The ISME Journal (2009) 3, 745–751

www.nature.com/ismej

SHORT COMMUNICATION

Microbiomic analysis of the bifidobacterial population in the human distal gut

Francesca Turroni[1], Julian R Marchesi[2], Elena Foroni[1], Miguel Gueimonde[3], Fergus Shanahan[4], Abelardo Margolles[3], Douwe van Sinderen[4] and Marco Ventura[1]
[1]*Department of Genetics, Biology of Microorganisms, Anthropology and Evolution, University of Parma, Parma, Italy;* [2]*Cardiff School of Biosciences, Cardiff University, Cardiff, Wales, UK;* [3]*Departamento de Microbiología y Bioquímica de Productos Lácteos, IPLA–CSIC, Ctra. Infiesto s/n, Villaviciosa, Asturias, Spain and* [4]*Department of Microbiology, Alimentary Pharmabiotic Centre, Bioscience Institute, National University of Ireland, Cork, Ireland*

One of the most complex microbial ecosystems is represented by the microbiota of the human gastrointestinal tract (GIT). Although this microbial consortium has been recognized to have a crucial effect on human health, its precise composition is still not fully established. Among the GIT bacteria, bifidobacteria represent an important commensal group whose presence is often associated with health-promoting effects. In this work, we assessed the complexity of the human intestinal bifidobacterial population by analysing the diversity of several 16S rRNA gene-based libraries. These analyses showed the presence of novel bifidobacterial phylotypes, which had not been found earlier and may thus represent novel taxa within the genus *Bifidobacterium*.
The ISME Journal (2009) **3,** 745–751; doi:10.1038/ismej.2009.19; published online 19 March 2009
Subject Category: microbial population and community ecology
Keywords: *Bifidobacterium*; human gut microbiota; microbiomics; gut biodiversity

The human gastrointestinal tract (GIT) is an extremely complex microbial ecosystem (Gill *et al.*, 2006), whose microbial component (also termed the microbiota) effects on human health (Kurokawa *et al.*, 2007). Within the human gut some microbial members are permanently established (mucosa-adherent components), whereas others may represent transient members (Ley and Peterson DA Gordon, 2006; Turroni *et al.*, 2008, 2009). It is noted that the mucosa-adherent community seems to be different from the luminal community, which in turn resembles the fecal microbiota (Eckburg *et al.*, 2005; Turroni *et al.*, 2009). However, several studies employing 16S rRNA-based fluorescent probes have suggested that commensal bacteria may live in suspension in the lumen without being in direct contact with the gut epithelium (Van der Waaij *et al.*, 2005; Swidsinski *et al.*, 2007).

Despite its importance, the precise composition and activities of the human GIT microbiota are still subject to a lot of speculation because of its complexity. For this reason, an accurate analysis of this microbial consortium is an essential step to understand how the various elements of the GIT microbiota interact and consequently affect the health status of the host (Comstock, 2007).

Among the GIT-resident (mucosa-adherent) bacteria, bifidobacteria represent a commensal group that constitutes less than 10% of the human adult microbiota, but whose presence is often associated with health-promoting effects (Ouwehand *et al.*, 2002; Turroni *et al.*, 2008). Bifidobacteria are high G+C Gram-positive bacteria belonging to the *Bifidobacteriaceae* family and the *Actinobacteria* class (Ventura *et al.*, 2007). Traditionally, estimates of microbial diversity were based solely on culturable microorganisms. However, microbial observations and mathematical models indicate that the majority of bacteria are non-culturable under standard laboratory conditions (Stach and Bull, 2005). Advances in microbial ecology including metagenomic studies of environmental samples have allowed microbial ecologists to access earlier unimaginable genetic diversity. Recently, metagenomic and microbiomic studies based on the analysis of genomic DNA- and rRNA gene libraries have highlighted the diversity of the GIT microbiota and have shown that it consists of many novel and, as yet, unculturable bacterial components (Eckburg *et al.*, 2005; Wang *et al.*, 2005; Gill *et al.*, 2006; Kurokawa *et al.*, 2007; Palmer *et al.*, 2007). However, relatively few bifidobacterial rRNA gene sequences have so far been identified in these metagenomic studies, which are

Correspondence: M Ventura, Department of Genetics, Biology of Microorganisms, Anthropology and Evolution, University of Parma, Parco Area delle Scienze 11A, Parma 43100, Italy.
E-mail: marco.ventura@unipr.it
Received 11 December 2008; revised 9 February 2009; accepted 9 February 2009; published online 19 March 2009

in striking contrast to what was described earlier in several bifidobacterial surveys, which had adopted culture-dependent approaches (Poxton *et al.*, 1997; Harmsen *et al.*, 2000; Fanaro *et al.*, 2003).

