APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Nov. 2010, p. 7373–7381 Vol. 76, No. 22
0099-2240/10/$12.00 doi:10.1128/AEM.00675-10

Broad Conservation of Milk Utilization Genes in *Bifidobacterium longum* subsp. *infantis* as Revealed by Comparative Genomic Hybridization▽†

Riccardo G. LoCascio,[1,2] Prerak Desai,[3,4] David A. Sela,[1,2] Bart Weimer,[4] and David A. Mills[1]*

Department of Viticulture and Enology, Robert Mondavi Center for Wine and Food Science,[1] and Microbiology Graduate Group,[2] University of California, Davis, California 95616; Department of Nutrition, Dietetics, and Food Sciences, Utah State University, Logan, Utah, 84322[3]; and Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, California 95616[4]

Received 15 March 2010/Accepted 16 August 2010

Human milk oligosaccharides (HMOs) are the third-largest solid component of milk. Their structural complexity renders them nondigestible to the host but liable to hydrolytic enzymes of the infant colonic microbiota. Bifidobacteria and, frequently, *Bifidobacterium longum* strains predominate the colonic microbiota of exclusively breast-fed infants. Among the three recognized subspecies of *B. longum*, *B. longum* subsp. *infantis* achieves high levels of cell growth on HMOs and is associated with early colonization of the infant gut. The *B. longum* subsp. *infantis* ATCC 15697 genome features five distinct gene clusters with the predicted capacity to bind, cleave, and import milk oligosaccharides. Comparative genomic hybridizations (CGHs) were used to associate genotypic biomarkers among 15 *B. longum* strains exhibiting various HMO utilization phenotypes and host associations. Multilocus sequence typing provided taxonomic subspecies designations and grouped the strains between *B. longum* subsp. *infantis* and *B. longum* subsp. *longum*. CGH analysis determined that HMO utilization gene regions are exclusively conserved across all *B. longum* subsp. *infantis* strains capable of growth on HMOs and have diverged in *B. longum* subsp. *longum* strains that cannot grow on HMOs. These regions contain fucosidases, sialidases, glycosyl hydrolases, ABC transporters, and family 1 solute binding proteins and are likely needed for efficient metabolism of HMOs. Urea metabolism genes and their activity were exclusively conserved in *B. longum* subsp. *infantis*. These results imply that the *B. longum* has at least two distinct subspecies: *B. longum* subsp. *infantis*, specialized to utilize milk carbon, and *B. longum* subsp. *longum*, specialized for plant-derived carbon metabolism.

The newborn infant not only tolerates but requires colonization by commensal microbes for its own development and health (3). The relevance of the gut microbiome in health and disease is reflected by its influence in a number of important physiological processes, from physical maturation of the developing immune system (28) to the altered energy homeostasis associated with obesity (51, 52).

Human milk provides all the nutrients needed to satisfy the neonate energy expenditure and a cadre of molecules with nonnutritional but biologically relevant functions (6). Neonatal health is likely dependent on the timely and complex interactions among bioactive components in human milk, the mucosal immune system, and specialized gut microbial communities (30). Human milk contains complex prebiotic oligosaccharides that stimulated the growth of select bifidobacteria (24, 25) and are believed to modulate mucosal immunity and protect the newborn against pathogens (23, 33, 41). These complex oligosaccharides, which are abundantly present in human milk (their structures are reviewed by Ninonuevo et al. [31] and LoCascio et al. [24]), arrive intact in the infant colon (5) and modulate the composition of neonatal gastrointestinal (GI) microbial communities.

Bifidobacteria and, frequently, *Bifidobacterium longum* strains often predominate the colonic microbiota of exclusively breast-fed infants (10, 11). Among the three subspecies of *B. longum*, only *B. longum* subsp. *infantis* grows robustly on human milk oligosaccharides (HMOs) (24, 25). The availability of the complete genome sequences of *B. longum* subsp. *infantis* ATCC 15697 (40) and two other *B. longum* subsp. *longum* strains (22, 39) made possible the analysis of whole-genome diversity across the *B. longum* species. Analysis of the *B. longum* subsp. *infantis* ATCC 15697 genome has identified regions predicted to enable the metabolism of HMOs (40); however, their distribution across the *B. longum* spp. remains unknown. We predict that these regions are exclusively conserved in *B. longum* strains adapted to colonization of the infant gut microbiome and are therefore capable of robust growth on HMOs. In this work, whole-genome microarray comparisons (comparative genomic hybridizations [CGHs]) were used to associate genotypic biomarkers among 15 *B. longum* strains exhibiting various HMO utilization phenotypes and host associations.

MATERIALS AND METHODS

Bacterial strains, HMO utilization, and DNA preparation. Bifidobacterial strains (Table 1) were obtained from the Japanese Collection of Microorganism (Riken BioResource Center, Japan), the American Type Culture Collection (Manassas, VA), and the University of California, Davis (UC-Davis), Viticulture and Enology Culture Collection (Davis, CA). Cultures were propagated in de-

* Corresponding author. Mailing address: Department of Viticulture and Enology, University of California, One Shields Avenue, Davis, CA 95616. Phone: (530) 754-7821. Fax: (530) 752-0382. E-mail: damills@ucdavis.edu.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 27 August 2010.

TABLE 1. Strains used in this study

B. longum strain	Species	Origin	OD_{600}[a]	
			Urease activity	HMO consumption
ATCC 15697	*B. longum* subsp. *infantis*	Feces of human infant	+	+
JCM 11346	*B. longum* subsp. *infantis*	Feces of human infant	+	+
JCM 1210	*B. longum* subsp. *infantis*	Feces of human infant	+	+
JCM 1260	*B. longum* subsp. *infantis*	Feces of human infant	+	±
JCM 1272	*B. longum* subsp. *infantis*	Feces of human infant	+	+
JCM 7007	*B. longum* subsp. *infantis*	Feces of human infant	+	+
JCM 7009	*B. longum* subsp. *infantis*	Feces of human infant	+	+
JCM 7011	*B. longum* subsp. *infantis*	Feces of human infant	−	+
ATCC 15708	*B. longum* subsp. *longum*	Feces of human infant	−	±
DJO10A	*B. longum* subsp. *longum*	Feces of human infant	−	−
UCD186	*B. longum* subsp. *longum*	Probiotic pill	−	−
UCD20	*B. longum* subsp. *suis*/*B. longum* subsp. *longum*	Feces of calf	−	−
UCD49	*B. longum* subsp. *suis*/*B. longum* subsp. *longum*	Feces of infant monkey	+	−
JCM 11347	*B. longum* subsp. *longum*[b]	Feces of human infant	+	±
JCM 7010	*B. longum* subsp. *longum*[b]	Feces of human infant	−	+
CCUG52486[c]	*B. longum* subsp. *longum*[d]	Feces of healthy elderly individual	NA[e]	NA
ATCC 55813[c]	*B. longum* subsp. *longum*[f]	Feces of human infant	NA	NA

[a] +, $OD_{600} > 0.9$; ±, $OD_{600} = 0.9$ to 0.4; −, $OD_{600} = 0.4$ to 0.0. From LoCascio et al. (25) and Ahrens (1).
[b] Strains originally typed as *B. longum* supbsp. *infantis* by the Japanese Collection of Microorganism (JCM) (37).
[c] *In silico* analysis only.
[d] Strain originally typed as *B. longum* subsp. *infantis* by Silvi et al. (42).
[e] NA, not available.
[f] Strain originally typed as *B. longum* subsp. *infantis* by Yang (56).

Man, Rogosa, and Sharpe (MRS) medium (Becton Dickinson, Sparks, MD) supplemented with 0.05% L-cysteine HCl and incubated at 37°C in an anaerobic chamber (Coy Laboratory Products, Grass Lake, MI). Data for HMO utilization were obtained by growth assays using 1.5% (wt/vol) HMO as the sole carbon source, and growth was measured by determination of the optical density at 600 nm (OD_{600}) in a microtiter plate. The data are available from LoCascio et al. (25). Genomic DNA was extracted using a Promega Wizard genomic DNA isolation system (Madison, WI) following standard procedures, indicated by the manufacturer, and using lysozyme (50 mg/ml in Tris-EDTA) and mutanolysin (1,000 U/ml) as lytic enzymes. DNA quality and yield were checked using a Nanodrop apparatus (Wilmington, DE), and the DNA was stored at −20°C until further use.

