

Exploration of the Genomic Diversity and Core Genome of the *Bifidobacterium adolescentis* Phylogenetic Group by Means of a Polyphasic Approach

Sabrina Duranti,^a Francesca Turroni,^b Christian Milani,^a Elena Foroni,^a Francesca Bottacini,^b Fabio Dal Bello,^c Alberto Ferrarini,^d Massimo Delle Donne,^d Douwe van Sinderen,^b Marco Ventura^a

Laboratory of Probiogenomics, Department of Genetics, Biology of Microorganisms, Anthropology and Evolution, University of Parma, Parma, Italy^a; Alimentary Pharmabiotic Centre and Department of Microbiology, Bioscience Institute, National University of Ireland, Cork, Ireland^b; Sacco Srl, Cadorago, Italy^c; Functional Genomic Center, Department of Biotechnology, University of Verona, Verona, Italy^d

In the current work, we describe genome diversity and core genome sequences among representatives of three bifidobacterial species, i.e., *Bifidobacterium adolescentis*, *Bifidobacterium catenulatum*, and *Bifidobacterium pseudocatenulatum*, by employing a polyphasic approach involving analysis of 16S rRNA gene and 16S-23S internal transcribed spacer (ITS) sequences, pulsed-field gel electrophoresis (PFGE), and comparative genomic hybridization (CGH) assays.

Bifidobacteria are high-G+C-content Gram-positive bacteria that are commonly found in the gastrointestinal tract (GIT) of humans, animals, and insects (1–3). Their presence in the human GIT has been associated with beneficial health effects, which has led to the widespread use of various bifidobacterial strains in health-promoting or probiotic foods (3). However, despite their generally accepted importance as probiotic components of the human GIT microbiota, there is a paucity of information about the physiology, phylogenetic relationships, and underlying genetics of bifidobacteria (4).

Bacterial genome sequencing has opened up a new era of biological investigation, which can shed light on the interactive genetics underlying all microbial properties. However, bifidobacterial genomes appear underexplored in this respect, since from a total of over 40 currently recognized species of this genus, only seven genomes have been completely decoded (5–14). Bifidobacteria are receiving ever-increasing levels of scientific interest as health-promoting bacteria, in particular, species commonly identified in the human adult intestine, such as those belonging to the *Bifidobacterium adolescentis* group, i.e., *Bifidobacterium adolescentis*, *Bifidobacterium catenulatum*, and *Bifidobacterium pseudocatenulatum* (15). However, members of the *B. adolescentis* group have recently been isolated in infants (16), suggesting that members of this taxon are able to adapt to different ecological niches. Various studies have explored the genetic adaptation of strains belonging to the *B. adolescentis* group (14, 17). However, no detailed information is available on the genomic variability of the main species of this phylogenetic group that might serve as a reference for the identification of strain-specific features. In this study, we explored the genomic diversity of strains belonging to *B. adolescentis*, *B. catenulatum*, and *B. pseudocatenulatum* by applying a polyphasic approach that consisted of 16S rRNA gene sequencing, 16S-23S internal transcribed spacer (ITS) sequence analysis, and DNA typing, complemented by a genome-based approach involving comparative genomic hybridization (CGH) analysis.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *Bifidobacterium* strains were cultivated in an anaerobic atmosphere (2.99% H₂, 17.01% CO₂, and 80%

N₂) in a chamber (Concept 400; Ruskin) on De Man-Rogosa-Sharpe (MRS) (Scharlau Chemie, Barcelona, Spain) medium supplemented with 0.05% (wt/vol) L-cysteine hydrochloride and incubated at 37°C.

PCR amplification and phylogenetic analyses. PCR was used to amplify part of the 16S rRNA gene of investigated *Bifidobacterium* strains using primers P0 (5'-GAAGAGTTTGATCCTGGCTCAG-3') and P6 (5'-CTACGGCTACCTTGTACGA-3') (18), while DNA fragments corresponding to the 16S-ITS region were amplified using the oligonucleotides BIF-specific (5'-GGTGTGAAAGTCCATCGCT-3') and 23S_bif (5'-GTCTGCCAAGGCATCCACCA-3') (15). Phylogenetic analysis and trees were calculated using the PHYLIP software package, version 3.5c (19). Bootstrap values were computed by 100 random resamplings.

PFGE. The pulsed-field gel electrophoresis (PFGE) protocol used was as previously described (22).

CGH microarray, description, labeling, hybridizations, data acquisition, and treatment. CGH analysis was performed with microarrays that were based on the genome sequences of *B. adolescentis* ATCC 15703 (GenBank accession number NC_008618) and *B. pseudocatenulatum* DSM20438 (ABXX00000000.2). A total of 39,249 probes of 35 bp in length were designed using OligoArray 2.1 software (23). Oligonucleotides were synthesized in triplicate on a 2-by-40,000 CombiMatrix array (CombiMatrix, Irvine, CA). Replicates were distributed on the chip at random, nonadjacent positions. A set of 74 negative-control probes designed on phage and plant sequences was also included on the chip. Seventeen micrograms of purified genomic DNA was labeled with Cy5-ULS using the Kreatech ULS array CGH labeling kit (Kreatech Diagnostics) according to the supplier's instructions. Hybridization of labeled test DNA to these microarrays was performed according to CombiMatrix protocols (13). Fluorescence scanning was performed employing a MicroArray scanner (InnoScan 700). CGH data were processed as previously described (13, 14). Hierarchical clustering was performed with average linkage and Euclidean distance (24) using TMEV 4.0 software (13).

Received 8 August 2012 Accepted 9 October 2012

Published ahead of print 12 October 2012

Address correspondence to Marco Ventura, marco.ventura@unipr.it. S.D. and F.T. contributed equally.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.02467-12>.

Copyright © 2013, American Society for Microbiology. All Rights Reserved. doi:10.1128/AEM.02467-12

TABLE 1 *Bifidobacterium* strains used in this study

Species	Strain	Origin
<i>B. adolescentis</i>	ATCC 15703	Intestine of adult
<i>B. adolescentis</i>	125B	Infant feces
<i>B. adolescentis</i>	369B	Infant feces
<i>B. adolescentis</i>	153B	Infant feces
<i>B. adolescentis</i>	703B	Adolescent feces
<i>B. adolescentis</i>	145B	Infant feces
<i>B. adolescentis</i>	22L	Human milk
<i>B. pseudocatenulatum</i>	DSM20438	Infant feces
<i>B. pseudocatenulatum</i>	249B	Infant feces
<i>B. pseudocatenulatum</i>	202B	Human intestine
<i>B. pseudocatenulatum</i>	318B	Adolescent intestine
<i>B. catenulatum</i>	853B	Adolescent intestine
<i>B. catenulatum</i>	DSM16992	Intestine of adult
<i>B. catenulatum</i>	660-1	Human gut
<i>B. catenulatum</i>	660-7	Human gut
<i>B. catenulatum</i>	676	Human gut
<i>B. catenulatum</i>	MB271	Human gut

Bioinformatic analyses. Comparative analyses were performed using BLAST analysis (30) and MCL (graph-theory-based Markov clusters algorithm) clustering analysis (31).

Microarray data accession numbers. The microarray data have been deposited in the Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/>) under accession numbers GSE39050 and GSE39047.

