

Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor

Mary O'Connell Motherway^{a,b,1,2}, Aldert Zomer^{a,1}, Sinead C. Leahy^{a,b,1}, Justus Reunanen^c, Francesca Bottacini^{a,b,d}, Marcus J. Claesson^a, Frances O'Brien^a, Kiera Flynn^a, Patrick G. Casey^a, Jose Antonio Moreno Munoz^a, Breda Kearney^a, Aileen M. Houston^{a,e}, Caitlin O'Mahony^a, Des G. Higgins^f, Fergus Shanahan^{a,e}, Airi Palva^c, Willem M. de Vos^{c,g}, Gerald F. Fitzgerald^{a,b,h}, Marco Ventura^d, Paul W. O'Toole^{a,b}, and Douwe van Sinderen^{a,b,2}

^aAlimentary Pharmabiotic Centre and Departments of ^bMicrobiology, ^cMedicine, and ^dFood and Nutritional Sciences, National University of Ireland, Cork, Ireland; ^eLaboratory of Probiogenomics, Department of Genetics, Biology of Microorganisms, Anthropology, and Evolution, University of Parma, 43100 Parma, Italy; ^fDivision of Microbiology and Epidemiology, Department of Basic Veterinary Medicine, University of Helsinki, FIN-00014, Helsinki, Finland; ^gConway Institute, University College Dublin, Belfield, Dublin 4, Ireland; and ^hLaboratory of Microbiology, Wageningen University, 6703 HB, Wageningen, The Netherlands

Edited* by Todd R. Klaenhammer, North Carolina State University, Raleigh, NC, and approved May 25, 2011 (received for review April 7, 2011)

Development of the human gut microbiota commences at birth, with bifidobacteria being among the first colonizers of the sterile newborn gastrointestinal tract. To date, the genetic basis of *Bifidobacterium* colonization and persistence remains poorly understood. Transcriptome analysis of the *Bifidobacterium breve* UCC2003 2.42-Mb genome in a murine colonization model revealed differential expression of a type IVb tight adherence (Tad) pilus-encoding gene cluster designated "tad₂₀₀₃." Mutational analysis demonstrated that the tad₂₀₀₃ gene cluster is essential for efficient in vivo murine gut colonization, and immunogold transmission electron microscopy confirmed the presence of Tad pili at the poles of *B. breve* UCC2003 cells. Conservation of the Tad pilus-encoding locus among other *B. breve* strains and among sequenced *Bifidobacterium* genomes supports the notion of a ubiquitous pili-mediated host colonization and persistence mechanism for bifidobacteria.

fimbriae | probiotic | prebiotic | genomics

The mammalian gastrointestinal tract (GIT) harbors a complex community of microorganisms, also referred to as the "intestinal microbiota" (1, 2). Although it is well recognized that the intestinal microbiota has a profound influence on health and disease, knowledge on the precise mechanism(s) by which this microbiota exerts such influence remains largely unknown. Bacterial colonization of the GIT of a newborn commences at birth and depends on many factors that include gestational age, mode of delivery, local environment, type of feeding, and antibiotic treatment (3, 4). *Bifidobacterium* species are among the first colonizers of the essentially sterile GIT of newborns and constitute one of the dominant genera of the microbiota of healthy breast-fed infants, although they become less abundant in weaned infants and adults (5, 6).

GIT colonization by microbes is believed to play an essential part in metabolism and energy balance, in resisting pathogen colonization, in the maturation of the intestine, and in the education of the immune system (7–10). The specific contribution of members of the *Bifidobacterium* genus to such activities is subject to much investigation and speculation (11), although it is widely accepted that their presence in the GIT may confer health benefits (12–14). The global analysis of the bifidobacterial gene expression in babies indicated the activities include the production of folic acid and exopolysaccharides (EPS) (15). Although the latter property could be involved in persistence, the details of the molecular mechanisms that allow bifidobacteria to colonize the GIT have not yet been elucidated.

Recently, Kankainen et al. (16) showed that adherence and colonization of *Lactobacillus rhamnosus* GG is mediated by sortase-dependent pili. In Gram-positive bacteria, these pili are the most frequently described nonflagellar, proteinaceous, multimeric surface appendages that are involved in adhesion to

other bacteria, to host cells, or to environmental surfaces (17, 18). Gram-positive pili are formed by a mechanism involving specific transpeptidases, called "sortases," which cross-link individual pilin monomers and join the resulting covalent polymer to the cell wall peptidoglycan (19). Recently, the Gram-positive pathogen *Mycobacterium tuberculosis* was found to express type IVb pili similar to those specified by the Gram-negative tight adherence (*tad*) locus (20). Homologs of *tad* genes have been identified in the genomes of *Corynebacterium diphtheria*, *Thermobifida fusca*, and *Streptomyces coelicolor* (21, 22). The Tad pili biosynthesis apparatus was first described in *Actinobacillus actinomycetemcomitans* and allows the production and assembly of pili, which in this bacterium mediate adhesion to surfaces and are essential for colonization and pathogenesis (23, 24). Because of their simple cell envelope structure, Gram-positive *tad* loci encode fewer pilus assembly functions than do their Gram-negative counterparts, namely TadZ, TadV, TadC, TadB, and TadA (21). TadA, an ATPase localized at the periphery of the cytoplasmic membrane, is believed to energize pilus assembly; TadB and TadC represent integral membrane proteins that form homo- or hetero-oligomeric structures in association with TadA to constitute the secretion apparatus of the pilus subunit. Based on similarity to the septum site-determining protein MinD, TadZ is thought to direct the Tad secretion and assembly apparatus to the cell poles. The Tad pilus is comprised of homopolymers of a single pilin subunit, although some pili possess an adhesive subunit at the pilus tip or can be decorated with pseudopilins along the pilus (21, 22). The pilin and pseudopilins are synthesized as precursors (prepilins) with a hydrophobic leader peptide terminating with a glycine residue that is removed by a dedicated peptidase (TadV).