Multilocus sequence typing (MLST) locus selection, gene amplification, and sequencing. Intragenic regions of seven housekeeping genes (*clpC*, *dnaB*, *dnaG*, *dnaJ1*, *purF*, *rpoC*, *xfp*) were selected on the basis of the findings of a previous study by Ventura et al. (54). These loci were mapped on the chromosome of *B. longum* subsp. *infantis* ATCC 15697 (GenBank accession number CP001095) to ensure even spacing across the entire genome and presence as a single copy (see Fig. S1 in the supplemental material). Primers from Ventura et al. (54) were optimized or redesigned on the basis of the known gene sequences of *B. longum* subsp. *infantis* ATCC 15697 and *B. longum* subsp. *longum* DJO10A (GenBank accession number CP000605), and amplification products were obtained for all the tested strains (see Table S1 in the supplemental material). PCR amplifications were performed on an Applied Biosystems (ABI) 2720 thermal cycler (Mountain View, CA) using the Promega GoTaq green master mixture. PCR cycling conditions were optimized for every primer set (see Table S1 in the supplemental material) and consisted of an initial denaturation at 95°C for 5 min, followed by 30 cycles of 95°C for 30 s, annealing at 59.5 to 52.6°C for 30 s, and elongation at 72°C for 46 to 60 s, followed by a final extension at 72°C for 8 min and a hold at 4°C. PCR mixtures (50 to 100 μl) were prepared according to the manufacturer's protocol. The resulting amplicons were separated on a 0.8% agarose gel, followed by ethidium bromide staining, and were purified using a Qiagen QIAquick PCR purification kit (Valencia, CA). Sequencing was performed on an ABI 3730 capillary electrophoresis genetic analyzer using BigDye Terminator chemistries at the University of California Sequencing Center.

MLST data analysis. Sequencing data for all the loci were edited using the BioEdit program (version 7.0; http://www.mbio.ncsu.edu/BioEdit/BioEdit.html) and were aligned using the CLUSTAL W program (50). Phylogenetic analysis and concatenations of the sequenced loci were performed using molecular evolutionary genetic analysis (MEGA) software, version 4 (http://megasoftware.net). Descriptive evolutionary analysis, including the mol% G+C content, number of polymorphic sites, nucleotide diversity (π), and average number of nucleotide differences (k), was performed using DnaSP software, version 4.0 (see Table S2 in the supplemental material). Allelic sequences were assigned as described previously (4) (see Table S3 in the supplemental material). A minimum-evolution tree was calculated using MEGA (version 4.0) (Fig. 1), and split decomposition analysis was performed on individual and concatenated loci using SplitsTree software (version 4.1) (17).

Urease assay. Urease activity was tested as described by Zotta et al. (57), modified as follows: cells were grown to an OD_{600} of 1.0 in MRS medium supplemented with 5 g/liter filtered-sterilized urea. The cells were then washed in phosphate-buffered saline (PBS) prior to and after permeabilization with 10% (vol/vol) toluene and immediately assayed for urease activity using BBL taxonomic differentiation urea discs (Becton Dickinson). Assays were performed in biological triplicates and validated using *B. longum* ATCC 15697 and *B. longum* DJO10A as the urease-positive and -negative control strains, respectively.

CGH. (i) Array design and hybridization. Total genomic DNA (1 μg) was isolated to a quality of an A_{260}/A_{280} ratio of ≥1.5 and an A_{260}/A_{230} ratio of ≥1.5. Subsequently, it was sheared for 10 min with 0.6 U of Promega DNase 1 at 37°C. Sheared genomic DNA was labeled with the GeneChip DNA labeling reagent (Affymetrix Inc., Santa Clara, CA), according to the manufacturer's recommendations. The labeled genomic DNA (50 ng) was hybridized onto custom Affymetrix DNA chips containing probe sets designed for all the annotated coding sequences (CDSs) of *B. longum* subsp. *infantis* ATCC 15697 (the control strain). This design resulted in 2,428 probe sets, each composed of 11 unique 17-mer probe sequences per CDS. The chips were hybridized and scanned at the Center for Integrated BioSystems (Utah State University, Logan, UT), according to the manufacturer's protocols for *Escherichia coli*. Hybridizations for each strain were done in two biological replicates.

(ii) CGH data normalization, visualization, and analysis. Raw intensities (.cel files) from all the chips were normalized together with the robust multichip average (RMA) method (18) using the xCluster R module (Center for Integrated BioSystems, Utah State University, Logan, UT) that resulted in a normalized file of $\log_2$-transformed intensities that were subsequently used for statistical analysis. Average $\log_2$ ratios (LRs) for each sample were calculated by determining the difference in $\log_2$ intensity of each CDS from that of control strain *B. longum* ATCC 15697. Heat maps were generated from LR data using The Institute for Genomic Research MultiExperiment Viewer (MeV), version 4.3 (36). Unsupervised data analysis was performed in MeV using the *k*-means clustering (KMC) (15, 44) and hierarchical clustering (HCL) (9) modules.

(iii) Array validation and probe set/SI determination. The performance of the array was tested by calculating the distribution of LRs for *B. longum* subsp.

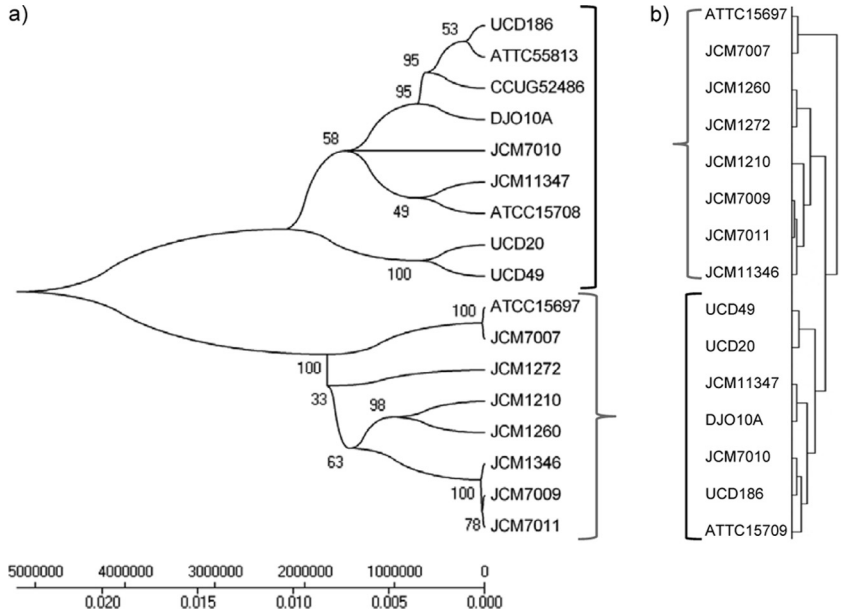


FIG. 1. Evolutionary relationship of *B. longum* spp. strains used in the study. (a) The MLST-based hierarchical clustering was inferred using the minimum-evolution method (35). The optimal tree with the sum of branch length of 0.12 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches (12). The phylogenetic tree was linearized assuming equal evolutionary rates in all lineages (47). The clock calibration to convert distance to time was 2.1×10^8 (time/node height). The tree is drawn to scale, with the branch lengths being in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the maximum-composite-likelihood method (49) and are in units of the number of base substitutions per site. Phylogenetic analyses were conducted in MEGA (version 4) (48). Sequence fragments for the seven MLST loci of *B. longum* ATCC 55813 and *B. longum* CCUG 52486 were obtained from NCBI. (b) Hierarchical clustering (Euclidian distance, average linking) based on the entire CGH data set and performed in MeV using the HCL algorithm.