RESULTS AND DISCUSSION

Phylogenetic analysis based on rRNA sequences. A phylogenetic analysis based on the 16S rRNA gene sequences of the 17 strains analyzed in this study (Table 1) was performed, while we also included relevant sequences of the type strains of other species of the genus *Bifidobacterium*. This analysis resulted in a phylogenetic tree (Fig. 1a), which was shown to be consistent with a previously described bifidobacterial taxonomic analysis (15). This analysis allowed species assignment of six isolates to *B. adolescentis*, three isolates to *B. pseudocatenulatum*, and six isolates to *B. catenulatum* (Fig. 1a; Table 1). Since the evolutionary rate of the 16S rRNA sequence in bifidobacteria is too low to delineate speciation among closely related taxa or for measuring intraspecies relationships, we decided to investigate a more variable rRNA-associated sequence, i.e., the 16S-23S ITS sequence, to assess intraspecific phylogenetic relationships in bifidobacteria. In fact, the high level of ITS sequence variation observed between members of the same bifidobacterial species (15, 20) renders this sequence a suitable molecular marker to infer phylogenetic distances between strains of the same species, as well as between closely related taxa. Thus, PCRs directed to amplify 16S-23S ITS sequences of all investigated bifidobacterial strains were carried out through the use of oligonucleotides BIF-specific and 23S_bif (15). The resulting 16S-23S ITS amplicons were subjected to DNA sequencing. The microvariability detected between the different ribosomal loci (16S and 16S-23S ITS sequences) occurring within each bacterial genome sequence had previously been shown not to affect the overall phylogenetic image determined by direct sequencing of the 16S- and 16S-23S ITS amplicons of various bifidobacterial or lactobacillus strains (15, 21, 22). Indeed, the phylogenetic clustering achieved using the ITS sequence-based tree is consistent with that obtained using 16S rRNA sequences (Fig. 1b), while it also provided a means to clearly distinguish all analyzed strains.

PFGE analyses. *B. adolescentis* and *B. pseudocatenulatum*/*B. catenulatum* DNA, when digested with SpeI, yielded fragments of approximately 230 kb to 20 kb and 300 kb to 20 kb, respectively (Fig. 2). According to the number and size of the fragments, we were able to distinguish seven different genomic fingerprints for both the *B. adolescentis* and *B. pseudocatenulatum*/*B. catenulatum* strains (Fig. 2a and b). The strain composition of each group is in good accordance with groupings achieved with the 16S-23S ITS sequence-based phylogenetic tree as well as with other typing tools employed. Altogether, these findings reinforce the notion that the strains used in this study are genetically different from each other and that their genomes may contain interesting genetic features that allow particular adaptations to the different ecological niches from which they were originally isolated.

***B. adolescentis* intraspecies differences as revealed by microarray analysis.** The degree of genetic diversity within a single species of the *B. adolescentis* group was further analyzed by CGH. In order to perform such an analysis, we used a DNA microarray design based on the publicly available genome sequences of *B. adolescentis* ATCC 15703 and *B. pseudocatenulatum* DSM20438.

Based on obtained hybridization signals, between 5% and 14% of the open reading frames (ORFs) identified in the *B. adolescentis* ATCC 15703 genome appeared to be absent in the six investigated *B. adolescentis* strains. These values are higher than those described for other bifidobacterial species such as *Bifidobacterium dentium* Bd1 (14) but similar to those for other bifidobacteria like *Bifidobacterium longum* subsp. *longum* (25). These findings suggest that the *B. adolescentis* genome is subject to relatively rapid diversification, as also observed in certain other enteric bifidobacteria (13). When projected on the genome map of ATCC 15703, the obtained CGH data highlight clusters of conserved ORFs but also genes/sequences that appear to be unique to ATCC 15703 (Fig. 3). These regions of genetic variability are randomly distributed along the genome, even if mostly located between the origin of replication and the terminus of replication in the clockwise direction. This characteristic was previously also noticed in *B. dentium* (14) and *Bifidobacterium bifidum* species (13).

According to the ATCC 15703 gene function annotation, the identified genomic diversity can be assigned to two classes: (i) mobile DNA that constitutes the *B. adolescentis* mobilome and (ii) plasticity regions of the *B. adolescentis* genome, which may underlie specific adaptations of the investigated strains and which may represent laterally acquired DNA or remnants of ancestral DNA that has not (yet) been lost. Detailed information about the variable regions identified in the genomes of *B. adolescentis* species is reported in Table S1 in the supplemental material. Plasticity regions which are preferred sites for acquisition of strain-specific DNA are well recognized in the genomes of pathogens such as *Helicobacter pylori* (26), where array-based CGH has been used to highlight regions involved in different pathogenic activities (27). Among the genome diversity regions that could be classified as the mobilome of *B. adolescentis* ATCC 15703, we found a presumed CRISPR element (BAD_1176-1186) and a possible integrated plasmid (BAD_1239-1291). Among the variable regions of the *B. adolescentis* CGH map, indicated as plasticity regions, we were also able to identify genes associated with bacterium-environment interactions and metabolic abilities. These include a number of regions predicted to be involved in the adaptation of ATCC 15703 to its ecological niche, such as two exopolysaccharide (*eps*) loci (*eps* I,

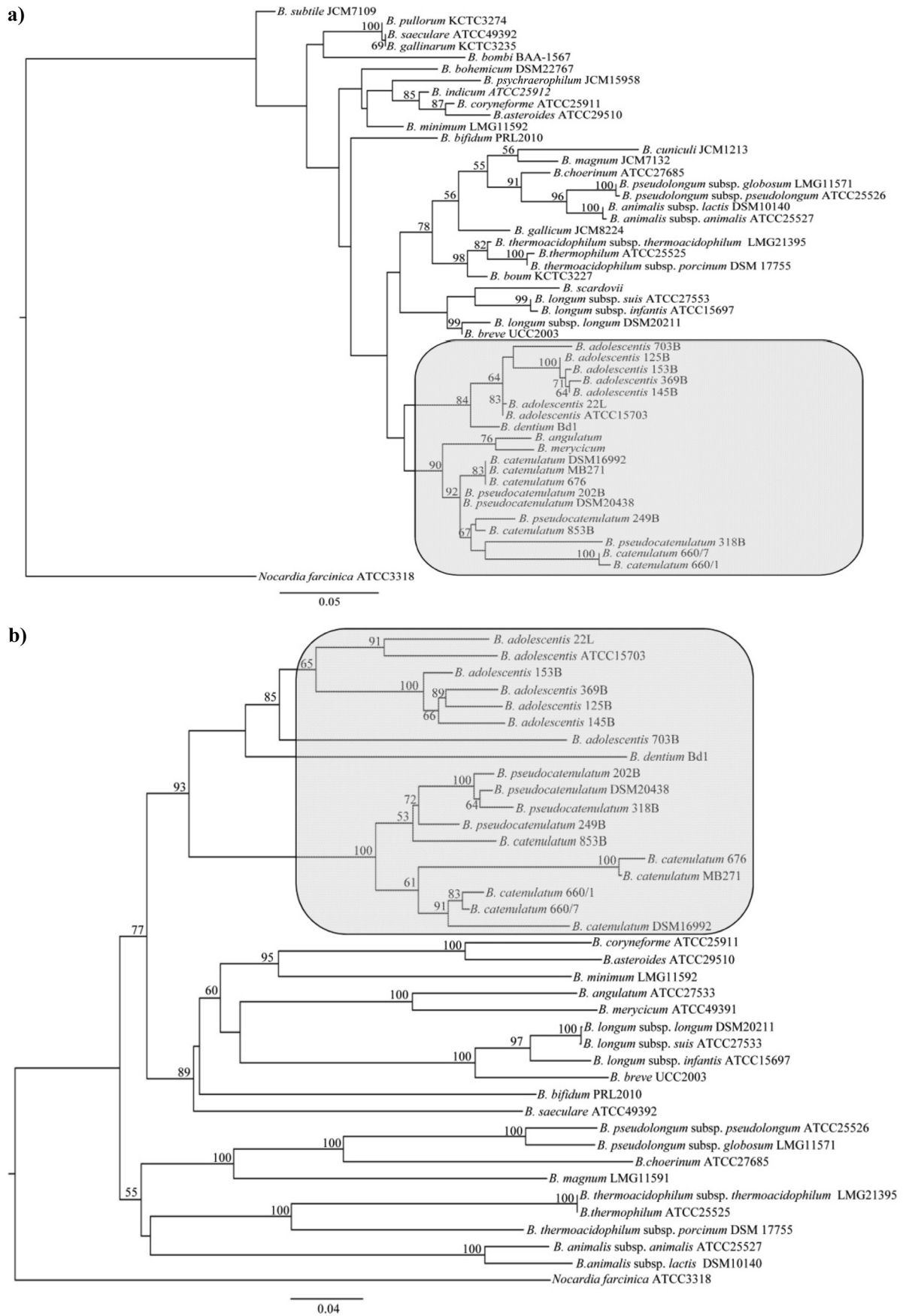


FIG 1 Phylogenetic tree of the genus *Bifidobacterium*, computed on the basis of (partial) 16S rRNA gene sequences (a) or ITS sequences (b). In panels a and b, the bar scale indicates phylogenetic distances. Bootstrap values are reported for a total of 100 replicates. The bacterial strains investigated in this study are highlighted in gray.