To enhance our understanding of bifidobacteria and the role they play in the intestinal ecosystem, we determined the genome

Author contributions: M.O.M., A.Z., S.C.L., J.R., D.G.H., F.S., W.M.d.v., G.F.F., and D.v.S. designed research; M.O.M., A.Z., S.C.L., J.R., F.O., K.F., P.G.C., J.A.M.M., and B.K. performed research; A.M.H., C.O., and A.P. contributed new reagents/analytic tools; M.O.M., A.Z., S.C.L., J.R., F.B., M.J.C., F.S., W.M.d.v., M.V., P.W.O., and D.v.S. analyzed data; and M.O.M., A.Z., J.R., F.S., W.M.d.v., G.F.F., M.V., P.W.O., and D.v.S. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The sequence reported in this article has been deposited in the GenBank database (accession no. CP000303). The array data for the comparative genome hybridization and transcription studies have been deposited in the GEO database (accession no. GSE27491).

¹M.O.M., A.Z., and S.L. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: m.oconnellmotherway@ucc.ie or d.vansinderen@ucc.ie.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1105380108/-DCSupplemental.

sequence of a *Bifidobacterium breve* strain isolated from a nursing stool. Analysis of this genome sequence coupled with *in vivo* transcriptome studies identified a *tad* locus essential for efficient host colonization of the murine intestine. This locus is conserved in all *B. breve* strains tested and in currently available *Bifidobacterium* genomes, so the elucidation of its function represents a significant step toward understanding the molecular details of the interaction between bifidobacteria and their host.

Results and Discussion

***B. breve* UCC2003 Genome.** The 2,422,684-bp genome of *B. breve* UCC2003 (Fig. 1) has a guanine-cytosine (G+C) content of 58.7%, similar to that of currently available *Bifidobacterium* genomes (Table S1), and is predicted to harbor 1,985 genes, of which 1,854 are protein-coding sequences with an average length of 1,099 bp. The genome contains two rRNA-encoding operons and 54 tRNA genes. A functional role or general function was assignable for 1,597 (86%) of the protein-coding sequences, 250 (14%) protein-coding sequences represent conserved genes for which no functional role could be assigned, and only seven protein-coding sequences lack similarity to entries in the public databases. Furthermore, 71 pseudogenes were detected.

A comparison of the general genome characteristics of a selection of relevant and completed *Bifidobacterium* genomes is provided in Table S1. Dot plot alignments of various representative *Bifidobacterium* genomes revealed a varying degree of colinearity (Fig. S1), which correlated with the phylogenetic distance between the compared genomes (25) and which frequently generated an X-shaped plot, suggestive of multiple rearrangements around the central axis following divergence

from a common ancestor (26). The UCC2003 genome contains a single prophage-like remnant and one CRISPR (clustered regularly interspaced short palindromic repeats) locus, which have been described previously (27–29). Extracellular proteins and structures are considered to be important for interaction of the bacterium with its environment, for example in adherence, degradation of complex carbohydrates, and microbe–microbe/microbe–host communication. *In silico* prediction of the bacterial secretome enables functional characterization of possible adhesins, as shown by the identification of sortase-anchored adhesins in *Lactobacillus salivarius* (30). The predicted *B. breve* UCC2003 secretome (Fig. S2) consists of 216 proteins with an Sec-type signal peptide (31), of which 41 contain an N-terminal lipoprotein motif for anchoring to the cell membrane (32) and 10 contain a C-terminal LPxTG-like motif for sortase-dependent anchoring to peptidoglycan or pilus biosynthesis (33–35). Examples of the latter category include seven genes that are predicted to specify sortase-dependent pilin subunits, organized into three gene clusters (see below), and two extracellular glycosidases, ApuB and GalA, which function in the hydrolysis of particular glucose- and galactose-containing polysaccharides, respectively (36, 37). Furthermore, *B. breve* UCC2003 encodes a secreted serine protease inhibitor, also known as a “serpin” (Bbr_1320), as previously described for *Bifidobacterium longum* subsp. *longum* NCC2705 (38) and *B. breve* 210B (39). The putative *B. breve* serpin thus may contribute to host interactions in the GIT as proposed for *B. longum* subsp. *longum* NCC2705 (38).

Two regions in the genome of *B. breve* UCC2003, Bbr_0430–Bbr_0474 and Bbr_1786–Bbr_1803, contain genes that are predicted to encode proteins involved in the production of exo- or capsular polysaccharides, which we refer to collectively as EPS. Interestingly, the Bbr_0430–Bbr_0474 region has a G+C DNA content (47.4%) markedly lower than the *B. breve* average (58.73%), perhaps because it was acquired by horizontal DNA transfer.

Gene clusters specifying pili are of particular interest in the case of *Bifidobacterium*, because such proteins may represent candidates mediating attachment to mucus or cells in the intestine (40, 16). Three different gene clusters predicted to encode sortase-dependent pili are present in the genome of UCC2003, although the first gene in two of these clusters, Bbr_0113 and Bbr_1889, respectively, appears to be a pseudogene because of a frameshift in a stretch of 11 or 10 guanine nucleotides. Both these gene clusters include a dedicated sortase gene that differs from the housekeeping sortase gene, suggesting that these pili may be produced if the effect of the homopolymer frameshifts is suppressed. Interestingly, a putative type IVb pilus-encoding gene cluster, similar to the *tad* locus in *A. actinomycetemcomitans*, was identified on the genome of *B. breve* UCC2003 (Bbr_0132–Bbr_0138, referred to here as “*tad*₂₀₀₃”) (Fig. 2 and Fig. S3), whose function was investigated further as described below. Characterized *tad* loci are responsible for the assembly of pili, which are essential for colonization of several pathogenic bacteria (21). Homologs of the genes that make up the proposed *B. breve* UCC2003 *tad*₂₀₀₃ locus are present in all examined *B. breve* strains [as determined by comparative genome hybridization (CGH); see below], as well as in all publicly available bifidobacterial genomes (Fig. S4), and they all appear to be organized as two adjacent gene clusters. The first of these clusters (Bbr_0132–Bbr_0135 in UCC2003) encodes proteins dedicated toward pilus assembly, location, and export; the second cluster is predicted to encode the Flp prepilin (Bbr_0136) and two pseudopilins, TadE (Bbr_0137) and TadF (Bbr_0138) (Fig. 2). Consistent with their processing requirement, a putative prepilin peptidase (TadV) cleavage site, GAATAEY/F (21), is readily identifiable in Flp (Fig. S5) and TadE, and a less obvious potential cleavage sequence (GVALSAT) is present in TadF. Based on sequence similarity, the putative UCC2003 TadV prepilin peptidase is specified by Bbr_0901, although this gene is not linked to the *tad*₂₀₀₃ locus. The TadV peptidase is believed to hydrolyze the leader peptide of its targets at the C-terminal end of the glycine within the cleavage site, whereas the conserved glutamate residue at position five in the