infantis JCM 7007, a strain that, with the exception of a missing bacteriophage insert (Blon_1780 to Blon_1832), is identical to the control strain. Array technical variation was calculated by comparing replicates of control/control hybridizations. LR thresholds for highly conserved, divergent, and absent genes were computed on binned probe set/sequence identity (SI) values, as described by Taboada et al. (46). Briefly, all predicted CDSs represented on the ATCC 15697 array were aligned against the predicted CDSs of *B. longum* subsp. *longum* DJO10A, and SI values for pair-wise alignments were determined with the BLAST program (blastall −p blastn −m 8 −a 2) (2). Strong sequence identity between the sequenced test strain (DJO10A) and the array control strain (ATCC 15697) yielded a stronger array signal. Previous CGH studies have used $100\% \geq SI \geq 80\%$ as a range for determining divergence between test and control strains (46). Over this range, the increased array signal correlated with stronger sequence identity in the test strain (see Fig. S3 in the supplemental material). The assumption was made that the function describing the shape of this curve is linearly increasing. This method was verified by testing the sequenced DJO10A strain and resulted in false-positive and false-negative rates similar to those of previously published CGH analysis methods (43, 46).

(iv) *In silico* analysis of additional *B. longum* genomes. Draft genome sequences for *B. longum* ATCC 55813 (GenBank accession number ACHI00000000) and *B. longum* CCUG52486 (GenBank accession number ABQQ00000000) were obtained from NCBI and aligned to the CDSs present on the ATCC 15697 array by BLAST analysis (Blastall −p blastn −m 8 −a 2 −e 0.001) (2). The resulting hits were manually cleared of alignments of <150 bp, and any remaining multiple hits were cleared of entries with the highest E values and shortest alignments. The *in silico* CGH was performed using Equation 1 in Fig. S3 in the supplemental material to convert blastn-generated percent sequence identity values into LR values. CDSs for which the SI call was missing (for blastn hits of <150 bp or a divergent gene) were assigned an intensity value of −1.3 and cannot be commented on in this analysis. Sequence data for the seven loci utilized for MLST were obtained from NCBI and analyzed as described above.

Genome sequencing of *B. longum* subsp. *infantis* JCM 1260. *B. longum* subsp. *infantis* JCM 1260 genomic DNA was extracted and used to generate a sequencing library, according to the manufacturer's protocols (NEBNext; New England Biolabs). The JCM 1260 library was sequenced by synthesis with a Genome Analyzer II apparatus (Illumina) for 40 cycles at the UC-Davis DNA Core Facility. The resultant 15.9 million reads were aligned to the sequence of the *B. longum* subsp. *infantis* ATCC 15697 HMO cluster using the MOSAIK program (version 0.9; http://bioinformatics.bc.edu/marthlab/Mosaik).

COGs enrichment analysis. Clusters of orthologous groups (COGs) enrichment analysis was done using gene set enrichment analysis (GSEA) (29, 45) (http://www.broad.mit.edu/gsea/) to identify genes that were significantly overrepresented within COGs (false discovery rate [FDR], <25%). COG annotations were obtained from NCBI, and enrichments using GSEA were done using the LR for each tested strain as input, weighted_p2 as the enrichment statistic, and the Diff_of_classes method as the metric for gene ranking. Statistical analysis was done using the SAM program (version 2.01) (53) to identify genes that were conserved in all the *B. longum* subsp. *infantis* strains but divergent in *B. longum* subsp. *infantis* strain JCM 1260. The data were first filtered to include only the probe sets containing LRs of ≤−0.4 for all *B. longum* subsp. *infantis* strains except JCM 1260. SAM analysis was done on the filtered data set using the LRs for the eight *B. longum* subsp. *infantis* strains as input. Probe sets with *q* values of ≤0.254 were considered significant and indicated genes that were conserved among all *B. longum* subsp. *infantis* strains but divergent in JCM 1260.

Nucleotide sequences accession numbers. The nucleotide sequences of the partial *clpC*, *dnaB*, *dnaG*, *dnaJ1*, *purF*, *rpoC*, and *xfp* genes have been deposited in the GenBank database under accession numbers HQ438317 to HQ438435.

Microarray data accession numbers. The microarray data reported here have been deposited in the Gene Expression Omnibus (GMO) database under accession number E-MEXP-2951.

RESULTS

MLST analysis. Prior to the CGH and to address the ongoing taxonomic uncertainty over subspecies designations for the *B. longum* species (27, 37), MLST was used to type and prescreen strains. *B. longum* strains previously isolated from the feces of human, calf, and monkey newborns were selected for analysis on the basis of their HMO consumption phenotype.

As previously reported, the HMO-positive (HMO^+) phenotype (i.e., the ability to grow on HMOs) was predominantly associated with *B. longum* subsp. *infantis* strains; in contrast, *B. longum* subsp. *longum* strains have generally shown mild ($HMO^\pm$ phenotype) or no growth (HMO^- phenotype) on HMOs (Table 1) (25). The MLST scheme adapted from Ventura et al. (54) successfully differentiated all strains at the subspecies level. A total of 287 single nucleotide polymorphisms (SNPs) were analyzed in the seven loci and generated between 5 (*xfp*) and 116 (*purF*) polymorphic sites (more details on the MLST analysis are provided in Table S2 in the supplemental material). A consensus phylogeny dendrogram was generated on concatenated MLST data and resulted in two major taxa with >58% bootstrap support (Fig. 1a). While the taxonomic and phylogenetic positions of *B. longum* subsp. *suis* are somewhat unclear (7), the segregation of nonhuman isolates *B. longum* subsp. *suis* UCD49 and UCD20 as a subgroup of the *B. longum* subsp. *longum* taxon confirms that *B. longum* subsp. *suis* is more closely related to *B. longum* subsp. *longum* than *B. longum* subsp. *infantis*. According to this analysis, strains JCM 11347 and JCM 7010 were erroneously classified as *B. longum* subsp. *infantis* by Sakata et al. (37) and should belong to *B. longum* subsp. *longum*. Furthermore, *in silico* MLST analysis of the draft genomes of *B. longum* subsp. *infantis* ATCC 55813 (GenBank accession number ACHI00000000) and *B. longum* subsp. *infantis* CCUG52486 (GenBank accession number ABQQ00000000) indicates that these organisms are in the *B. longum* subsp. *longum* clade and were also incorrectly classified as *B. longum* subsp. *infantis* (42, 56) (Table 1). HCL analysis of the CGH probe hybridization data resulted in a hierarchical dendrogram highly similar to the MLST-derived linearized minimum-evolution tree (Fig. 1b). In these dendrograms, one group contained HMO^- or $HMO^\pm$ strains clustered with *B. longum* subsp. *longum* DJO10A and was classified as *B. longum* subsp. *longum*. The other group contained HMO^+ strains and clustered with the sequenced type strain *B. longum* subsp. *infantis* ATCC 15697. Strains in this group were classified as *B. longum* subsp. *infantis*. MLST can be employed as a fast and cost-effective method for unequivocally distinguishing among the three *B. longum* subspecies.

CGH array validation and probe set/SI determination. The technical variability of the array platform was measured in replicates of control/control hybridizations (ATCC 15697/ATCC 15697), resulting in a mean variability in probe set LR signals of ±0.07% (data not shown). To evaluate array performance and sensitivity, an approach described by Paustian et al. (34) was implemented, and LR distributions were measured in hybridizations with JCM 7007. JCM 7007/ATCC 15697 hybridizations resulted in 97% of the probe sets having LR signals that were normally distributed between −0.6 and 0.6 (see Fig. S6 in the supplemental material), supporting the suggestion that the array platform was well suited for CGHs.