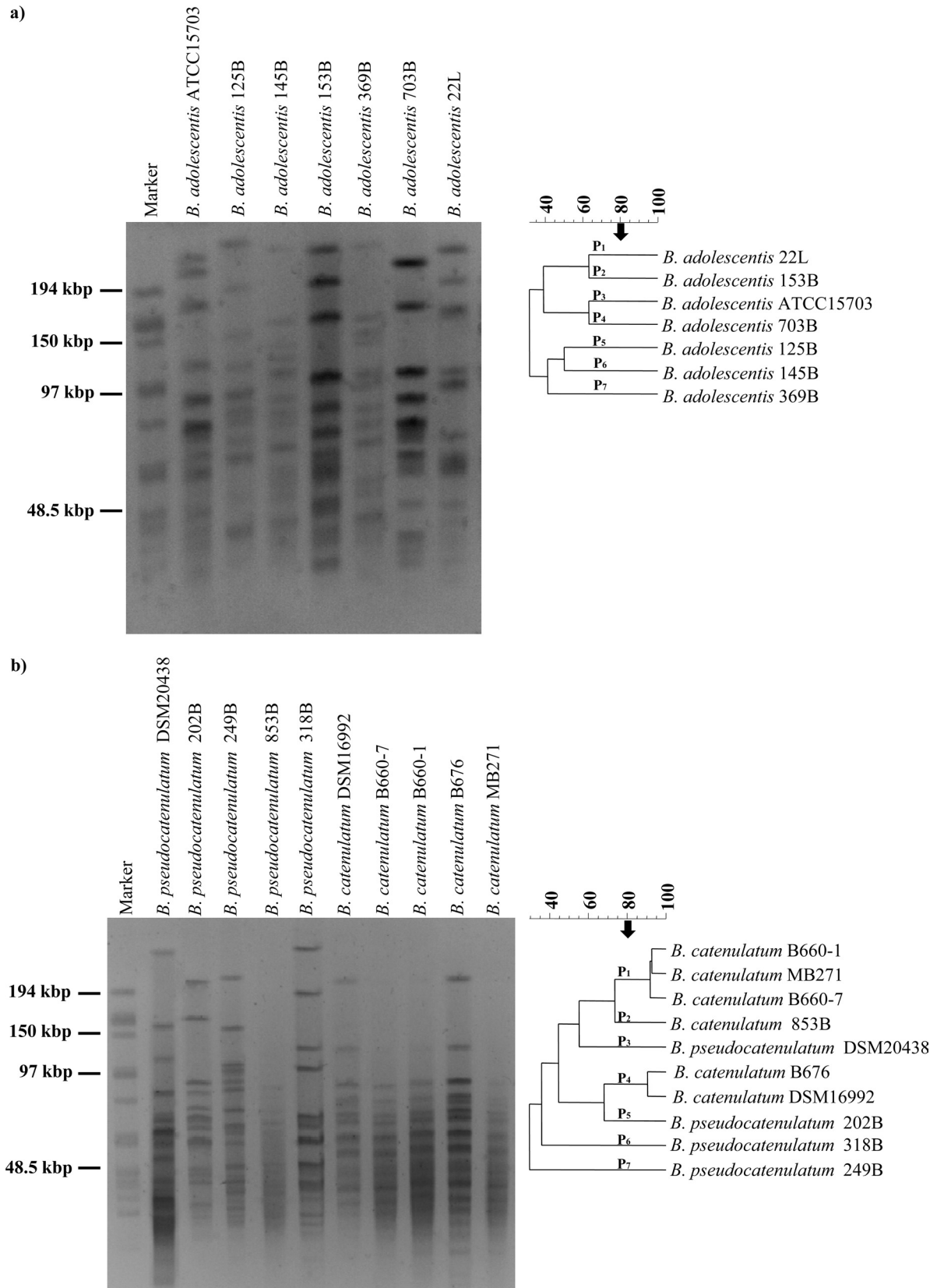


FIG 2 PFGE analysis of *B. adolescentis* (a) or *B. pseudocatenulatum*/*B. catenulatum* (b) strains as determined by the use of restriction enzyme *Spe*I. The strains analyzed are indicated on the right-hand margin of each panel. The arrow denotes the cutting level for separation clusters according to the UPGMA similarity values described previously (37).

BAD_1349-1398, and *eps* II, BAD_1485-1511), putative surface proteins (BAD_0393-0402), a predicted pilus-biosynthesis gene cluster (BAD_1467-1470), and membrane-associated transporters (BAD_0478-0489). The *eps* II locus is associated with genes that are predicted to synthesize dTDP-rhamnose (BAD_1507-BAD_1509) and represents the largest genome segment with substantial interstrain genetic variability. All genes identified on the basis of their variability at intraspecies level in *B. adolescentis* taxon were classified according to COG functional categories (<http://www.ncbi.nlm.nih.gov/COG>). A preliminary *in silico* prediction was first performed on the reference genome *B. adolescentis* ATCC 15703. CGH data were then used to identify the presence of enriched functional categories in each of the two analyses. Compared to their abundance as a functional category in the genome, the variable genome regions of *B. adolescentis* appear to be enriched in functions related to defense mechanisms against invading DNA (e.g., genes encoding a restriction-modification system and an Abi-like protein), cell wall biogenesis, and carbohydrate transport and metabolism (Fig. 4).

Based on these CGH data, it was possible to identify truly unique genes (TUG) of ATCC 15703, whose microarray probes failed to hybridize to any of the other *B. adolescentis* genomic DNA. Such TUG of ATCC 15703 encompass six ORFs, predominantly characterized by mobile elements (e.g., displaying a consistent deviation in the GC content of the reference genome), a predicted type II restriction-modification system (BAD_1280), various hypothetical proteins, and one glycosyltransferase (BAD_0975). Furthermore, 1,162 genes, representing approximately 73.6% of the *B. adolescentis* ATCC 15703 genome, were shown to be shared among all *B. adolescentis* strains tested, probably representing the core genome content of *B. adolescentis*. Other than genes associated with housekeeping functions, this deduced *B. adolescentis* core genome also encodes enzymes involved in carbohydrate metabolism and transport.

A simplified representation of the CGH data is presented in Fig. 3c. Raw CGH results of *B. adolescentis* strains come in pairs of intensity values connected to each *B. adolescentis* ATCC 15703 probe on the array. These values are expressed as mean ratios of the results from the test to the reference ATCC 15703 strain, normalized in such a manner that a ratio of one indicates the perfect conservation of the gene. Notably, a major peak is noted at the same position for all tested *B. adolescentis* strains, indicative of very similar DNA sequences (Fig. 3). In addition, a clustering of the microarray data was performed in order to extract qualitative information about the presence of each gene. The CGH-based information obtained in this study reinforces the notion of *B. adolescentis* being a moderately evolved bifidobacterial species, consistent with previously published CGH results concerning the *B. adolescentis* species (17). However, compared with the latter study, we have expanded knowledge on the identity of the plasticity regions present in representative genomes of this species, by including also *B. adolescentis* strains from human milk (Table 1).