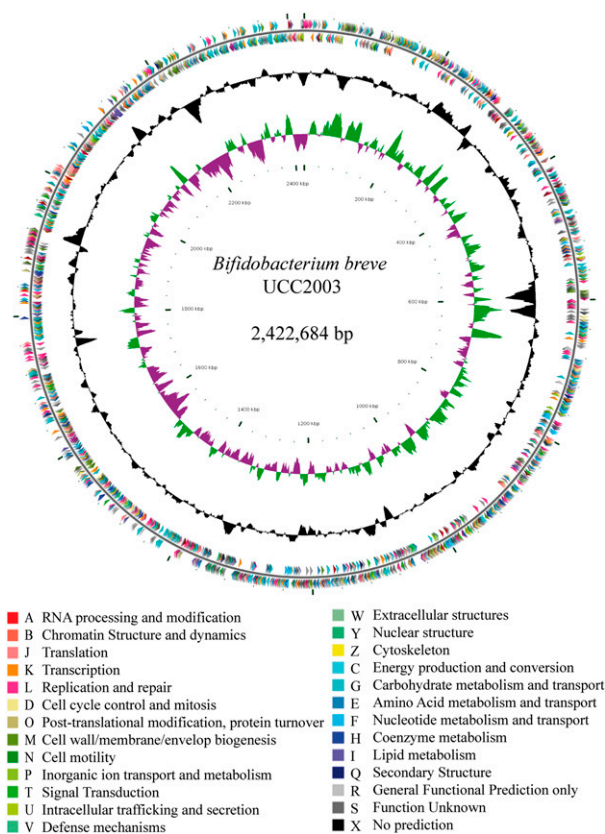


Fig. 1. Circular genome map of *B. breve* UCC2003. The innermost circle illustrates GC skew, shown in green on the forward strand and in purple on the reverse strand. The middle circle highlights G+C deviation from the mean of 58.7%. The outer circle displays the ORF distribution by strand with color corresponding to the Cluster of Orthologous Genes functional assignment.

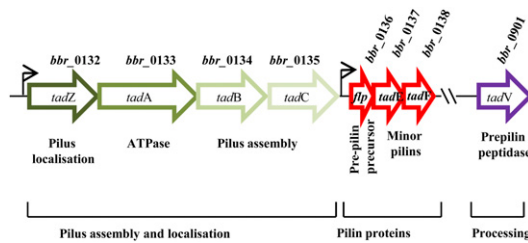


Fig. 2. Schematic representation of the *tad* locus from *B. breve* UCC2003. Each arrow represents an ORF with the gene number given above the arrow and the gene name given within the arrow. The functions of the encoded proteins are indicated below the arrow. The positions of the predicted promoters are indicated by arrowheads.

mature pilin protein is thought to be essential for the recognition of incoming pilin subunits during pilus assembly (21).

Comparative Genome Hybridization Analysis of *B. breve* Strains. To investigate genome variability in 18 *B. breve* strains obtained from various sources and *B. breve* UCC2003, we performed CGH experiments using an array designed on 1,864 of the 1,985 annotated genes of *B. breve* UCC2003. Overall, based on the criteria applied (*Materials and Methods*), 1,393 genes were shown to be present in all *B. breve* strains tested. Variably present genes were classified into three groups that were predicted to be (i) phage related or encode phage resistance, (ii) involved in biosynthesis of extracellular structures, or (iii) involved in carbon source metabolism (Fig. 3A). For the first group, variations are observed for genes in the CRISPR locus (Bbr_1405–Bbr_1411), restriction/modification systems (Bbr_0214–Bbr_0216 and Bbr_1118–Bbr_1121) (41), a cluster containing a putative abortive infection protein (Bbr_1050–Bbr_1053), and a (partial) prophage (Bbr_0291–Bbr_0321) (28). Variation in the second group, genes involved in biosynthesis of extracellular structures, is shown for the two EPS regions (Bbr_0430–Bbr_0474 and Bbr_1786–Bbr_1803, mentioned above) and the sortase-dependent pili-encoding genes Bbr_1889–Bbr_1887. No variation, however, was observed for the sortase-dependent pili-encoding genes Bbr_0113–Bbr_0115 and Bbr_0365–Bbr_0366, the Tad-like type IVb pili-encoding system (Bbr_0132–Bbr_0138) including *tadV* (Bbr_0901), or the serpin-encoding gene (Bbr_1320). Finally, variability was observed in the third group of genes, those involved in the utilization of starch (42) and galactan (37), ribose (43), sucrose isomers (44), cellodextrin (Fig. S6) (45), and raffinose (46). Differences between *B. breve* strains as determined by CGH were used to infer the phylogenetic relationships among the tested *B. breve* strains using CGHdist (47). *B. longum* was used as an outgroup using BLAST-derived data (cutoff $1E^{-20}$). Six different clusters were produced, with a limited diversity between the *B. breve* strains, thus supporting the notion of a closed pangenome architecture of the *B. breve* species (Fig. 3A).

