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 Delledonne,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;* Alimentary Pharmacology Centre and Department of Microbiology, Bioscience Institute, National University of Ireland, Cork, Ireland;# Sacco Srl, Codagoro, Italy;^ Functional Genomic Center, Department of Biotechnology, University of Verona, Verona, Italy

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′-CTACGGCTACCTTGTTACGA-3′) (18), while DNA fragments corresponding to the 16S-ITS region were amplified using the oligonucleotides BIF-specific (5′-GGTGTGAAAGTCATCGCT-3′) and 23S_bif (5′-GT CTGCCAAGGCATCCACA-3′) (15). Phylogenetic analysis and trees were calculated using the PHYLIP software package, version 3.5c (19).

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@univr.it. S.D. and F.T. contributed equally.

Copyright © 2013, American Society for Microbiology. All Rights Reserved. doi:10.1128/AEM.02467-12
using the ITS sequence-based tree is consistent with that obtained
strains (15, 21, 22). Indeed, the phylogenetic clustering achieved
16S-23S ITS amplicons of various bifidobacterial or lactobacillus
sequence had previously been shown not to affect the overall phy-
variability detected between the different ribosomal loci (16S and
PCRs directed to amplify 16S-23S ITS sequences of all investigated
the same species, as well as between closely related taxa. Thus,
molecular marker to infer phylogenetic distances between strains
Bioinformatic analyses. Comparative analyses were performed using
BLAST analysis (30) and MCL (graph-theory-based Markov clusters al-
Microarray data accession numbers. The microarray data have been
deposited in the Gene Expression Omnibus database (http://www.ncbi
RESULTS AND DISCUSSION
Phylogenetic analysis based on rRNA sequences. A phylogenetic
analysis based on the 16S rRNA gene sequences of the 17 strains
alyzed 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 oligo-
nucleotides BIF-specific and 235_bif (15). The resulting 16S-23S
ITS amplicons were subjected to DNA sequencing. The micro-
variability 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 phy-
logenetic 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.

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

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 micro-
array 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 se-
quences of B. adolescentis ATCC 15703 and B. pseudocatenula-
tum 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 de-
scribed 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 sug-

gest that the B. adolescentis genome is subject to relatively rapid
diversification, as also observed in certain other enteric bifidobac-
teria (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 distrib-
uted 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. den-
tium (14) and Bifidobacterium bifidum species (13).

According to the ATCC 15703 gene function annotation, the
identifed 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 under-
lie 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 vari-
able regions identified in the genomes of B. adolescentis species is
reported in Table S1 in the supplemental material. Plasticity re-

gions 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 in-
teractions and metabolic abilities. These include a number of re-

gions predicted to be involved in the adaptation of ATCC 15703 to
its ecological niche, such as two exopolysaccharide (eps) loci (eps,
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 SpeI. 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 colocate the CRISPR loci (BIFP SEUDO_03113-BIFPSEUDO_03119) and terW locus (BIFP SEUDO_03503-BIFPSEUDO_03506). DNA regions associated with bacterium-environment interaction include two sortase-dependent pilus loci (BIFPSEUDO_04099-BIFPSEUDO_04102 and BIFPSEUDO_04336-BIFPSEUDO_04438), 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 mobiliome 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. pseudocat-
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.
Genomic Diversity of *Bifidobacterium adolescentis*

The genomes of functional class based on MCL clustering analysis (31). Represents, which represent gene sets, each of which belongs to the same category as detected in the sequenced genomes of *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 been previously 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). These 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 robustness 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-

![Image](http://aem.asm.org/)
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.
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. (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.

REFERENCES

Duranti et al.