***B. pseudocatenulatum*-*B. catenulatum* inter-/intraspecies differences as revealed by CGH analysis.** As previously reported, *B. catenulatum* and *B. pseudocatenulatum* are very closely related species, with high DNA homology of >88% (28). This high level of DNA sequence similarity allowed the investigation of genomic diversity of these species by CGH analysis using a microarray designed as described above. The number of ORFs of the spotted DSM20438-derived probes that failed to efficiently hybridize with

DNA of the nine investigated strains belonging to the *B. pseudocatenulatum* and *B. catenulatum* taxa (Table 1) was shown to range from 3% to 13%. The observed values that quantify the genomic divergence among tested strains of these two species are similar to those identified at intraspecific level for other bifidobacterial taxa such as *B. adolescentis* (see above), *B. bifidum* (13), and *B. longum* subsp. *longum* (25), which reinforces the notion of a very close phylogenetic relationship between the *B. catenulatum* and *B. pseudocatenulatum* species. Nevertheless, the obtained CGH profiles showed a clear distinction between members of the *B. catenulatum* and *B. pseudocatenulatum* species (Fig. 5a), also supported by the phylogenetic tree based upon these CGH scores, generating a separation in two main clusters that correspond to *B. catenulatum* and *B. pseudocatenulatum* species (Fig. 5b). According to the *B. pseudocatenulatum* DSM20438 genome annotation, the identified genomic diversity identified can be assigned to two functional classes, i.e., mobile DNA and plasticity regions, which might underline the specific ecological adaptation of the investigated strains. In the first group, it is possible to collocate the CRISPR locus (BIFPSEUDO_03113-BIFPSEUDO_03119) and *tetW* locus (BIFPSEUDO_03503-BIFPSEUDO_03506). DNA regions associated with bacterium-environment interaction include two sortase-dependent pilus loci (BIFPSEUDO_04099-BIFPSEUDO_04102 and BIFPSEUDO_04336-BIFPSEUDO_04338), as well as two *eps* loci (*eps* I, BIFPSEUDO_03257-BIFPSEUDO_03271, and *eps* II, BIFPSEUDO_04141-BIFPSEUDO_04153). The *eps* II locus of the DSM20438 strain is associated with the dTDP-rhamnose biosynthesis locus (BIFPSEUDO_04144-BIFPSEUDO_04146) and represents the largest genome segment with substantial interstrain genetic variability. Notably, *eps* I consists of DNA with a lower GC content (46%) than the average of the DSM20438 genome sequence (56%), indicating that this region was acquired by horizontal gene transfer (HGT). Further variable genes include a glycosyl hydrolase-encoding gene (BIFPSEUDO_02845) and three genes specifying a putative ABC-type carbohydrate-uptake system (BIFPSEUDO_02841-BIFPSEUDO_02843). BLAST searches identified that this ABC-type transporter system is homologous to an uptake system encoded by *B. bifidum* PRL2010, which has been linked to turanose uptake (BBPR_1353) (29). Detailed information on the variable regions identified in the genomes of *B. catenulatum*/*B. pseudocatenulatum* species is reported in Table S2 in the supplemental material.

Similarly to the TUG repertoire of *B. adolescentis* ATCC 15703, the TUG of *B. pseudocatenulatum* DSM20438 include three ORFs encoding two hypothetical proteins (BIFPSEUDO_03506 and BIFPSEUDO_03267) and a dihydrodipicolinate synthase (BIFPSEUDO_03274) displaying a significant deviation in the GC content and thus representing a member of the mobilome of this strain. In contrast, 1,579 ORFs were shown to be shared between the tested *B. pseudocatenulatum* strains and 1,431 genes are present in all tested *B. catenulatum* strains, while 1,355 of the latter genes are present in both species (Fig. 4). Such genes might represent the core genome of the *B. pseudocatenulatum*/*B. catenulatum* species. A functional analysis of this core genetic repertoire suggests the existence of shared COG categories encompassing housekeeping genes and also genes encoding various carbohydrate metabolic and transport systems (Fig. 4).

Compared to their abundance as a functional category in the genome, the diversity regions of the *B. catenulatum* and *B. pseudo-*

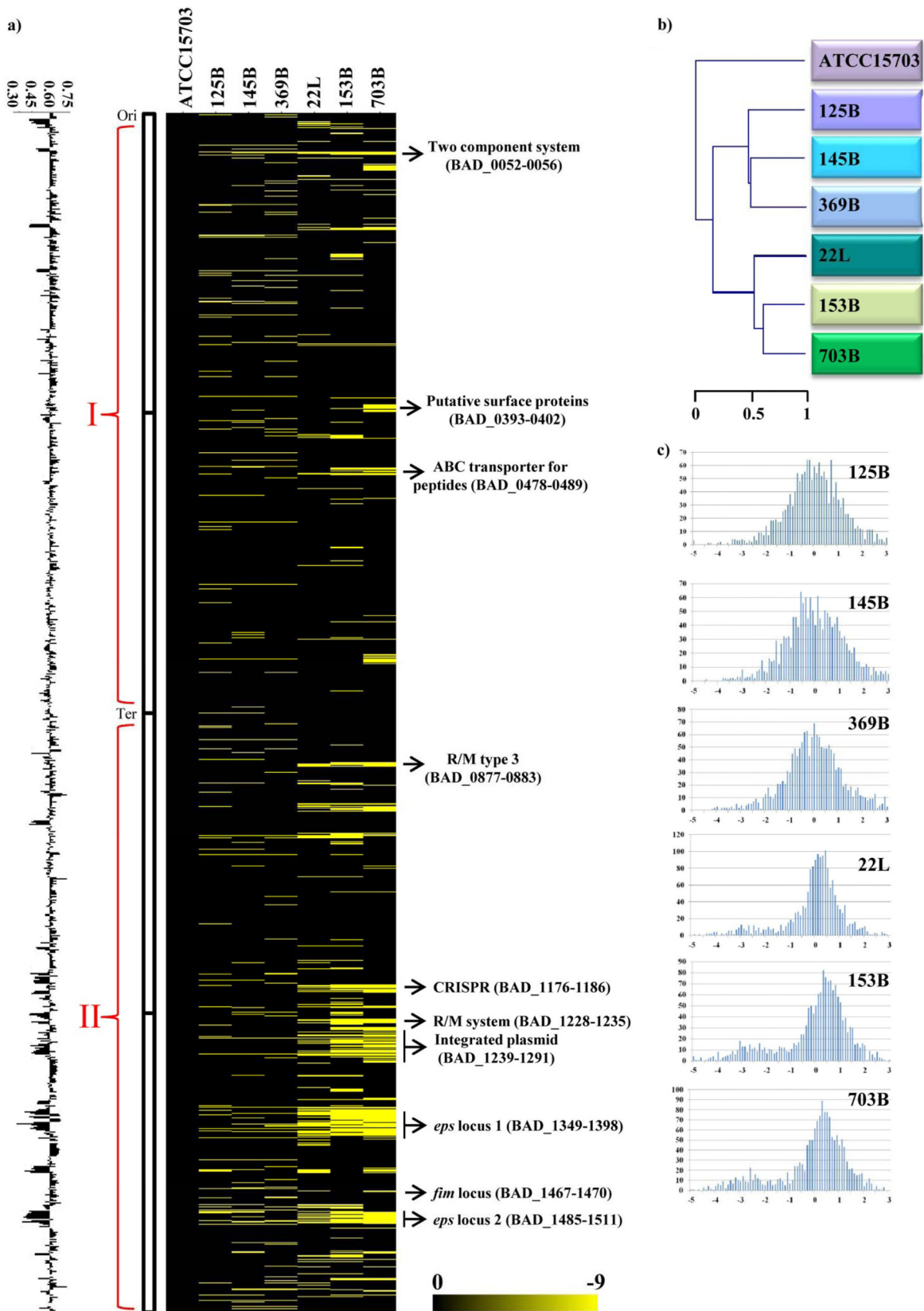
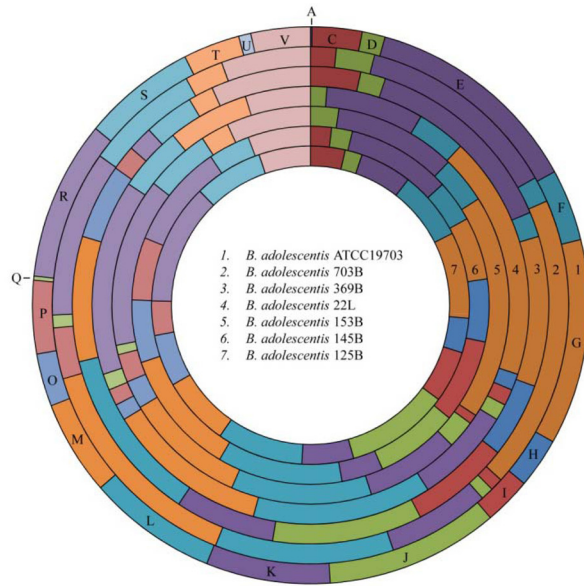
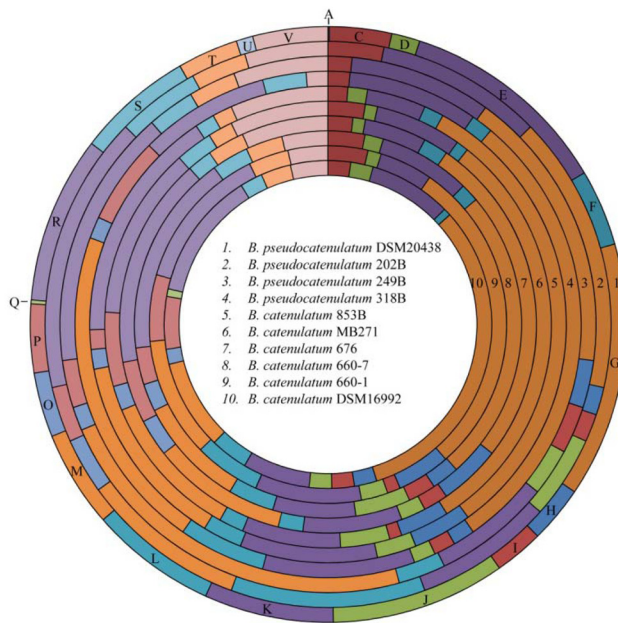