***B. breve* UCC2003 Transcriptome in a Murine Colonization Model.** Previously, Cronin et al. (48, 49) demonstrated that *B. breve* UCC2003 can colonize the murine GIT and persist for at least 30 d. To determine which *B. breve* UCC2003 genes are differentially transcribed in a murine GIT relative to its transcriptional profile under laboratory growth conditions, total bacterial RNA was isolated from ceca of Balb/c mice that had been stably colonized by *B. breve* UCC2003. The RNA was amplified and used to determine the in vivo transcriptome compared with that of an exponential-phase culture in de Man, Rogosa, Sharpe (MRS) medium supplemented with glucose (*Materials and Methods*). A total of 105 genes was significantly up-regulated in vivo, and 102 genes were down-regulated (\geq fivefold, $P < 0.0001$) in the murine-derived *B. breve* UCC2003 transcriptome relative to the control. The most highly up-regulated genes include the prophage locus, a locus with genes encoding a protein with a fibronectin III and

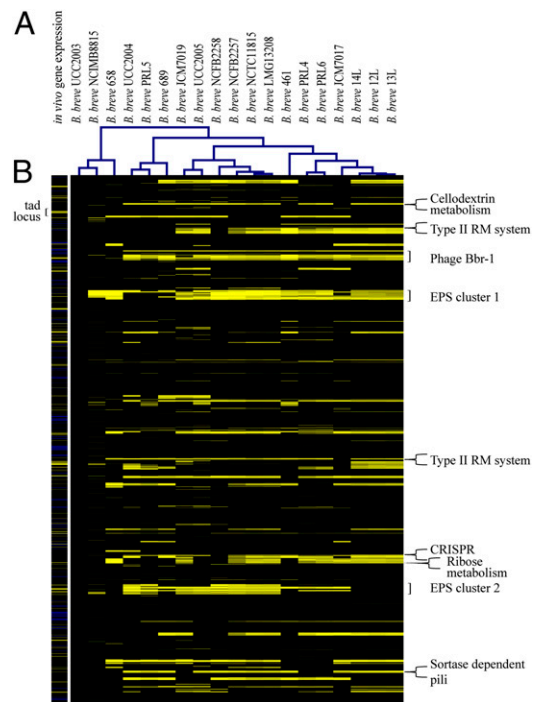


Fig. 3. Genomic diversity within *B. breve* species compared with *B. breve* UCC2003 and *B. breve* UCC2003 in vivo versus in vitro gene expression. (A) Each horizontal row corresponds to a gene on the array, and genes are ordered according to their position on the UCC2003 genome (with *dnaA* taken as the first gene). Each column represents an analyzed strain. The color code, which goes from black to yellow to indicate the presence, divergence, or absence of a gene sequence. The predicted function of some relevant genes is shown in the right margin of the figure. The visual output of a CGH-based clustering analysis for these *B. breve* strains is shown at the top of this CGH heatmap. (B) Differential transcriptional patterns of *B. breve* UCC2003 following mouse cecum colonization compared with UCC2003 grown in MRS medium supplemented with glucose. Tested genes are in the same order as in the CGH pattern. Yellow and blue colors represent *B. breve* genes that exhibit an increased or decreased transcription level, respectively, in the cecum of *B. breve*-colonized mice compared with growth in the rich medium.

cadherin domains (50, 51), and genes encompassing the *tad*₂₀₀₃ locus (Fig. 3B). Other significantly up-regulated genes include a putative lipase-encoding gene, genes related to carbon metabolism, and a number of genes encoding for transport systems for unknown substrates. The up-regulation of genes related to carbon metabolism also was observed when a select number of species was used to colonize the gut of a germ-free mouse model (52, 53). This gut-specific increase in gene expression may represent specific adaptations of UCC2003 to its natural environment. For example, increased lipase expression may provide UCC2003 with fatty acids, whereas the gut-induced expression of Tad pili and the protein containing the fibronectin/cadherin domain may promote host adhesion and thus colonization (as demonstrated below for Tad). Down-regulated genes primarily encoded proteins involved in replication, transcription, and translation, suggesting that, as might be expected, *B. breve* UCC2003 divides at a slower rate in the murine GIT than in rich media. The microarray experiments were verified for the most relevant genes by quantitative RT-PCR (qRT-PCR), which confirmed that the *tad*₂₀₀₃ genes are up-regulated 25- to 62-fold in vivo (Table S2).

***tad*₂₀₀₃ Locus Is Essential for Colonization of and Persistence In Murine Gut.** To establish whether the predicted type IVb pili specified by the *tad*₂₀₀₃ locus are involved in host interactions, the *tadA*₂₀₀₃ gene was inactivated in *B. breve* UCC2003 by gene disruption (the resulting *tadA*₂₀₀₃ mutant was designated “*B. breve*

UCC2003-tadA") (*Materials and Methods*). The *tadA*₂₀₀₃ gene (Bbr_0133) homolog was selected for mutagenesis because its protein product is proven to be essential for Tad pilus assembly in *A. actinomycetemcomitans* (54). In addition, several studies on pathogenic bacterial species have demonstrated that TadA activity is crucial for Tad pilus-mediated colonization and pathogenicity (21). Furthermore, *tadA*₂₀₀₃ is the largest gene in the Tad-pilus assembly operon and therefore is the most suitable for insertional inactivation (41). Seven-week-old ($n = 5$) Balb/c mice were administered 1×10^9 cfu of *B. breve* UCC2003PK1 or *B. breve* UCC2003-tadAPK1 (*B. breve* UCC2003 or *B. breve* UCC2003-tadA harboring plasmid pPKCM1, respectively) (*Materials and Methods*) for 3 d by oral gavage. Fecal pellets were collected at intervals over 25 d to enumerate *B. breve* UCC2003PK1 or UCC2003-tadAPK1 shedding. After 25 d the animals were killed, the small intestine, cecum, and large intestines were homogenized, and bifidobacterial numbers were enumerated. In contrast to the parent strain *B. breve* UCC2003PK1, which was stably colonized at a level of $4.38 \times 10^5 \pm 6.1 \times 10^4$ cfu g⁻¹, its derivative UCC2003-tadAPK1 was shown to colonize and persist in the Balb/c mouse at a level of $5.95 \times 10^3 \pm 1 \times 10^3$ cfu g⁻¹, exhibiting an approximately 70-fold reduction in shedding level between days 8 and day 25 (Fig. 4A). This finding was substantiated further when *B. breve*

