)	+
BBPR_0562	hexokinase	7.3 (3,35E-10)	–
BBPR_0561	L-fucose permease	7.5 (2,45E-10)	–
BBPR_0517	Linoleic acid hydratase	19.3 (0)	–
BBPR_0482	LacZ; β-galactosidase	73.3 (2,16E-09)	+
BBPR_0479	Narrowly conserved hypothetical membrane spanning protein	5.8 (5,69E-07)	–
BBPR_0460	Conserved hypothetical protein	84.7 (2,71E-12)	+
BBPR_0337	Conserved hypothetical membrane	17.5 (3,10E-10)	+
BBPR_0264	Endo-α-*N*-acetylgalactosaminidase, cell wall anchor protein family	10.6 (7,75E-07)	+
BBPR_0228	Transcriptional regulator, LacI family	10.5 (5,30E-10)	+
BBPR_0193	1,2-α-L-fucosidase	80.5 (2,47E-09)	+
BBPR_0167	Hypothetical protein	20.4 (9,91E-06)	–
BBPR_0134	Hypothetical protein	7.1 (9,53E-09)	–
BBPR_0132	Hypothetical protein	5.7 (1,27E-09)	–
BBPR_0122	Conserved hypothetical protein with DUF74 domain	6.1 (1,50E-08)	–
BBPR_0054	IS3521 family transposase	6.8 (2,07E-08)	–
BBPR_0049	VirS two-component sensor kinase	8.2 (1,11E-15)	–
BBPR_0032	Ascorbate-specific PTS system enzyme IIC	101.6 (2,30E-09)	–
BBPR_0031	PTS system, lactose/cellobiose-specific IIB component	79.6 (9,80E-08)	–
BBPR_0030	PTS system, lactose/cellobiose specific IIA subunit	28.3 (4,09E-11)	+
BBPR_0025	AbfA alpha L-arabinofuranosidase	53.3 (7,38E-13)	+

AAAN in the region upstream of the gene. Genes that are grouped in bold are predicted to form an operon in which only the first gene contains the upstream-regulated element. The gray-shaded gene names represent genes that elicited increased transcription in *B. bifidum* PRL2010 when cultivated in mucin vs. lactose and which that not identified by the proteomics approach (Table S3).

[†]Genes up-regulated in *B. bifidum* PRL2010 cells grown in MRS plus mucin as compared with growth on lactose. Values in parenthesis indicate *P* value.

[‡]Presence (+) or absence (–) of the putative regulatory element TTTTGTNAANNNNNTTNACA.