FIG 3 Genomic diversity within representatives of the *B. adolescentis* species as determined by CGH and with reference to the *B. adolescentis* ATCC 15703 genome. (a) CGH data. Each horizontal row corresponds to a probe on the array, and genes are ordered vertically according to their position on the ATCC 15703 genome. Columns represent analyzed strains, which are identified by their code numbers. The color code, which varies from black to yellow to indicate the presence (black), divergence, or absence (yellow) of a gene sequence, is given near the bottom. The predicted functions of some relevant genes are shown in the right-hand margin. Ori, origin of replication; Ter, terminus of replication; R/M, restriction-modification. (b) CGH-based clustering data. A CGH-based clustering analysis was performed for the seven *B. adolescentis* strains analyzed. (c) Signal ratio distribution of the CGH data. The reference is *B. adolescentis* strain ATCC 15703. Ratios are expressed in a \log_2 scale. The x axis of each plot indicates the \log_2 ratio, whereas the y axis indicates observed frequency.

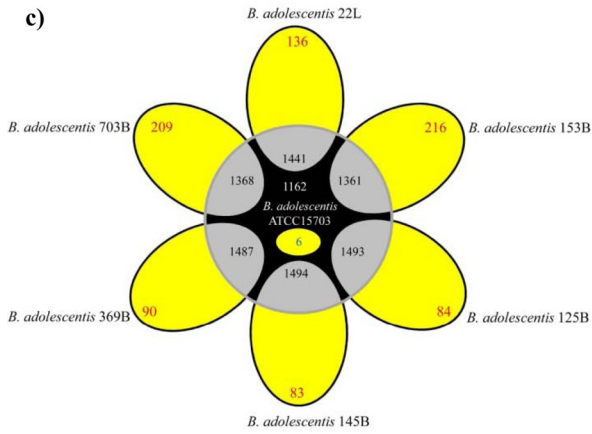
a)



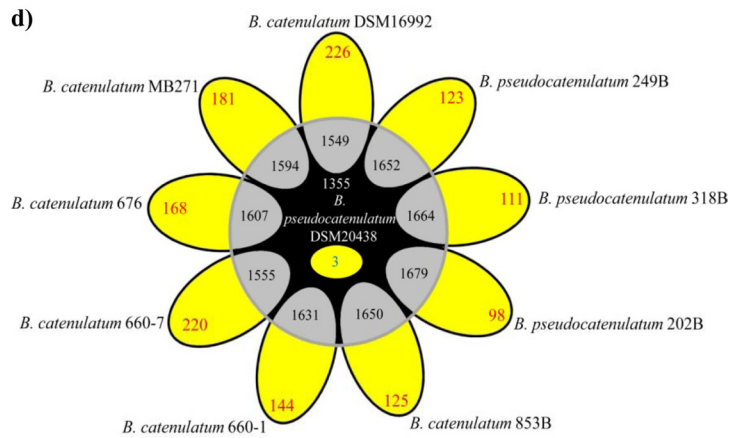
b)



c)



d)



catenulatum strains were somewhat enriched in functions related to cell wall/membrane/envelope biogenesis and carbohydrate metabolism (Fig. 4).

Identification of the core genome of the *B. adolescentis* phylogenetic group. Comparative analyses were performed between the core genes derived by CGH analyses and available complete genome sequences of the *B. adolescentis* group genomes (*B. dentium* Bd1, *B. adolescentis* ATCC 15703, and *B. pseudocatenulatum* DSM20438).

The acquired CGH data for *B. adolescentis* and *B. pseudocatenulatum*/*B. catenulatum* allowed the identification of the shared genetic repertoire of these taxa, represented by 766 genes (Fig. 6a). In addition, the availability of the genome sequences as well as CGH data for other members of the *B. adolescentis* phylogenetic group, such as the chromosome sequence of *B. dentium* Bd1 (GenBank accession number NC_013714) and the CGH data for 10 additional *B. dentium* strains (14), allowed us to further evaluate if this genetic arsenal is also conserved in these bacteria. The performed analyses revealed the presence of 685 gene families, which represent gene sets, each of which belongs to the same functional class based on MCL clustering analysis (31). Representatives of each of these 685 gene families are commonly present in the genomes of *B. adolescentis*, *B. pseudocatenulatum*, *B. catenulatum*, and *B. dentium* and might represent the core genome of the *B. adolescentis* phylogenetic group (Fig. 6a). COG classification of the sequences of this presumed core genome highlighted, as expected, a clear dominance of housekeeping functions such as those involved in DNA replication, transcription, and translation but also the presence of genes encoding carbohydrate breakdown and transport as well as amino acid metabolism and transport (Fig. 6b). Such findings further reinforce the notion that the latter gene categories represent a general genome feature of bifidobacteria (32). Notably, *in silico* analyses of these core genes of the *B. adolescentis* group predicted to be involved in carbohydrate hydrolysis, in accordance with the Carbohydrate Active Enzymes (CAZy) system of Coutinho and Henrissat (<http://www.cazy.org/>), highlighted the existence of 29 carbohydrate-active proteins, including glycoside hydrolases (GHs) and glycosyltransferases (GTs), distributed in 20 GH and 9 GT families (see Table S3 in the supplemental material). These enzymes are predicted to be involved in the breakdown of plant-derived carbohydrates as well as host glycan. Furthermore, among the core genes of the *B. adolescentis* group and involved in carbohydrate metabolism were three genes encoding carbohydrate transporters (see Table S3), which according to the Transporter Classification Database (TCDB;

www.tcdb.org) are active in the uptake of dietary carbohydrates such as raffinose/stachyose, maltose/trehalose, and arabinosaccharide.