UCC2003PK1 and UCC2003-tadAPK1 numbers were determined in the small intestine, cecum, and large intestine: These numbers ranged from 10^4 – 10^5 cfu per organ for the parent strain UCC2003PK1, but relatively low numbers of the mutant UCC2003-tadAPK1, 10^2 – 10^3 cfu per organ, were detected in the cecum, and the mutant strain was undetectable in the small and large intestine (Fig. 4B).

To investigate their relative colonization ability in germ-free mice, 1×10^9 cfu of *B. breve* UCC2003PK1 or *B. breve* UCC2003-tadAPK1 were administered by oral inoculation to 8-wk-old germ-free Swiss Webster mice ($n = 10$) for 5 d. Fecal samples were collected twice weekly for the next 3 wk to enumerate bacterial numbers. Unlike the outcome in conventional Balb/c mice, the UCC2003-tadAPK1 strain was shed at high numbers by the monoassociated Swiss Webster mice, although the numbers were significantly lower than those of the WT strain UCC2003PK1 in this germ-free murine model at the end of the fourth week (Fig. 4C). The high level of UCC2003-tadAPK1 shedding may be caused by the structure of the GIT in the animals that were germ-free or monoassociated with bifidobacteria; the cecum is five times larger in these mice than in conventional animals (55). Therefore this organ may function as a suitable niche for *B. breve* UCC2003 in the absence of any competing