LR thresholds for highly conserved, divergent, and absent genes were determined to explore the divergence among the strains in the *B. longum* clade. Using an approach developed by Taboada et al. (46) and Snipen et al. (43), four discrete percent SI categories describing absent, divergent, partially conserved, and highly conserved CDSs were established (see Fig. S2 in the supplemental material). The increase of sequenced bifidobacterial genomes enabled supervised learning techniques to be utilized to establish high-accuracy CGH binning calls, as described previously (43). The proportion of genes from *B. longum* DJO10A belonging to these four categories was calculated on the basis of the binned LR data, and error rates were calculated as described previously (46). Conserved genes were assigned at LR values of ≥ -0.4. At this threshold, only 2.6% of strain ATCC 15697 CDSs absent in strain DJO10A (identity [ID], ≤40%) were incorrectly binned as conserved (20 of 751 CDSs), and all false-positive results corresponded to hypothetical proteins. Absent genes were assigned at LR values of ≤ -2.6. At this threshold, 0.6% of genes highly conserved between DJO10A and ATCC 15697 (ID, ≥97%) were incorrectly binned as absent (6 of 811 CDSs), with three of the false-negative results corresponding to hypothetical proteins. The region with values of $-2.6 \leq LR \geq -0.4$ corresponded to genes that were either absent, divergent, or partially conserved; however, this region has significant overlap among SI categories. Thus, in this range misclassification of the CGH data is likely to occur. For example, at an LR of ~−0.8, there is an equal proportion of genes being classified as absent, divergent, or partially conserved; and in regions proximal to those with an LR of less than ~−0.8, there is a markedly high probability of absent genes (see Fig. S2 in the supplemental material).

Genome-wide features of CGH data. Except for a single bacteriophage-related insertion, M-1, control strain ATCC 15697 and JCM 7007 were otherwise identical. Seventeen other regions containing remnants of bacteriophage insertions were found (see Table S4 in the supplemental material). These regions rich in mobile elements contained hypothetical proteins with unknown functions, and a significant number of them were exclusively conserved in ATCC 15697 and JCM 7007. An example is M-3, an ~164-Mb region rich in mobile elements. In addition, M-6 and M-7 are almost exclusively conserved in *B. longum* subsp. *infantis* and contain genes annotated as solute binding proteins (SBPs).

COGs enrichment analysis. GSEA is a computational method to establish statistically significant enrichment of COGs within concordant groups on the basis of the defined set of genes in a group (i.e., a COG) (45). Using this technique, cellular functions that are significantly overrepresented at levels beyond those due to chance can provide insight into the specific functions influenced by a treatment, in this case, evolutionary differences between strains. To test the COG evolution between strains and subspecies, we conducted a GSEA analysis using the LR signal intensities from their respective CDS and COG annotations. This analysis resulted in a list of COGs having statistically significant overrepresentation in genes associated with these functions, or enrichment, between the subspecies. A total of 5/19 COGs were significantly enriched in *B. longum* subsp. *infantis* (FDR, <25%) (see Table S5 in the supplemental material). The enrichment score (ES), which reflects the amount of gene overrepresentation, was highest for COG G (carbohydrate transport and metabolism). Thirty-one of 51 genes that were enriched in COG G belonged to COG 1653 (SBP; family 1), COG 2814 (major facilitator superfamily [MFS]; MFS_1), COG 0395, and COG 1175 (ABC-type sugar transport systems, permease components). These COGs were significantly enriched in *B. longum* subsp.

infantis strains and underrepresented in *B. longum* subsp. *longum* strains ($q \leq 0.254$).

Unsupervised clustering of CGH data. KMC clustered the CGH data in three groups: CDSs conserved exclusively in ATCC 15697 (group 1); CDSs conserved in *B. longum* subsp. *infantis*, forming a *B. longum* subsp. *infantis* core genome set (group 2); and CDSs shared by both subspecies, which formed the *B. longum* core genome set (group 3). Group 3 retained the largest number of CDSs, corresponding to 51% ($n = 1,235$) of the total CDSs on the array. Groups 1 and 2 retained 25% ($n = 605$) and 24% ($n = 588$) of the CDSs, respectively. The average G+C skew from the ATTC 15696 genome in these three groups was ±5% for group 1 and ranged from −35% to 16% for groups 2 and 3 (data not shown).

(i) Group 1: genes unique to strain ATCC 15697. A total of 605 CDSs were conserved in ATCC 15697 and JCM 7007 (see Fig. S7 in the supplemental material). This group was predominantly characterized by mobile elements composed of hypothetical proteins, bacteriophage-related proteins, and transposases. In this analysis, strains JCM 1210, JCM 1260, and JCM 1272 clustered in close proximity with ATCC 15697, partially sharing among each other two regions, R1-1 and R1-2, featuring mobile elements containing hypothetical proteins but also a notable number of genes related to carbohydrate transport and metabolism.

(ii) Group 2: genes unique to *B. longum* subsp. *infantis*. The *B. longum* subsp. *infantis* core genome has 588 conserved CDS (see Fig. S8 in the supplemental material). An earlier enumeration of the core genome of the sequenced *B. longum* subsp. *infantis* type strain had identified 702 genes (40). Group 2 contains gene regions where conservation and divergence among the strains are highly informative to better understand the underlying genetic components separating *B. longum* subsp. *infantis*, *B. longum* subsp. *longum*, and *B. longum* subsp. *suis*. For example, this analysis further supports the suggestion that while *B. longum* subsp. *suis* UCD20 and UCD49 are phylogenetically similar to *B. longum* subsp. *longum* (7), they possess a significant number of functionally important gene regions (R2-1, R2-6, R2-7, R2-4, R2-5) exclusively shared with *B. longum* subsp. *infantis*. These include mobile element insertions containing a gene (Blon_1750) predicted to be a bacteriocin of the lactococcin 972 family, several ABC-type oligosaccharide transporters, and SBP-related genes, particularly in R2-4. Interestingly, R2-2 and R2-3 contain several CDSs annotated as ABC transporters, oligosaccharide symporters, SBPs, and alpha-fucosidase. This fucosidase, conserved only among *B. longum* subsp. *infantis* strains, may play a role in the metabolism of fucosylated glycans present on milk oligosaccharides and glycosylated proteins/peptides.

(iii) Group 3: genes common to *B. longum* species. Group 3 contained 1,235 CDSs (see Fig. S9 in the supplemental material), representing approximately 50% of the *B. longum* genome shared between the two subspecies. Other than CDSs associated with basic cellular functions, this shared genome contains carbohydrate metabolism-related CDSs, predicted to be responsible for the ability of *B. longum* to grow on a wide variety of mono- and disaccharides (1, 37).