Conclusions. The era of genomic exploration is only in its infancy for bifidobacteria, and the limited number of species for which genome sequences have been completed has prompted us to investigate the genetic composition and diversity of this bacterial genus. Whole-genome sequencing of multiple strains of the same species is likely to be the most straightforward strategy in order to achieve this aim. However, even though DNA sequencing technologies are rapidly advancing by providing novel and less expensive sequencing approaches, processing and analyses of the DNA sequence data are still time-consuming and require considerable bioinformatic skills. Thus, a valuable alternative to delineate genomic evolution is to use a polyphasic approach involving classical molecular typing tools such as ITS sequencing and PFGE analyses complemented by comparative genomic approaches. The usefulness of such a polyphasic approach has previously been demonstrated for many microbial groups such as lactobacilli (21, 33, 34), as well as for various bifidobacterial species, such as *Bifidobacterium breve* (9), *Bifidobacterium longum* subsp. *infantis* (35), *Bifidobacterium dentium* (14), and *Bifidobacterium bifidum* (13). The current study represents an example of the application of this polyphasic approach allowing insights into the genetic diversity of the *B. adolescentis* phylogenetic group, which includes key components of the gut microbiota of adults (*B. adolescentis*, *B. pseudocatenulatum*, and *B. catenulatum*) (4). Such analyses have highlighted the existence of extensive genetic variability within the *B. adolescentis* as well as *B. pseudocatenulatum* and *B. catenulatum* species, with a level of genetic diversity that seemed higher than that observed among other members of the *B. adolescentis* phylogenetic group, such as *B. dentium* (14). These analyses allowed the identification of groups of *B. adolescentis* strains that cocluster in the same group and that have also been isolated from the same ecological niche, thus reinforcing the previously proposed concept of a high level of genetic specialization of bifidobacteria to their ecological niche (14, 23). Compared to other enteric bifidobacterial species (9, 13, 14), the high genetic variability detected within the *B. adolescentis* species suggests that this taxon should be further subdivided in other taxonomic units (e.g., subspecies). A number of branching discrepancies were noticed when the 16S-23S ITS-based tree was compared with PFGE- and CGH-based dendrograms, which may have been due to the different robustnesses of the molecular markers/genetic traits used as well as by the fact that only the 16S-ITS sequence-based tree is a real estima-

FIG 4 (a and b) Number and functional analysis according to COG categories of identified *B. adolescentis* (a) or *B. catenulatum*/*B. pseudocatenulatum* (b) genetic functions, as based on variable genes identified by CGH analyses. Each COG family is identified by a one-letter abbreviation: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control and mitosis; E, amino acid metabolism and transport; F, nucleotide metabolism and transport; G, carbohydrate metabolism and transport; H, coenzyme metabolism; I, lipid metabolism; J, translation; K, transcription; L, replication and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, posttranslational modification, protein turnover, chaperone functions; P, inorganic ion transport and metabolism; Q, secondary structure; T, signal transduction; U, intracellular trafficking and secretion; Y, nuclear structure; V, defense mechanisms; Z, cytoskeleton; R, general functional prediction only; S, function unknown. For each category, the black bar represents the percentage of genes in that category as detected in the sequenced genomes of *B. adolescentis* ATCC 15703 and *B. pseudocatenulatum* DSM20438. (c) Venn diagram displaying the number of conserved genes in *B. adolescentis* species (white type), variable (red type), unique genes in *B. adolescentis* ATCC 15703 (blue type), and genes shared between ATCC 15703 and each of the other *B. adolescentis* strains tested (black type) determined by CGH experiments. A detailed description of the variable regions identified in the genomes of *B. adolescentis* species is reported in Table S1 in the supplemental material. (d) Venn diagram displaying the number of conserved genes in *B. catenulatum*/*B. pseudocatenulatum* species (indicated in white), variable genes (depicted in red), unique genes in *B. pseudocatenulatum* DSM20438 (blue type), and genes shared between DSM20438 and each of the other *B. catenulatum*/*B. pseudocatenulatum* strains tested (black type) determined by CGH experiments. A detailed description of the variable regions identified in the genomes of *B. catenulatum*/*B. pseudocatenulatum* species is presented in Table S1 in the supplemental material.

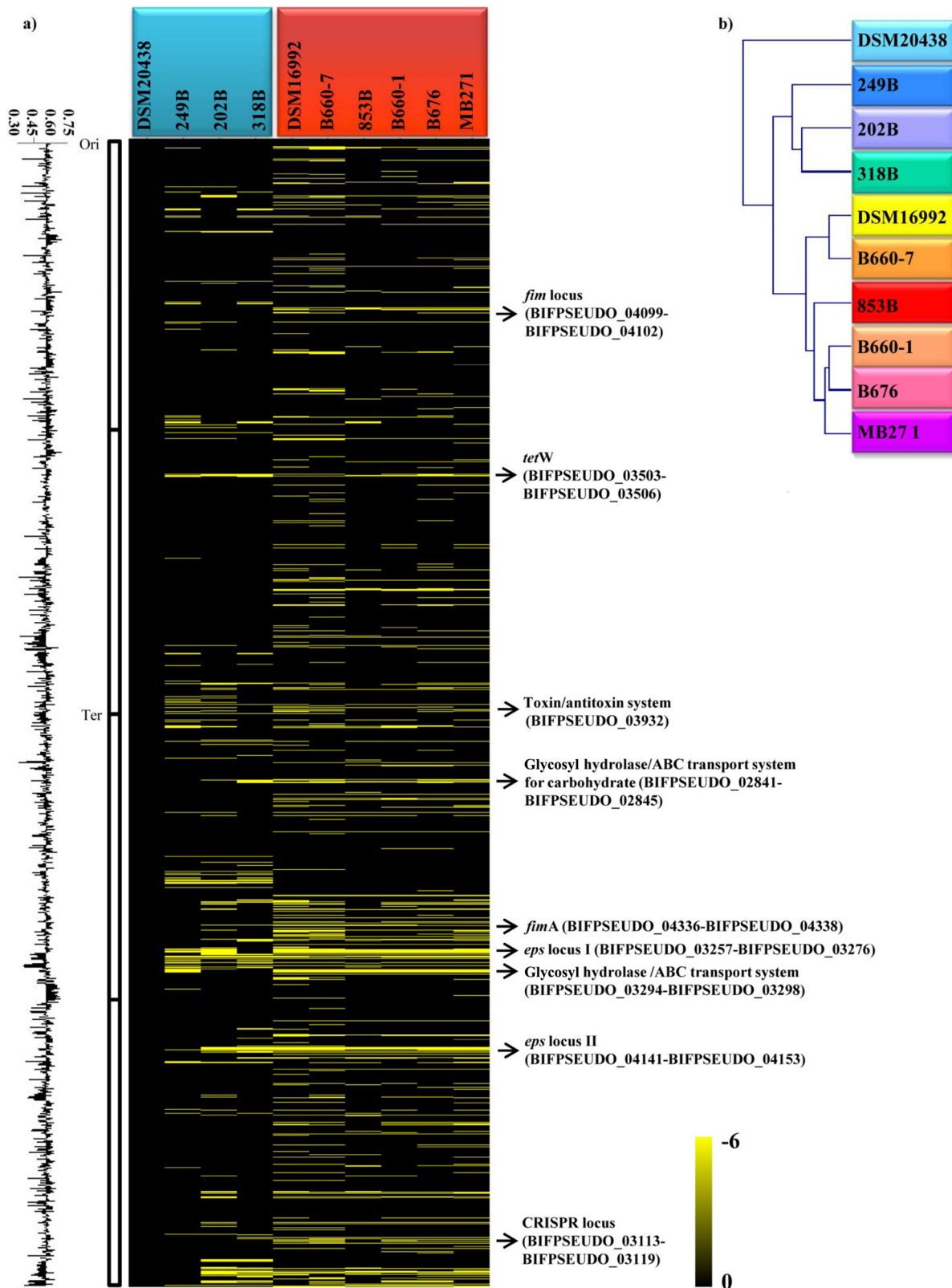


FIG 5 Genomic inter-/intraspecies diversity among *B. pseudocatenulatum* and *B. catenulatum* strains with reference to the *B. pseudocatenulatum* strain DSM20438 genome as identified by CGH assays. (a) Obtained CGH data. Each row corresponds to a probe on the array, and genes are ordered vertically according to their position on the DSM20438 genome. Columns represent analyzed strains, which are identified by their code numbers. The color code, which goes from black to yellow to indicate the presence (black), divergence, or absence (yellow) of a gene sequence, is given near the bottom. The predicted functions of some relevant genes are shown in the right-hand margin. Ori, origin of replication; Ter, terminus of replication. (b) CGH-based clustering data. A CGH-based clustering analysis was performed for the four *B. pseudocatenulatum* and six *B. catenulatum* strains analyzed.