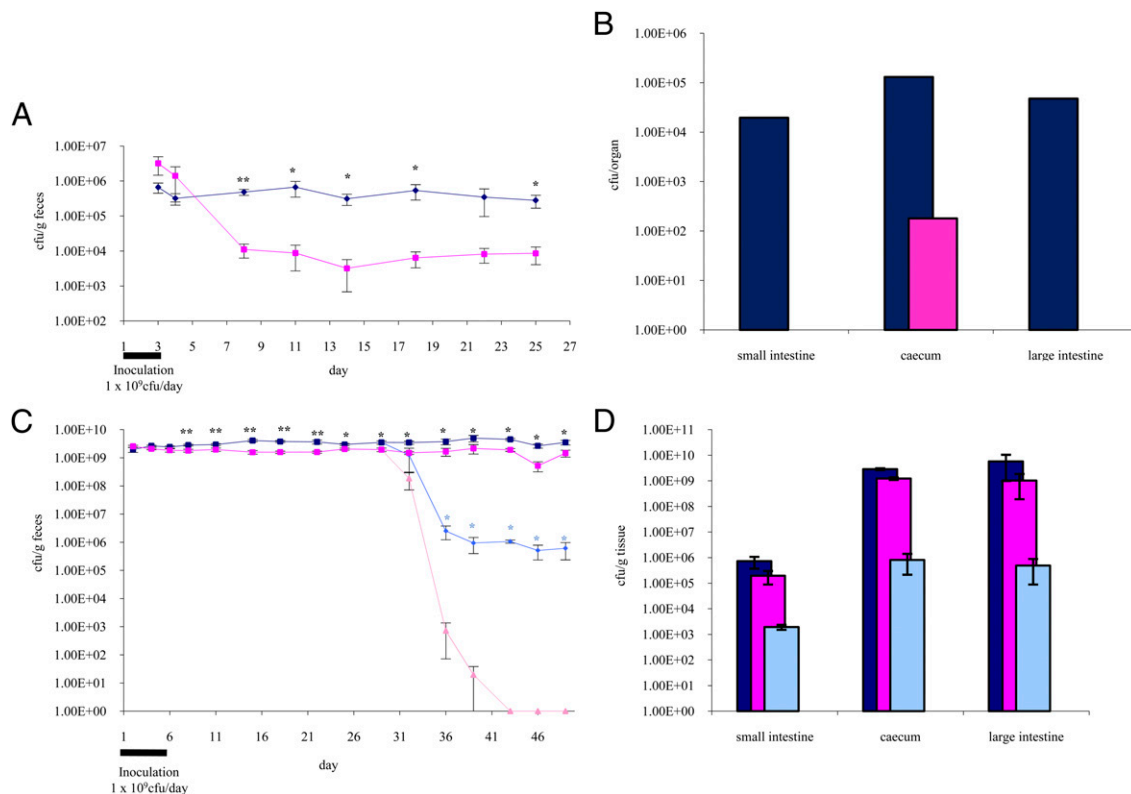


Fig. 4. *B. breve* UCC2003PK1 and *B. breve* UCC2003-tadAPK1 colonization of Balb/c and germ-free murine models. (A) Recovery of *B. breve* UCC2003PK1 (dark blue) or *B. breve* UCC2003-tadAPK1 (pink) from Balb/c murine fecal samples over a 25-d trial period. The data are average representative cfu g⁻¹ feces from five fecal samples per group for each time point (days). Error bars represent SEM. Statistically significant differences (as determined by student's *t*-test) between the groups are indicated by asterisks (* $P < 0.05$; ** $P < 0.01$). (B) Comparison of *B. breve* UCC2003PK1 (dark blue) and *B. breve* UCC2003-tadAPK1 (pink) recovered from the small intestine, the caecum, and the large intestine. Each set of data is representative of the average from five mice per group (except caecum *B. breve* UCC2003PK1, where $n = 2$ because three ceca were retained for RNA isolations). (C) Recovery of *B. breve* UCC2003PK1 or *B. breve* UCC2003-tadAPK1 from murine fecal samples of Swiss Webster mice monoassociated with *B. breve* UCC2003PK1 (dark blue) or *B. breve* UCC2003-tadAPK1 (pink) over a 7-wk trial period. Enumeration of each strain from the recolonized animals after day 29 is indicated by pale blue (*B. breve* UCC2003PK1) or light pink (*B. breve* UCC2003-tadAPK1). The data are average representative cfu g⁻¹ feces from 10 fecal samples per group for each time point up to day 29, after which there were five fecal samples per group for each time point. Error bars represent SEM. Statistically significant differences (as determined by student's *t*-test) between the monoassociated groups are indicated by black asterisks; statistically significant differences between the recolonized groups are indicated by blue asterisks (* $P < 0.05$; ** $P < 0.01$). (D) Comparison of *B. breve* UCC2003PK1 (monoassociated, dark blue; recolonized, pale blue) and *B. breve* UCC2003-tadAPK1 (monoassociated, pink; recolonized, light pink) recovered from the small intestine, the caecum, and the large intestine of monoassociated and recolonized animals. Each set of data is representative of the average from five mice per group.

microbiota, allowing both *B. breve* UCC2003PK1 and *B. breve* UCC2003-*tad*APK1 to be retained at high numbers.

To determine the effect on *B. breve* colonization and persistence following the introduction of a competitive gut microbiota, five mice from each monoassociated group (i.e., with *B. breve* UCC2003PK1 or *B. breve* UCC2003-*tad*APK1) were transferred to a non-germ-free area, and fecal pellets from conventional mice were introduced into the cages to promote microbial gut colonization. *B. breve* UCC2003PK1 or UCC2003-*tad*APK1 numbers in the fecal pellets were determined twice weekly for 3 wk. Although the numbers of UCC2003PK1 and UCC2003-*tad*APK1 remained high at $\sim 10^9$ cfu g⁻¹ feces in the mono-associated animals throughout the 7-wk trial period, a dramatic drop in the numbers of these two strains was observed 7 d after their murine hosts were moved to the non-germ-free area. The numbers of *B. breve* UCC2003PK1 fell to $\sim 10^6$ cfu g⁻¹ feces and remained at this level for the ensuing 2 wk of the trial. *B. breve* UCC2003-*tad*APK1 was barely detectable 7 d after transfer to the non-germ-free area and was undetectable in the fecal pellets thereafter (Fig. 4C). These numbers also were reflected when *B. breve* UCC2003PK1 and UCC2003-*tad*APK1 numbers were determined in the small intestine, cecum, and large intestine (Fig. 4D). These data substantiate our finding in the Balb/c mouse model and highlight the crucial importance of the *tad*₂₀₀₃ locus in the colonization and persistence of *B. breve* UCC2003 as part of a competing GIT microbiota.

Immunogold Labeling of *B. breve* UCC2003PK1 Reveals the Presence of Tad Pili. Immunogold electron microscopy was performed on cecal extracts from germ-free mice that had been monoassociated with *B. breve* UCC2003PK1 using antibodies that had been raised against a 20-amino acid sequence of Flp (*Materials and Methods*). The electron micrographs showed that *B. breve* UCC2003PK1 in cecal samples have pilus structures extending outward from the cell surface that could be labeled with Bbr0136 antiserum, whereas the same cells treated with Bbr0136 preimmuneserum showed no antibody labeling (Fig. 5). Furthermore, in agreement with the expression studies, no such immunolabeling could be observed with either the WT *B. breve* UCC2003PK1 or the UCC2003-*tad*APK1 strain when grown in MRS broth (Fig. S7). Collectively these data clearly implicate the *tad*₂₀₀₃ locus in GIT colonization and persistence, most likely through direct interaction between the observed extracellular Tad appendages and a host tissue component such as mucin in analogy with other pili (16).

Conclusion

To augment our understanding of bifidobacterial genomics and elucidate key determinants of *Bifidobacterium*–host interactions in the GIT, we determined the complete genome sequence of *B. breve* UCC2003, a nursing stool isolate. CGH of various *B. breve* isolates revealed a high level of sequence homology among the 19 *B. breve* strains examined; however, differences were

observed for each strain when compared with the genome of *B. breve* UCC2003. These differences were found primarily in genes encoding phage or phage-resistance systems, the genomic islands encoding proteins for EPS biosynthesis, and genes encoding extracellular proteins. Furthermore, the CGH analysis suggests that each *B. breve* strain may have different carbohydrate utilization capabilities, an observation that is supported by our published work and that highlights the importance of carbohydrate utilization capabilities for bifidobacterial colonization of the GIT (37, 42–45). The *in vivo* gene expression analysis clearly highlighted a number of genes whose expression was up-regulated when *B. breve* UCC2003 had colonized the mouse GIT. Many of these up-regulated genes are predicted to encode functions associated with bifidobacterial colonization and persistence, interaction with the host, and adaptation to the GI environment. The observed up-regulation of neuroserpin gene expression *in vivo* is consistent with predictions from *in vitro* experiments (39) and further supports the notion that such bifidobacteria produce the neuroserpin protease inhibitor to protect their surface-exposed and extracellular proteins, which are involved in colonization and host interactions, when in the GI environment.