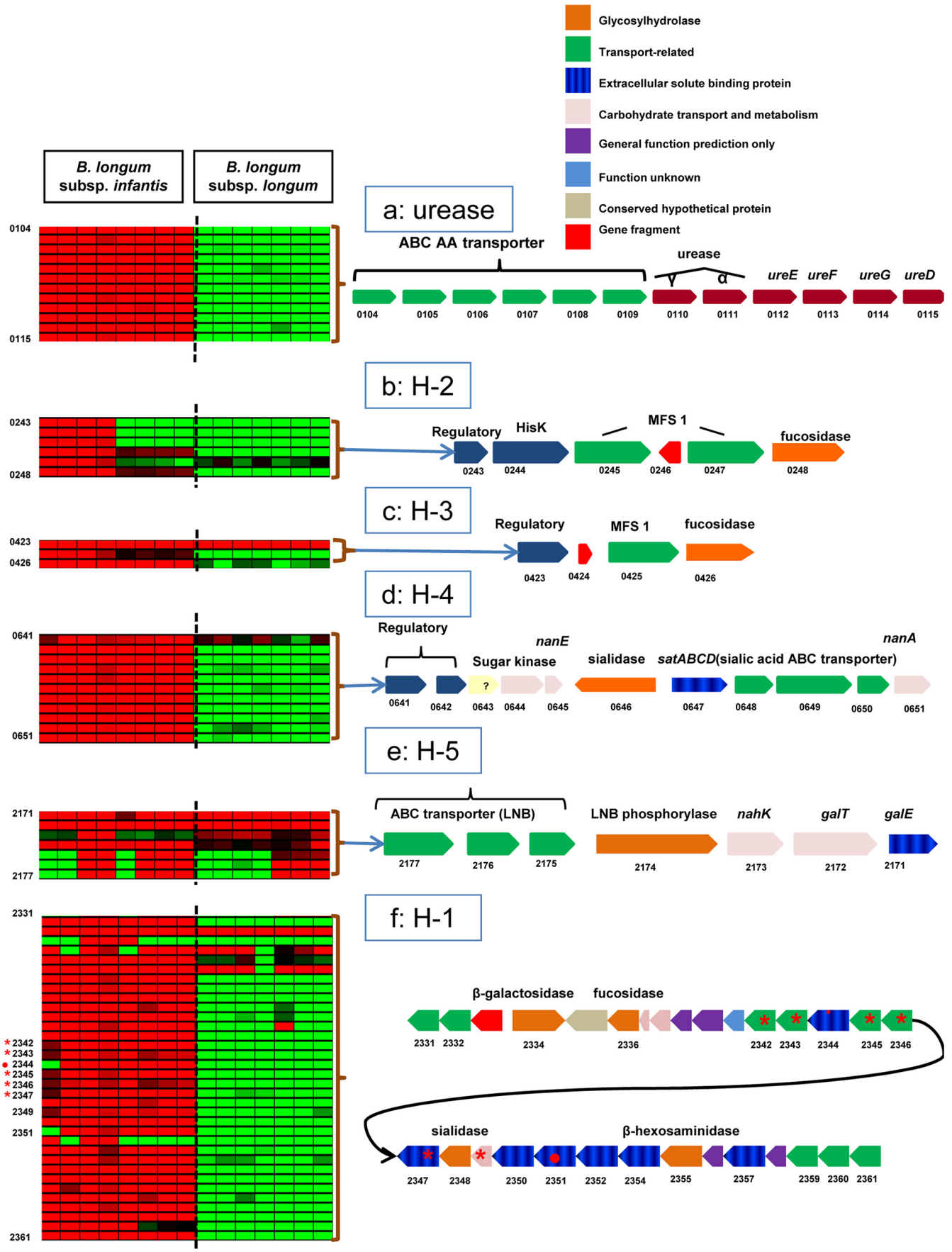
CGH analysis of HMO utilization genes. A distinctive feature of the *B. longum* subsp. *infantis* ATCC 15697 genome is the presence and conservation of five distinct gene clusters (Fig. 2) with a predicted capacity to bind, cleave, and transport free milk sugars (40). CGH provided a platform to associate the HMO^+, $HMO^\pm$, or HMO^- phenotype with the genome-wide distribution of milk utilization genes among *B. longum* strains.

The largest of these clusters, H-1 (Blon_2331 to Blon_2361), contains 30 contiguous genes annotated as sialidases, fucosidases, β-galactosidases, SBPs, and ABC transport permeases. Their colocation as a single 43-kbp fragment suggests the possible involvement of a coregulation mechanism controlling the binding, import, and deconstruction of oligosaccharide milk components (40). With the exception of two SBPs (Blon_2351 and Blon_2352), the H-1 cluster is conserved in all *B. longum* subsp. *infantis* HMO^+ strains. Conversely, with the exception of a sugar transporter (Blon_2331) and a glycosyl hydrolase (Blon_2334), the H-1 cluster is absent in all $HMO^\pm$ and HMO^- *B. longum* subsp. *longum* strains. Association of the H-1 cluster with HMO^+ *B. longum* subsp. *infantis* strains and their absence in $HMO^\pm$ and HMO^- *B. longum* subsp. *longum* strains provide corroboratory evidence that genes in this cluster are responsible for the utilization of HMO among *B. longum* strains.

Two fucosidase-containing gene clusters, H-2 (Blon_0243 to Blon_0218) and H-3 (Blon_0423 to Blon_0426), were conserved in four *B. longum* subsp. *infantis* strains. These clusters arose from a recent gene duplication event displacing arabinose metabolism genes in the ATTC 15697 type strain (40) and have rendered some *B. longum* subsp. *infantis* strains unable to metabolize this substrate (37). The fucosidase-containing cluster, H-3, was entirely conserved among all *B. longum* subsp. *infantis* strains, while H-2 was conserved in four strains (see Fig. S10 in the supplemental material). *B. longum* subsp. *infantis* strains retaining intact H-2 clusters were closely grouped by HCL analysis, suggesting that conservation of HMO genes has contributed to speciation events among *B. longum* strains. All *B. longum* subsp. *longum* strains tested in this study, including the sequenced strains ATCC 55813 and CCUG52486, lacked the H-2 and H-3 fucosidase-containing gene clusters. The absence of fucosidase-containing clusters in *B. longum* subsp. *longum* strains supports previous observations that fucosidase activity was absent in *B. longum* subsp. *longum* DJO10A (24). The fucosidase-containing gene clusters, H-2 and H-3, are conserved in *B. longum* subsp. *infantis* and are likely important for enabling growth on fucosylated milk oligosaccharides.

Sialic acid moieties are common in milk glycoproteins, in milk oligosaccharides (31), on the surface of the GI tract mucosa, and on pathogens and viruses (38). A complete sialic acid metabolic pathway was described in *B. longum* ATCC 15697, and sialidase activity was detected during growth on HMOs (24). The sialic acid cluster H-4 (Blon_0641 to Blon_0651) (Fig. 2) was conserved among *B. longum* subsp. *infantis* strains, suggesting the ability for this subspecies to bind and catabolize sialylated moieties.

Themetabolic pathway for lacto-*N*-biose [LNB; Gal(β1-3)GlcNac] has been characterized in *B. longum* JCM 1217 and associated with LNB metabolism and galacto-*N*-biose [GNB; Gal(β1-3)GalNac] in mucin sugars (19, 32). This seven-gene operon identified in ATCC 15697, region H-5 in Fig. 2 (Blon_2171 to Blon_2177) (40), was conserved in its entirety in


Glycosylhydrolase
Transport-related
Extracellular solute binding protein
Carbohydrate transport and metabolism
General function prediction only
Function unknown
Conserved hypothetical protein
Gene fragment
B. longum subsp. infantis
B. longum subsp. longum
a: urease
ABC AA transporter
urease
γ
α
ureE
ureF
ureG
ureD
0104
0105
0106
0107
0108
0109
0110
0111
0112
0113
0114
0115
b: H-2
Regulatory
HisK
MFS 1
fucosidase
0243
0244
0245
0246
0247
0248
c: H-3
Regulatory
MFS 1
fucosidase
0423
0424
0425
0426
d: H-4
Regulatory
nanE
Sugar kinase
sialidase
satABCD(sialic acid ABC transporter)
nanA
0641
0642
0643
0644
0645
0646
0647
0648
0649
0650
0651
e: H-5
ABC transporter (LNB)
LNB phosphorylase
nahK
galT
galE
2177
2176
2175
2174
2173
2172
2171
f: H-1
β-galactosidase
fucosidase
2331
2332
2334
2336
2342
2343
2344
2345
2346
sialidase
β-hexosaminidase
2347
2348
2350
2351
2352
2354
2355
2357
2359
2360
2361

eight additional *B. longum* strains. In seven strains with an incomplete LNB H-5 cluster, gene divergence occurred specifically in three family 1 SBPs (Blon_2175, Blon_2176, and Blon_2177), while the rest of the cluster, including the LNB phosphorylase (LnpA; Blon_2174) remained conserved across most strains. The H-5 LNB cluster is the only milk consumption-related cluster identified to be conserved in three *B. longum* subsp. *longum* and *B. longum* subsp. *suis* strains. Lacto-*N*-tetraose [LNT; Gal(β1-3)GlcNac(β1-3)Gal(β1-4)Glc] comprises 40% of the total amount of HMO in milk (24). Several bifidobacteria consume LNT, which is either transported entirely or catabolized via an LNB intermediate (24, 25). With the exception of the H-5 LNB cluster, the remaining HMO-related clusters (H-1, H-2, H-3, H-4) are either highly divergent or absent in $HMO^{\pm/-}$ *B. longum* subsp. *longum* strains, suggesting an important role of these gene regions for complete utilization of milk oligosaccharides other than LNT/LNB.