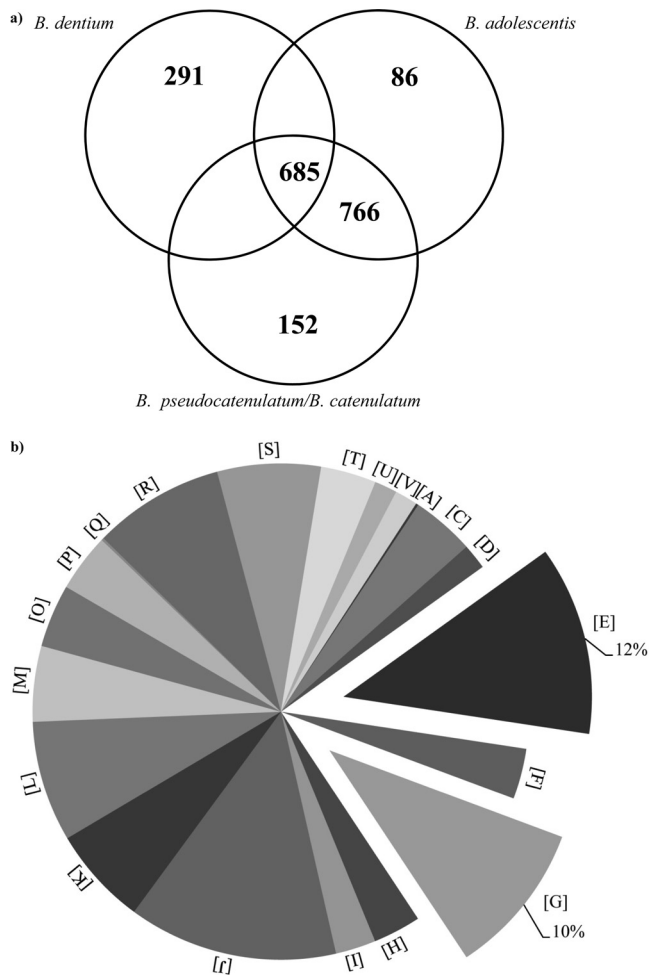


FIG 6 Identification of the core genome sequences of the *B. adolescentis* phylogenetic group. (a) Venn diagram of homologs shared between *B. adolescentis*, *B. pseudocatenulatum*, *B. catenulatum*, and *B. dentium* genomes identified based on CGH data. (b) Pie chart displaying the proportions of proteins encoded by the core genome sequences of the *B. adolescentis* phylogenetic group assayed by CGH according to general functional categories. Each COG family is identified by a one-letter abbreviation as described in the legend to Fig. 4.

tion of the phylogenetic relationships among the investigated strains, in contrast to PFGE- and CGH-based dendrograms, which are merely an evaluation of the phenetic relationships of the analyzed strains based on UPGMA (unweighted-pair group method using average linkages) analyses of banding patterns. Comparison of the results achieved by the various molecular typing tools used in this study showed that the CGH technique had the highest discriminatory power. In fact, in contrast to the other techniques applied here, which all are based on a single molecular marker (e.g., 16S rRNA and 16S-23S ITS) or on a restriction profile (PFGE), CGH data provide an overview on the complete gene repertoire of a specific strain. Furthermore, grouping of strains assayed by CGH may provide information regarding their genetic adaptation to a specific ecological niche. This notion is corroborated by our finding that strains which were isolated from the same ecological niche were shown to branch together on the CGH-derived UPGMA-based tree. Nevertheless, microarray-based techniques are incapable of identifying regions present in

the strains of interest but absent from the strain from which the array was constructed. Consequently, although CGH-based analyses may represent a solid basis for the assessment of genetic variability within the *B. adolescentis* group, these analyses are not expected to provide a complete picture of genomic diversity.

Furthermore, complementation of the phenetic clustering as achieved by PFGE with genome-based approaches like CGH allowed us to assess the genomic constraints related to core genome sequences of the *B. adolescentis* group, truly unique genes (TUG), i.e., genes present only in a reference genome and absent in any of the other bifidobacterial genomes, and variable genes, i.e., genes present that do not occur in all bifidobacterial genomes. Interestingly, among the core genes of the *B. adolescentis* group we found those that specify proteins that are involved in the breakdown and uptake of complex carbohydrates, derived either from the diet (e.g., plant material) or from host-produced glycans. Such findings reinforce the notion that members of the *B. adolescentis* group are common inhabitants of the adult human colon, where they can feed either on dietary components that are rich in complex plant-derived carbohydrates that are insensitive to digestion by mammalian enzymes (36) or on carbohydrates that are derived from host-derived N- or O-linked glycoproteins.

Interestingly, many of the encoded products of the here-identified TUG genes as well as variable genes, i.e., putative adhesion-mediating proteins similar to pilus subunits and sugar-metabolizing enzymes/transporters, are predicted to modulate/support the interaction with the human host, thus facilitating adaptation of bifidobacteria to the human gut.

ACKNOWLEDGMENTS

We thank GenProbio srl for the financial support of the Laboratory of Probiogenomics. This work was financially supported by Fondazione Cariparma to M.V. and by a FEMS Advanced Fellowship 2011 and an IRCSET Embark postdoctoral fellowship to F.T. D.V.S. is a member of The Alimentary Pharmabiotic Centre, which is a Centre for Science and Technology (CSET) funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan (grant no. 02/CE/B124 and 07/CE/B1368). This work was also financially supported by a Ph.D. fellowship (Spinner 2013, Regione Emilia Romagna) to S.D.

We thank also Sacco Srl for providing us with the CHEF DR II apparatus. Finally, we are gratefully to Diego Matteuzzi for providing us with bifidobacterial strains.

REFERENCES

- Cronin M, Ventura M, Fitzgerald GF, van Sinderen D. 2011. Progress in genomics, metabolism and biotechnology of bifidobacteria. *Int. J. Food Microbiol.* 149:4–18.
- Lee JH, O'Sullivan DJ. 2010. Genomic insights into bifidobacteria. *Microbiol. Mol. Biol. Rev.* 74:378–416.
- Ventura M, O'Flaherty S, Claesson MJ, Turrioni F, Klaenhammer TR, van Sinderen D, O'Toole PW. 2009. Genome-scale analyses of health-promoting bacteria: probiogenomics. *Nat. Rev. Microbiol.* 7:61–71.
- Turrioni 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.
- Barrangou R, Briczinski EP, Traeger LL, Loquasto JR, Richards M, Horvath P, Couste-Monvoisin AC, Leyer G, Rendulic S, Steele JL, Broadbent JR, Oberg T, Dudley EG, Schuster S, Romero DA, Roberts RF. 2009. Comparison of the complete genome sequences of *Bifidobacterium animalis* subsp. *lactis* DSM 10140 and BI-04. *J. Bacteriol.* 191:4144–4151.
- Bottacini F, Dal Bello F, Turrioni F, Milani C, Duranti S, Foroni E, Viappiani A, Strati F, Mora D, van Sinderen D, Ventura M. 2011. Complete genome sequence of *Bifidobacterium animalis* subsp. *lactis* BLC1. *J. Bacteriol.* 193:6387–6388.