Intriguingly, a locus dedicated to the production of type IVB or Tad pili was among the most highly up-regulated genes *in vivo*. Homologous *tad* loci are found in all sequenced Archaea and in both Gram-negative and Gram-positive bacteria (although in Gram-positive bacteria the homologous *tad* loci are found only in members of the *Actinobacteria*) (21). In addition, clear homologs of this *tad* locus and the associated *tadV* gene encoding the prepilin peptidase are present in all *B. breve* strains examined by CGH as well as in all completed bifidobacterial genome sequences. Interestingly, among different bifidobacterial genome sequences the observed level of identity between flp pili proteins or the predicted pseudopilins can be as little as 33%. Pili are known to adhere to carbohydrate moieties that are present in glycoprotein or glycolipid receptors (16, 21), and it has been suggested that pili form the first opportunity for bacteria to attach to their host; this opportunity is followed by a much closer and tighter association between the bacterium and the host cell surface (56). The sequence variation among the predicted bifidobacterial Tad pili proteins therefore may reflect differential receptor specificity and may determine the affinity of certain *Bifidobacterium* strains/species for particular hosts or host cell components while also reducing competition for attachment sites between related bifidobacterial species.

B. breve UCC2003 has become the most comprehensively studied bifidobacterial strain from a functional genomics perspective (36, 37, 41–45, 57, 58). The genetic accessibility and murine colonization capacity of this strain make it a valuable model for understanding bifidobacterial–host interactions in the gut. Our findings have shown the *tad*₂₀₀₃ locus to be an essential host-colonization factor. Future studies will focus on the specific role of each of the *tad* genes and on the identification of the epitopes to which the Tad pili adhere.

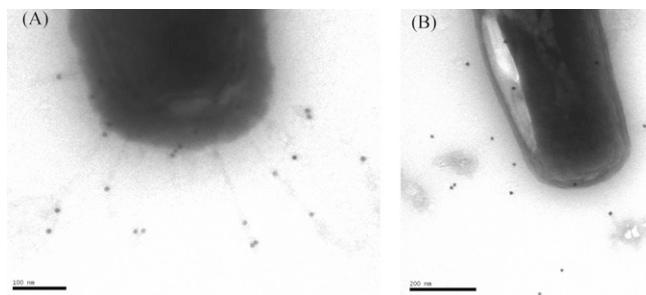


Fig. 5. Identification of pili in *B. breve* UCC2003 by immunogold electron microscopy. Cecal contents were suspended in PBS and subjected to low-speed centrifugation to sediment the chyme particles. The supernatant was collected, treated with anti-Flp₂₀₀₃ antibodies (A) and the corresponding preimmune serum (B), and labeled with protein A-conjugated gold particles (10 nm).

Materials and Methods

B. breve UCC2003 was manipulated, sequenced and analyzed as described in *SI Materials and Methods*. The genomic sequence of *B. breve* UCC2003 has been deposited in the GenBank database under the accession number CP000303. Detailed descriptions of bacterial strains and plasmids (Table S3) and oligonucleotides (Tables S4 and S5) used in this study as well as methods for comparative genome hybridizations, gene expression analyses, insertional mutant construction, murine trials, and immunogold electron microscopy are provided in *SI Materials and Methods*. The CGH and transcriptional array data have been deposited in the GEO database under accession number GSE27491.