CGH analysis and sequencing of JCM 1260. Although *B. longum* subsp. *infantis* strains are characterized by an HMO^{+} phenotype, *B. longum* subsp. *infantis* JCM 1260 had a limited HMO utilization capacity (25) (Table 1). CDSs in the HMO cluster (H-1) associated with binding and transport of oligosaccharides have diverged in JCM 1260. Three SBPs (Blon_2342, Blon_2347, and Blon_2351) were divergent, and two other SBPs (Blon_2344 and Blon_2351) appeared to be absent in JCM 1260 (R-5 in Fig. S11 in the supplemental material). To further investigate the divergence in the JCM 1260 H-1 HMO cluster, the genome was sequenced. Of interest, sequencing reads were absent for four permease subunits (Blon_2342, Blon_2343, Blon2345, and Blon_2346) and one SBP (Blon_2351), and two additional SBPs (Blon_2344 and Blon_2347) were missing portions of the 3′ and 5′ flanking regions, respectively (see Fig. S13 in the supplemental material). The absence of these transport-related genes within the HMO cluster is consistent with the suboptimal growth exhibited by JCM 1260 when HMO is fermented as the sole carbon source. This suggests that the integrity of these ABC transporter genes is essential in the efficient utilization of milk oligosaccharides by *B. longum* subsp. *infantis*.

CGH analysis of other relevant regions. Human milk contains approximately only 15% of its nitrogen in the form of urea (8). It is therefore believed that total nitrogen in the infant lower GIT may be present at suboptimal levels. An increase in postnatal nitrogen levels is likely necessary to satisfy the growth and metabolism requirements of the infant and the GIT microbiota. Urea nitrogen salvaging (UNS; reviewed in reference 13) is a bacterially mediated process, whereas colonic urea is made available to the GIT microbiota while also generating ammonia to satisfy the host nitrogen needs. *B. longum* subsp. *infantis* ATCC 15697 contains the complete gene cluster for urea uptake and metabolism (Blon_0104 to Blon_0115) (40). CGH analysis and urease assays established that the urease gene cluster and its activity are conserved in all *B. longum* subsp. *infantis* and absent in all *B. longum* subsp. *longum* strains (Fig. 2 and Table 1).

Bacterial ABC transporter systems are utilized for oligosaccharide and oligopeptide binding and transport into the cytosol. These systems normally contain an SBP component binding the target substrates and facilitate their transfer to membrane-associated components for intracellular transport (16). The *B. longum* subsp. *infantis* ATCC 15697 genome contains 20 copies of family 1 SBPs (SBP family 1; pfam01547) and, after *B. dentium* ATTC 27678, has the second-highest number of copies of family 1 SBPs among bifidobacteria sequenced to date (26). Comparatively, *B. longum* subsp. *longum* DJO10A and *B. longum* subsp. *longum* ATTC 55813 and CCUG 52486 possess 15, 14, and 13 family 1 SBPs, respectively (26). Glycoconjugates are found on mucins on the gut epithelium and as free or bound components of human milk components. Therefore, the distribution and specificity of SBPs in a given organism have the potential of influencing colonization and proliferation in the infant GIT.

CGH data and COGs enrichment analysis illustrate that family 1 SBPs are conserved in HMO^{+} *B. longum* subsp. *infantis* strains and are largely absent in $HMO^{\pm/-}$ *B. longum* subsp. *longum* strains (see Fig. S12 in the supplemental material). Specifically, among eight *B. longum* subsp. *infantis* strains in this analysis, four have conserved all but one SBP (Blon_2352), while JCM 1260, a *B. longum* subsp. *infantis* strain with limited HMO utilization ability, is missing four additional SBPs, three of which belong to the H-1 HMO gene cluster and are predicted to bind oligosaccharides (Blon_2344, Blon_2347, and Blon_2351). Proteomic analysis demonstrated that these three SBPs are expressed in *B. longum* subsp. *infantis* ATCC 15697 during HMO fermentation (40). Two SBPs (Blon_2444 and Blon_2458) are conserved across all strains, and analysis of the surrounding gene regions suggested that these SBPs are involved in the potential uptake of amino acids and carbohydrate/oligosaccharide transport, respectively.

DISCUSSION

The establishment of bifidobacterium-rich microbial communities in the infant GIT has been associated with milk-borne factors (10, 20). Previous studies demonstrated that HMO utilization is not equally observed across all bifidobacteria and varies within the *B. longum* clade. Among closely related *B. longum* strains, robust growth on HMOs has been observed mainly in *B. longum* subsp. *infantis* strains (25). The metabolic strategy adopted by this subspecies suggests that it coevolved with milk components as a means to outcompete colonization among other members of the infant microbial consortium. The notion of coevolution between nutrients in milk and GIT commensals is reflected in the genome of *B. longum* subsp. *infantis* ATCC 15697 (40), the archetype HMO^{+} strain containing

FIG. 2. CGH analysis of loci related to milk glycans utilization and gut adaptation. Strains are grouped hierarchically on the basis of probe intensities. (a) Urease operon; (b and c) putative fucose utilization regions (H-2 and H-3); (d) putative sialic acid utilization region (H-4); (e) LNB metabolism gene region (H-5); (f) a 43-kb cluster associated with utilization of human milk oligosaccharides (H-1). ●, locus absent; ∗, locus highly divergent in strain JCM 7010.

several gene regions intended for milk utilization. Conversely, *B. longum* subsp. *longum* DJO10A and *B. longum* subsp. *longum* NCC2705 have largely maintained the metabolic machinery required for utilization of plant-based carbon sources (40). The conservation of genes in *B. longum* subsp. *infantis* providing a metabolic capacity toward milk carbon is a genotypic biomarker linking the adaptation of this subspecies to a unique nutrient ecological niche.

CGH was used to query the distribution of ATCC 15697 genes and, specifically, the milk utilization and GIT adaptation regions across 14 additional *B. longum* strains. Putative HMO utilization gene regions have remained conserved across all *B. longum* subsp. *infantis* strains with HMO^+ phenotypes and have diverged in *B. longum* subsp. *longum* strains with HMO^- phenotypes. These regions contain genes annotated as fucosidases, sialidases, glycosyl hydrolases, ABC transporters, and family 1 SBPs and are likely needed for efficient metabolism of HMOs.

The H-5 LNB cluster was partially conserved among *B. longum* subsp. *longum* strains, and some of its genes are upregulated during growth on human milk, formula milk supplemented with galacto-oligosaccharides, and plant oligosaccharides (14). *B. longum* subsp. *longum* strains with an $HMO^{\pm/-}$ phenotype may target LNT metabolism through the H-5 LNB cluster and with other glycosyl transferases, family 1 SBPs, and transporters absent in ATCC 15697 and are therefore not traceable in this study. Lacto-*N*-biosidase activity and the LNB H-5 clusters were detected in *B. longum* subsp. *longum* and *B. bifidum* strains (55) and in the infant gut metagenome (21). LNT metabolism may enable these bifidobacteria to achieve basal growth on HMO, as observed by their characteristic colonization to the infant gut microbiome.

The full conservation of urease activity and its relative gene cluster in *B. longum* subsp. *infantis* strains suggests a microbiome-driven role in supplementing the host nitrogen requirements. The tripartite relationship established between nitrogen-containing milk components, a urease-producing microbiota, and its host is yet another example of a systemwide strategy chosen to guide early colonization of the infant GIT by a selected and highly specialized microbial consortia.