7. Garrigues C, Johansen E, Pedersen MB. 2010. Complete genome sequence of *Bifidobacterium animalis* subsp. *lactis* BB-12, a widely consumed probiotic strain. *J. Bacteriol.* 192:2467–2468.
8. Lee JH, Karamychev VN, Kozyavkin SA, Mills D, Pavlov AR, Pavlova NV, Polouchine NN, Richardson PM, Shakhova VV, Slesarev AI, Weimer B, O'Sullivan DJ. 2008. Comparative genomic analysis of the gut bacterium *Bifidobacterium longum* reveals loci susceptible to deletion during pure culture growth. *BMC Genomics* 9:247. doi:10.1186/1471-2164-9-247.
9. O'Connell Motherway M, Zomer A, Leahy SC, Reunanan J, Bottacini F, Claesson MJ, O'Brien F, Flynn K, Casey PG, Munoz JA, Kearney B, Houston AM, O'Mahony C, Higgins DG, Shanahan F, Palva A, de Vos WM, Fitzgerald GF, Ventura M, O'Toole PW, van Sinderen D. 2011. Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc. Natl. Acad. Sci. U. S. A.* 108:11217–11222.
10. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, Mende DR, Li J, Xu J, Li S, Li D, Cao J, Wang B, Liang H, Zheng H, Xie Y, Tap J, Lepage P, Bertalan M, Batto JM, Hansen T, Le Paslier D, Linneberg A, Nielsen HB, Pelletier E, Renault P, Sicheritz-Ponten T, Turner K, Zhu H, Yu C, Jian M, Zhou Y, Li Y, Zhang X, Qin N, Yang H, Wang J, Brunak S, Dore J, Guarner F, Kristiansen K, Pedersen O, Parkhill J, Weissenbach J, Bork P, Ehrlich SD. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65.
11. Schell MA, Karmirantzou M, Snel B, Vilanova D, Berger B, Pessi G, Zwahlen MC, Desiere F, Bork P, Delley M, Pridmore RD, Arigoni F. 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.
12. Sela DA, Chapman J, Adeuya A, Kim JH, Chen F, Whitehead TR, Lapidus A, Rokhsar DS, Lebrilla CB, German JB, Price NP, Richardson PM, Mills DA. 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.
13. Turrioni F, Bottacini F, Foroni E, Mulder I, Kim JH, Zomer A, Sanchez B, Bidossi A, Ferrarini A, Giubellini V, Delle Donne M, Henrissat B, Coutinho P, Oggioni M, Fitzgerald GF, Mills D, Margolles A, Kelly D, van Sinderen D, Ventura M. 2010. Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging. *Proc. Natl. Acad. Sci. U. S. A.* 107:19514–19519.
14. Ventura M, Turrioni F, Zomer A, Foroni E, Giubellini V, Bottacini F, Canchaya C, Claesson MJ, He F, Mantzourani M, Mulas L, Ferrarini A, Gao B, Delle Donne M, Henrissat B, Coutinho P, Oggioni M, Gupta RS, Zhang Z, Beighton D, Fitzgerald GF, O'Toole PW, van Sinderen D. 2009. The *Bifidobacterium dentium* Bd1 genome sequence reflects its genetic adaptation to the human oral cavity. *PLoS Genet.* 5:e1000785. doi:10.1371/journal.pgen.1000785.
15. Turrioni F, Foroni E, Pizzetti P, Giubellini V, Ribbera A, Merusi P, Cagnasso P, Bizzarri B, de'Angelis GL, Shanahan F, van Sinderen D, Ventura M. 2009. Exploring the diversity of the bifidobacterial population in the human intestinal tract. *Appl. Environ. Microbiol.* 75:1534–1545.
16. Turrioni F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, Kerr C, Hourihane J, Murray D, Fuligni F, Gueimonde M, Margolles A, De Bellis G, O'Toole PW, van Sinderen D, Marchesi JR, Ventura M. 2012. Diversity of bifidobacteria within the infant gut microbiota. *PLoS One* 7:e36957. doi:10.1371/journal.pone.0036957.
17. Yasui K, Tabata M, Yamada S, Abe T, Ikemura T, Osawa R, Suzuki T. 2009. Intra-species diversity between seven *Bifidobacterium adolescentis* strains identified by genome-wide tiling array analysis. *Biosci. Biotechnol. Biochem.* 73:1422–1424.
18. Ventura M, Elli M, Reniero R, Zink R. 2001. Molecular microbial analysis of *Bifidobacterium* isolates from different environments by the species-specific amplified ribosomal DNA restriction analysis (ARDRA). *FEMS Microbiol. Ecol.* 36:113–121.
19. Felsenstein J. 1989. Mathematics vs. evolution: mathematical evolutionary theory. *Science* 246:941–942.
20. Leblond-Bourget N, Philippe H, Mangin I, Decaris B. 1996. 16S rRNA and 16S to 23S internal transcribed spacer sequence analyses reveal inter- and intraspecific *Bifidobacterium* phylogeny. *Int. J. Syst. Bacteriol.* 46:102–111.
21. Berger B, Pridmore RD, Barretto C, Delmas-Julien F, Schreiber K, Arigoni F, Brussow H. 2007. Similarity and differences in the *Lactobacillus acidophilus* group identified by polyphasic analysis and comparative genomics. *J. Bacteriol.* 189:1311–1321.
22. Ventura M, Zink R. 2002. Rapid identification, differentiation, and proposed new taxonomic classification of *Bifidobacterium lactis*. *Appl. Environ. Microbiol.* 68:6429–6434.
23. Rouillard JM, Zuker M, Gulari E. 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res.* 31:3057–3062.
24. Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95:14863–14868.
25. Klijn A, Mercenier A, Arigoni F. 2005. Lessons from the genomes of bifidobacteria. *FEMS Microbiol. Rev.* 29:491–509.
26. Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, Carmel G, Tummino PJ, Caruso A, Uria-Nickelsen M, Mills DM, Ives C, Gibson R, Merberg D, Mills SD, Jiang Q, Taylor DE, Vovis GF, Trust TJ. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397:176–180.
27. Romo-Gonzalez C, Salama NR, Burgeno-Ferreira J, Ponce-Castaneda V, Lazcano-Ponce E, Camorlinga-Ponce M, Torres J. 2009. Differences in genome content among *Helicobacter pylori* isolates from patients with gastritis, duodenal ulcer, or gastric cancer reveal novel disease-associated genes. *Infect. Immun.* 77:2201–2211.
28. Scardovi CF. 1974. *Bifidobacterium catenulatum*, *Bifidobacterium dentium* and *Bifidobacterium angulatum*: three new species and their deoxyribonucleic acid homology relationships. *Int. J. Syst. Bacteriol.* 24:6–20.
29. Turrioni F, Strati F, Foroni E, Serafini F, Duranti S, van Sinderen D, Ventura M. 2012. Analysis of predicted carbohydrate transport systems encoded by *Bifidobacterium bifidum* PRL2010. *Appl. Environ. Microbiol.* 78:5002–5012.
30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
31. Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30:1575–1584.
32. Bottacini F, Medini D, Pavesi A, Turrioni F, Foroni E, Riley D, Giubellini V, Tettelin H, van Sinderen D, Ventura M. 2010. Comparative genomics of the genus *Bifidobacterium*. *Microbiology* 156:3243–3254.
33. Molenaar D, Bringel F, Schuren FH, de Vos WM, Siezen RJ, Kleerebezem M. 2005. Exploring *Lactobacillus plantarum* genome diversity by using microarrays. *J. Bacteriol.* 187:6119–6127.
34. Raftis EJ, Salvetti E, Torriani S, Felis GE, O'Toole PW. 2011. Genomic diversity of *Lactobacillus salivarius*. *Appl. Environ. Microbiol.* 77:954–965.
35. LoCascio RG, Desai P, Sela DA, Weimer B, Mills DA. 2010. Broad conservation of milk utilization genes in *Bifidobacterium longum* subsp. *infantis* as revealed by comparative genomic hybridization. *Appl. Environ. Microbiol.* 76:7373–7381.
36. Ventura M, Canchaya C, Fitzgerald GF, Gupta RS, van Sinderen D. 2007. Genomics as a means to understand bacterial phylogeny and ecological adaptation: the case of bifidobacteria. *Antonie Van Leeuwenhoek* 91:351–372.
37. Ventura M, Zink R. 2002. Specific identification and molecular typing analysis of *Lactobacillus johnsonii* by using PCR-based methods and pulsed-field gel electrophoresis. *FEMS Microbiol. Lett.* 217:141–154.