- Qin J, et al.; MetaHIT Consortium (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65.
- Biagi E, et al. (2010) Through ageing, and beyond: Gut microbiota and inflammatory status in seniors and centenarians. *PLoS ONE* 5:e10667.
- Dominguez-Bello MG, et al. (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci USA* 107:11971–11975.
- Koenig JE, et al. (2011) Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci USA* 108:4578–4585.
- Roger LC, Costabile A, Holland DT, Hoyles L, McCartney AL (2010) Examination of faecal *Bifidobacterium* populations in breast- and formula-fed infants during the first 18 months of life. *Microbiology* 156:3329–3341.
- Favier CF, Vaughan EE, De Vos WM, Akkermans ADL (2002) Molecular monitoring of succession of bacterial communities in human neonates. *Appl Environ Microbiol* 68:219–226.
- O'Toole PW, Claesson MJ (2010) Changes throughout the lifespan from infancy to elderly. *Int Dairy J* 20:281–291.
- Ventura M, et al. (2009) Genome-scale analyses of health-promoting bacteria: Probiogenomics. *Nat Rev Microbiol* 7:61–71.
- Chow J, Lee SM, Shen Y, Khosravi A, Mazmanian SK (2010) Host-bacterial symbiosis in health and disease. *Adv Immunol* 107:243–274.
- Turroni F, Ribbera A, Foroni E, van Sinderen D, Ventura M (2008) Human gut microbiota and bifidobacteria: From composition to functionality. *Antonie van Leeuwenhoek* 94:35–50.
- Lee JH, O'Sullivan DJ (2010) Genomic insights into bifidobacteria. *Microbiol Mol Biol Rev* 74:378–416.
- Shanahan F (2010) The colonic microflora and probiotic therapy in health and disease. *Curr Opin Gastroenterol* 139:1808–1812.
- Roberfroid M, et al. (2010) Prebiotic effects: Metabolic and health benefits. *Br J Nutr* 104(Suppl 2):S1–S63.
- Fukuda S, et al. (2011) Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* 469:543–547.
- Klaassens ES, et al. (2009) Mixed-species genomic microarray analysis of fecal samples reveals differential transcriptional responses of bifidobacteria in breast- and formula-fed infants. *Appl Environ Microbiol* 75:2668–2676.
- Kankainen M, et al. (2009) Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human- mucus binding protein. *Proc Natl Acad Sci USA* 106:17193–17198.
- Kline KA, Dodson KW, Caparon MG, Hultgren SJ (2010) A tale of two pili: Assembly and function of pili in bacteria. *Trends Microbiol* 18:224–232.
- Proft T, Baker EN (2009) Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. *Cell Mol Life Sci* 66:613–635.
- Mandlik A, Swierczynski A, Das A, Ton-That H (2008) Pili in Gram-positive bacteria: Assembly, involvement in colonization and biofilm development. *Trends Microbiol* 16:33–40.
- Danelishvili L, Yamazaki Y, Selker J, Bermudez LE (2010) Secreted *Mycobacterium tuberculosis* Rv3654c and Rv3655c proteins participate in the suppression of macrophage apoptosis. *PLoS ONE* 5:e10474.
- Tomich M, Planet PJ, Figurski DH (2007) The Tad locus: Postcards from the widespread colonization island. *Nat Rev Microbiol* 5:363–375.
- Pellic V (2008) Type IV pili: E pluribus unum? *Mol Microbiol* 68:827–837.
- Kachlany SC, Planet PJ, DeSalle R, Fine DH, Figurski DH (2001) Genes for tight adherence of *Actinobacillus actinomycetemcomitans*: From plaque to plague to pond scum. *Trends Microbiol* 9:429–437.
- Schreiner HC, et al. (2003) Tight-adherence genes of *Actinobacillus actinomycetemcomitans* are required for virulence in a rat model. *Proc Natl Acad Sci USA* 100:7295–7300.
- Bottacini F, et al. (2010) Comparative genomics of the genus *Bifidobacterium*. *Microbiology* 156:3243–3254.
- Eisen JA, Heidelberg JF, White O, Salzberg SL (2000) Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biology* 1:RESEARCH0011.
- Haft DH, Selengut J, Mongodin EF, Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLOS Comput Biol* 1:e60.
- Ventura M, et al. (2005) Prophage-like elements in bifidobacteria: Insights from genomics, transcription, integration, distribution, and phylogenetic analysis. *Appl Environ Microbiol* 71:8692–8705.
- Ventura M, et al. (2009) Comparative analyses of prophage-like elements present in bifidobacterial genomes. *Appl Environ Microbiol* 75:6929–6936.
- van Pijkeren JP, et al. (2006) Comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *Lactobacillus salivarius* UCC118. *Appl Environ Microbiol* 72:4143–4153.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795.
- Sutcliffe IC, Russell RR (1995) Lipoproteins of gram-positive bacteria. *J Bacteriol* 177:1123–1128.
- Boekhorst J, de Been MW, Kleerebezem M, Siezen RJ (2005) Genome-wide detection and analysis of cell wall-bound proteins with LPxTG-like sorting motifs. *J Bacteriol* 187:4928–4934.
- Navarre WW, Schneewind O (1999) Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol Mol Biol Rev* 63:174–229.
- Clancy KW, Melvin JA, McCafferty DG (2010) Sortase transpeptidases: Insights into mechanism, substrate specificity, and inhibition. *Biopolymers* 94:385–396.
- 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.
- O'Connell Motherway M, Fitzgerald GF, van Sinderen D (2011) Metabolism of a plant derived galactose-containing polysaccharide by *Bifidobacterium breve* UCC2003. *Microb Biotechnol* 4:403–416.
- Ivanov D, et al. (2006) A serpin from the gut bacterium *Bifidobacterium longum* inhibits eukaryotic elastase-like serine proteases. *J Biol Chem* 281:17246–17252.
- Turroni F, et al. (2010) Characterization of the serpin-encoding gene of *Bifidobacterium breve* 210B. *Appl Environ Microbiol* 76:3206–3219.
- 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.
- 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.
- Ryan SM, Fitzgerald GF, van Sinderen D (2006) Screening for and identification of starch-, amylopectin-, and pullulan-degrading activities in bifidobacterial strains. *Appl Environ Microbiol* 72:5289–5296.
- Pokusaeva K, et al. (2010) Ribose utilization by the human commensal *Bifidobacterium breve* UCC2003. *Microb Biotechnol* 3:311–323.
- Pokusaeva K, O'Connell-Motherway M, Zomer A, Fitzgerald GF, van Sinderen D (2009) Characterization of two novel alpha-glucosidases from *Bifidobacterium breve* UCC2003. *Appl Environ Microbiol* 75:1135–1143.
- Pokusaeva K, et al. (2011) Cellodextrin utilization by *bifidobacterium breve* UCC2003. *Appl Environ Microbiol* 77:1681–1690.
- Aslanidis C, Schmid K, Schmitt R (1989) Nucleotide sequences and operon structure of plasmid-borne genes mediating uptake and utilization of raffinose in *Escherichia coli*. *J Bacteriol* 171:6753–6763.
- Davey R, Savva G, Dicks J, Roberts IN (2007) MPP: A microarray-to-phylogeny pipeline for analysis of gene and marker content datasets. *Bioinformatics* 23:1023–1025.
- Cronin M, Sleanor RD, Hill C, Fitzgerald GF, van Sinderen D (2008) Development of a luciferase-based reporter system to monitor *Bifidobacterium breve* UCC2003 persistence in mice. *BMC Microbiol* 8:161.
- Cronin M, et al. (2010) Orally administered bifidobacteria as vehicles for delivery of agents to systemic tumors. *Mol Ther* 18:1397–1407.
- Takeichi M (1990) Cadherins: A molecular family important in selective cell-cell adhesion. *Annu Rev Biochem* 59:237–252.
- Skorstengaard K, Jensen MS, Petersen TE, Magnusson S (1986) Purification and complete primary structures of the heparin-, cell-, and DNA-binding domains of bovine plasma fibronectin. *Eur J Biochem* 154:15–29.
- 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:e113.
- Sonnenburg ED, et al. (2010) Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* 141:1241–1252.
- Bhattacharjee MK, Kachlany SC, Fine DH, Figurski DH (2001) Nonspecific adherence and fibril biogenesis by *Actinobacillus actinomycetemcomitans*: TadA protein is an ATPase. *J Bacteriol* 183:5927–5936.
- Derrien M, et al. (2010) Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut Microbes* 1:254–268.
- Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G (2006) Pili in Gram-positive pathogens. *Nat Rev Microbiol* 4:509–519.
- 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.
- Zomer A, et al. (2009) An interactive regulatory network controls stress response in *Bifidobacterium breve* UCC2003. *J Bacteriol* 191:7039–7049.