Another important feature characteristic of *B. longum* subsp. *infantis* strains is the frequent conservation of numerous family 1 SBPs predicted to bind oligosaccharides. Glycoconjugates on milk proteins and HMOs share many common motifs with epithelial GIT mucins. It is possible that bifidobacteria and, in particular, *B. longum* subsp. *infantis* possess SBPs capable of binding both milk and mucin-bound glycans, serving a dual function needed for carbon sequestration and attachment to epithelial cell surfaces.

CGH analysis of 15 *B. longum* strains has determined that *B. longum* subsp. *infantis* has preserved the predicted gene regions and metabolic capacity toward utilization of milk-borne nutrients and adaptation to the infant GIT. *B. longum* subsp. *longum* strains, although taxonomically similar to *B. longum* subsp. *infantis*, lack these gene regions and exhibit an $HMO^{\pm}$ or HMO^- phenotype. In the absence of effective genetic tools amenable to bifidobacteria, CGH can provide important clues about the association between genotypic biomakers and phenotypes.

The comparative genomics data and MLST-based taxonomy presented here support a notion set forth by Sela et al. (40) that *B. longum* has at least two distinct subspecies: *B. longum* subsp. *infantis*, adapted to utilize milk carbon in the infant GIT, and *B. longum* subsp. *longum*, specialized for plant-derived carbon metabolism and associated with the adult GIT. A larger MLST study has recently suggested that *B. longum* subsp. *infantis* has emerged as a lineage from within *B. longum* subsp. *longum* (7). The reclassification of four strains previously typed using 16S rRNA gene sequences and enzymatic assays (37, 42, 56) indicates that MLST should be employed to correctly type *B. longum* organisms at the subspecies level.

Human colostrum and early milk contain large amounts of free oligosaccharides and glycoproteins. Although early gut colonization is likely dependent on a multitude of dietary and nondietary factors, the delivery of complex oligosaccharides through milk creates an ideal and unique nutrient niche for the establishment of, and colonization by, *B. longum* subsp. *infantis* strains. The succession of infant GIT communities correlates with the functional needs dictated by the host diet composition. During weaning, a gradual transitioning from milk-based to plant-based diets generates a shift in carbon availability in the GIT favorable for the expansion and formation of an adult-like GIT microbiota.

ACKNOWLEDGMENTS

R.G.L. was supported in part by a UC—Davis biotechnology training grant. D.A.S. was supported in part by a predoctoral training grant (NIH-NIGMS T32-GM08799). This publication was made possible by grant support from the University of California Discovery Grant Program, the California Dairy Research Foundation, USDA NRI-CSREES award 2008-35200-18776, Dairy Management Inc., USDA CSREES 2006-34526-17001 (to B.W.), and NIH-NICID awards 5R01HD059127 and 1R01HD061923.

We thank A. Adamson for technical assistance.

REFERENCES

1. **Ahrens, K.** 2009. Consumption of human milk oligosaccharides by Bifidobacterium. MS thesis. University of California—Davis, Davis, CA.
2. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.
3. **Backhed, F., R. E. Ley, J. L. Sonnenburg, D. A. Peterson, and J. I. Gordon.** 2005. Host-bacterial mutualism in the human intestine. Science **307:**1915–1920.
4. **Cai, H., B. T. Rodriguez, W. Zhang, J. R. Broadbent, and J. L. Steele.** 2007. Genotypic and phenotypic characterization of Lactobacillus casei strains isolated from different ecological niches suggests frequent recombination and niche specificity. Microbiology **153:**2655–2665.
5. **Chaturvedi, P., C. D. Warren, C. R. Buescher, L. K. Pickering, and D. S. Newburg.** 2001. Survival of human milk oligosaccharides in the intestine of infants, p. 315–323. *In* D. S. Newburg (ed.), Bioactive components of human milk. Springer, New York, NY.
6. **Coppa, G. V., S. Bruni, L. Morelli, S. Soldi, and O. Gabrielli.** 2004. The first prebiotics in humans—human milk oligosaccharides. J. Clin. Gastroenterol. **38:**S80–S83.
7. **DelÈtoile, A., V. Passet, J. Aires, I. Chambaud, M.-J. Butel, T. Smokvina, and S. Brisse.** 2010. Species delineation and clonal diversity in four Bifidobacterium species as revealed by multilocus sequencing. Res. Microbiol. **161:**82–90.
8. **Donovan, S. M., and B. Lonnerdal.** 1989. Non-protein nitrogen and true protein in infant formulas. Acta Paediatr. Scand. **78:**497–504.
9. **Eisen, M. B., P. T. Spellman, P. O. Brown, and D. Botstein.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U. S. A. **95:**14863–14868.
10. **Favier, C. F., W. M. de Vos, and A. D. L. Akkermans.** 2003. Development of bacterial and bifidobacterial communities in feces of newborn babies. Anaerobe **9:**219–229.
11. **Favier, C. F., E. E. Vaughan, W. M. De Vos, and A. D. L. Akkermans.** 2002. Molecular monitoring of succession of bacterial communities in human neonates. Appl. Environ. Microbiol. **68:**219–226.
12. **Felsenstein, J.** 1985. Confidence-limits on phylogenies—an approach using the bootstrap. Evolution **39:**783–791.

13. **Fuller, M. F., and P. J. Reeds.** 1998. Nitrogen cycling in the gut. Annu. Rev. Nutr. **18:**385–411.
14. **Gonzalez, R., E. S. Klaassens, E. Malinen, W. M. de Vos, and E. E. Vaughan.** 2008. Differential transcriptional response of *Bifidobacterium longum* to human milk, formula milk, and galactooligosaccharide. Appl. Environ. Microbiol. **74:**4686–4694.
15. **Herwig, R., A. J. Poustka, C. Muller, C. Bull, H. Lehrach, and J. O'Brien.** 1999. Large-scale clustering of cDNA-fingerprinting data. Genome Res. **9:**1093–1105.
16. **Higgins, C. F.** 2001. ABC transporters: physiology, structure and mechanism—an overview. Res. Microbiol. **152:**205–210.
17. **Huson, D. H., and D. Bryant.** 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23:**254–267.
18. **Irizarry, R. A., B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed.** 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics **4:**249–264.
19. **Kitaoka, M., J. S. Tian, and M. Nishimoto.** 2005. Novel putative galactose operon involving lacto-*N*-biose phosphorylase in *Bifidobacterium longum*. Appl. Environ. Microbiol. **71:**3158–3162.
20. **Kleessen, B., H. Bunke, K. Tovar, J. Noack, and G. Sawatzki.** 1995. Influence of two infant formulas and human milk on the development of the faecal flora in newborn infants. Acta Paediatr. **84:**1347–1356.
21. **Kurokawa, K., T. Itoh, T. Kuwahara, K. Oshima, H. Toh, A. Toyoda, H. Takami, H. Morita, V. K. Sharma, T. P. Srivastava, T. D. Taylor, H. Noguchi, H. Mori, Y. Ogura, D. S. Ehrlich, K. Itoh, T. Takagi, Y. Sakaki, T. Hayashi, and M. Hattori.** 2007. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res. **14:**169–181.
22. **Lee, J. H., V. N. Karamychev, S. A. Kozyavkin, D. Mills, A. R. Pavlov, N. V. Pavlova, N. N. Polouchine, P. M. Richardson, V. V. Shakhova, A. I. Slesarev, B. Weimer, and D. J. O'Sullivan.** 2008. Comparative genomic analysis of the gut bacterium Bifidobacterium longum reveals loci susceptible to deletion during pure culture growth. BMC Genomics **9:**247.
23. **Lievin, V., I. Peiffer, S. Hudault, F. Rochat, D. Brassart, J. R. Neeser, and A. L. Servin.** 2000. Bifidobacterium strains from resident infant human gastrointestinal microflora exert antimicrobial activity. Gut **47:**646–652.
24. **LoCascio, R. G., M. R. Ninonuevo, S. L. Freeman, D. A. Sela, R. Grimm, C. B. Lebrilla, D. A. Mills, and J. B. German.** 2007. Glycoprofiling of bifidobacterial consumption of human milk oligosaccharides demonstrates strain specific, preferential consumption of small chain glycans secreted in early human lactation. J. Agric. Food Chem. **55:**8914–8919.
25. **LoCascio, R. G., M. R. Niñonuevo, S. R. Kronewitter, S. L. Freeman, J. B. German, C. B. Lebrilla, and D. A. Mills.** 2009. A versatile and scalable strategy for glycoprofiling bifidobacterial consumption of human milk oligosaccharides. Microb. Biotechnol. **2:**333–342.
26. **Markowitz, V. M., E. Szeto, K. Palaniappan, Y. Grechkin, K. Chu, I. M. Chen, I. Dubchak, I. Anderson, A. Lykidis, K. Mavromatis, N. N. Ivanova, and N. C. Kyrpides.** 2008. The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. Nucleic Acids Res. **36:**D528–D533.
27. **Mattarelli, P., C. Bonaparte, B. Pot, and B. Biavatil.** 2008. Proposal to reclassify the three biotypes of Bifidobacterium longum as three subspecies: Bifidobacterium longum subsp. longum subsp. nov., Bifidobacterium longum subsp. infantis comb. nov. and Bifidobacterium longum subsp. suis comb. nov. Int. J. Syst. Evol. Microbiol. **58:**767–772.
28. **Mazmanian, S. K., C. H. Liu, A. O. Tzianabos, and D. L. Kasper.** 2005. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. Cell **122:**107–118.
29. **Mootha, V. K., C. M. Lindgren, K. F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstrale, E. Laurila, N. Houstis, M. J. Daly, N. Patterson, J. P. Mesirov, T. R. Golub, P. Tamayo, B. Spiegelman, E. S. Lander, J. N. Hirschhorn, D. Altshuler, and L. C. Groop.** 2003. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. **34:**267–273.
30. **Newburg, D. S., and W. A. Walker.** 2007. Protection of the neonate by the innate immune system of developing gut and of human milk. Pediatr. Res. **61:**2–8.
31. **Ninonuevo, M. R., Y. Park, H. Yin, J. Zhang, R. E. Ward, B. H. Clowers, J. B. German, S. L. Freeman, K. Killeen, R. Grimm, and C. B. Lebrilla.** 2006. A strategy for annotating the human milk glycome. J. Agric. Food Chem. **54:**7471–7480.
32. **Nishimoto, M., and M. Kitaoka.** 2007. Identification of *N*-acetylhexosamine 1-kinase in the complete lacto-*N*-biose I/galacto-*N*-biose metabolic pathway in *Bifidobacterium longum*. Appl. Environ. Microbiol. **73:**6444–6449.
33. **Ouwehand, A. C., S. Salminen, and E. Isolauri.** 2002. Probiotics: an overview of beneficial effects. Antonie Van Leeuwenhoek Int. J. Gen. Mol. Microbiol. **82:**279–289.
34. **Paustian, M. L., X. Zhu, S. Sreevatsan, S. Robbe-Austerman, V. Kapur, and J. P. Bannantine.** 2008. Comparative genomic analysis of Mycobacterium avium subspecies obtained from multiple host species. BMC Genomics **9:**135.
35. **Rzhetsky, A., and M. Nei.** 1992. Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. J. Mol. Evol. **35:**367–375.
36. **Saeed, A. I., V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, and J. Quackenbush.** 2003. TM4: a free, open-source system for microarray data management and analysis. Biotechniques **34:**374–378.
37. **Sakata, S., M. Kitahara, M. Sakamoto, H. Hayashi, M. Fukuyama, and Y. Benno.** 2002. Unification of Bifidobacterium infantis and Bifidobacterium suis as Bifidobacterium longum. Int. J. Syst. Evol. Microbiol. **52:**1945–1951.
38. **Schauer, R., S. Kelm, G. Reuter, P. Roggentin, and L. Shaw.** 1995. Biochemistry and role of sialic acids, p. 12–34. *In* A. Rosenberg (ed.), Biology of sialic acids. Springer, New York, NY.
39. **Schell, M. A., M. Karmirantzou, B. Snel, D. Vilanova, B. Berger, G. Pessi, M. C. Zwahlen, F. Desiere, P. Bork, M. Delley, R. D. Pridmore, and F. Arigoni.** 2002. The genome sequence of Bifidobacterium longum reflects its adaptation to the human gastrointestinal tract. Proc. Natl. Acad. Sci. U. S. A. **99:**14422–14427.
40. **Sela, D. A., J. Chapman, A. Adeuya, J. H. Kim, F. Chen, T. R. Whitehead, A. Lapidus, D. S. Rokhsar, C. B. Lebrilla, J. B. German, N. P. Price, P. M. Richardson, and D. A. Mills.** 2008. The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. Proc. Natl. Acad. Sci. U. S. A. **105:**18964–18969.
41. **Servin, A. L.** 2004. Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. FEMS Microbiol. Rev. **28:**405–440.
42. **Silvi, S., M. C. Verdenelli, C. Orpianesi, and A. Cresci.** 2003. EU project Crownalife: functional foods, gut microflora and healthy ageing: isolation and identification of Lactobacillus and Bifidobacterium strains from faecal samples of elderly subjects for a possible probiotic use in functional foods. J. Food Eng. **56:**195–200.
43. **Snipen, L., O. L. Nyquist, M. Solheim, A. Aakra, and I. F. Nes.** 2009. Improved analysis of bacterial CGH data beyond the log-ratio paradigm. BMC Bioinformatics **10:**91.
44. **Soukas, A., P. Cohen, N. D. Socci, and J. M. Friedman.** 2000. Leptin-specific patterns of gene expression in white adipose tissue. Genes Dev. **14:**963–980.
45. **Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov.** 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. **102:**15545–15550.
46. **Taboada, E. N., R. R. Acedillo, C. C. Luebbert, W. A. Findlay, and J. H. Nash.** 2005. A new approach for the analysis of bacterial microarray-based comparative genomic hybridization: insights from an empirical study. BMC Genomics **6:**78.
47. **Takezaki, N., A. Rzhetsky, and M. Nei.** 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. **12:**823–833.
48. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596–1599.
49. **Tamura, K., M. Nei, and S. Kumar.** 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. U. S. A. **101:**11030–11035.
50. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
51. **Turnbaugh, P. J., M. Hamady, T. Yatsunenko, B. L. Cantarel, A. Duncan, R. E. Ley, M. L. Sogin, W. J. Jones, B. A. Roe, J. P. Affourtit, M. Egholm, B. Henrissat, A. C. Heath, R. Knight, and J. I. Gordon.** 2009. A core gut microbiome in obese and lean twins. Nature **457:**480–484.
52. **Turnbaugh, P. J., R. E. Ley, M. A. Mahowald, V. Magrini, E. R. Mardis, and J. I. Gordon.** 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature **444:**1027–1031.
53. **Tusher, V. G., R. Tibshirani, and G. Chu.** 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. U. S. A. **98:**5116–5121.
54. **Ventura, M., C. Canchaya, A. Del Casale, F. Dellaglio, E. Neviani, G. F. Fitzgerald, and D. van Sinderen.** 2006. Analysis of bifidobacterial evolution using a multilocus approach. Int. J. Syst. Evol. Microbiol. **56:**2783–2792.
55. **Wada, J., T. Ando, M. Kiyohara, H. Ashida, M. Kitaoka, M. Yamaguchi, H. Kumagai, T. Katayama, and K. Yamamoto.** 2008. *Bifidobacterium bifidum* lacto-*N*-biosidase, a critical enzyme for the degradation of human milk oligosaccharides with a type 1 structure. Appl. Environ. Microbiol. **74:**3996–4004.
56. **Yang, Y.-S.** January 1998. Bifidobacteria strains with acid, bile salt and oxygen tolerance and their culture method. U.S. patent 5711977.
57. **Zotta, T., A. Ricciardi, R. Rossano, and E. Parente.** 2008. Urease production by Streptococcus thermophilus. Food Microbiol. **25:**113–119.