Here, we assessed the molecular complexity and biodiversity of the human, mucosa-adherent bifidobacterial population by analysing the diversity of bifidobacteria-specific 16S rRNA-encoding sequences.

Construction of 16S rRNA gene libraries

Five colonic mucosa samples from healthy adult volunteers of different ages, who all provided their written consent, were recovered for this study by means of colonoscopy as described earlier (Table 1). The fresh samples were transported to the laboratory in sterile containers and under anaerobic conditions and were processed within 1 h on receipt.

Biopsies were centrifuged at 3 000 *g* and resuspended in 600 µl of water. Three glass beads (5 mm in diameter) were added to every tube containing a sample, and the tube was shaken for 1 min and placed in ice, to remove microorganisms from colonic tissue. Total DNA from biopsies was isolated and purified using the Qiagen DNA mini kit according to the manufacturer's instructions (Qiagen, Valencia, CA, USA).

Amplification of the 16S rRNA gene was carried out using a nested PCR approach involving a PCR amplification of the 16S rRNA-encoding gene using the eubacterial PCR primers P0 (5′-GAAGAGTTTG ATCCTGGCTCAG-3′) and P6 (5′-CTACGGCTACCTT GTTACGA-3′) (the targeted amplicon corresponds to bases 4 through to 1494 of the 16S rRNA gene of *Bifidobacterium longum* subsp. *longum* DJO10A) followed by a second PCR amplification with primers specific for 16S rRNA genes found in all members of the genus *Bifidobacterium*: LM3 (5′-CG GGTGCTIcCCCACTTTCATG-3′) and LM26 (5′-GAT TCTGGCTCAGGATGAACG-3′) (the targeted amplicon corresponds to bases 12 to 1400 of the 16S rRNA gene of *B. longum* subsp. *longum* DJO10A) according to the PCR protocols described earlier (Kaufmann *et al.*, 1997; Ventura & Zink, 2002).

The copy number of 16S rRNA genes is known to vary between different *Bifidobacterium* species (for example, four in *B. longum* subsp. *longum*, five in *B. adolescentis* and two in *B. breve*) (Ventura *et al.*, 2007). Micro-heterogeneity, which is restricted to a small number of nucleotide differences, has been noticed between the different 16S rRNA gene copies found in the same microorganism (Bourget *et al.*, 1993). However, such mismatches are not expected to affect the reliability of the 16S rRNA gene as a marker for molecular identification and phylogenetic analyses of bifidobacterial populations (Stackebrandt *et al.*, 1997).

The purified PCR products were ligated into the pGEM-T easy vector system (Promega, Madison, WI, USA) and transformed into electrocompetent *Escherichia coli* M15 cells according to the manufacturer's instructions (Promega). Transformants were selected by blue–white screening methods on Luria–Bertani agar supplemented with ampicillin (100 µg ml^{-1}, Sigma, St Louis, MO, USA) and X-gal (100 µg ml^{-1}) and checked for plasmids with appropriately sized inserts by PCR amplification using the LM3 and LM26 primers.

About 1500 clones containing a putative 16S rRNA gene fragment were randomly selected and submitted to insert sequencing. Nucleotide sequencing of both strands from each selected clone was performed by Agencourt Bioscience Corporation (Boston, MA, USA) using primers, LM3 and LM26. Sequence data assembly was carried out using DNASTAR software (version 5.05 DNAstar, Madison, WI, USA).

Diversity measures

To calculate the diversity measures, the 16S rRNA gene clones were fully sequenced and clones with $\geqslant$ 97% sequences similarity were grouped with the same phylotype, as defined earlier (Suau *et al.*, 1999). A total of 188 clones were either found to contain chimeric sequences according to the software of the Ribosomal Database Project II website (Cole *et al.*, 2003) or they were too short for proper assembly and analysis, and were therefore excluded from the subsequent analysis. The remaining 1312 clone sequences were uniformly distributed within the mucosa biopsies retrieved from the five different individuals (Table 1). Rarefaction curves were

Table 1 Diversity indices for 16S rDNA libraries obtained from different individuals

Mucosal sample	*Details of volunteers*	*Dmean*[a]	*Chao-1*	*Shannon–Wiener index*	*Number of clones*	*Number of bifidobacterial phylotypes*
32F	Female, 32 years old	0.115457	774.4500	3.9974	294	136
FS29	Female, 29 years old	0.098082	636.0500	4.2382	233	126
M	Female, 40 years old	0.087123	231.8000	3.9871	125	75
MS51	Male, 51 years old	0.100568	2902.0455	5.2008	351	262
R	Female, 60 years old	0.088305	433.0000	4.6737	171	127

[a](Watve and Gangal., 1996).

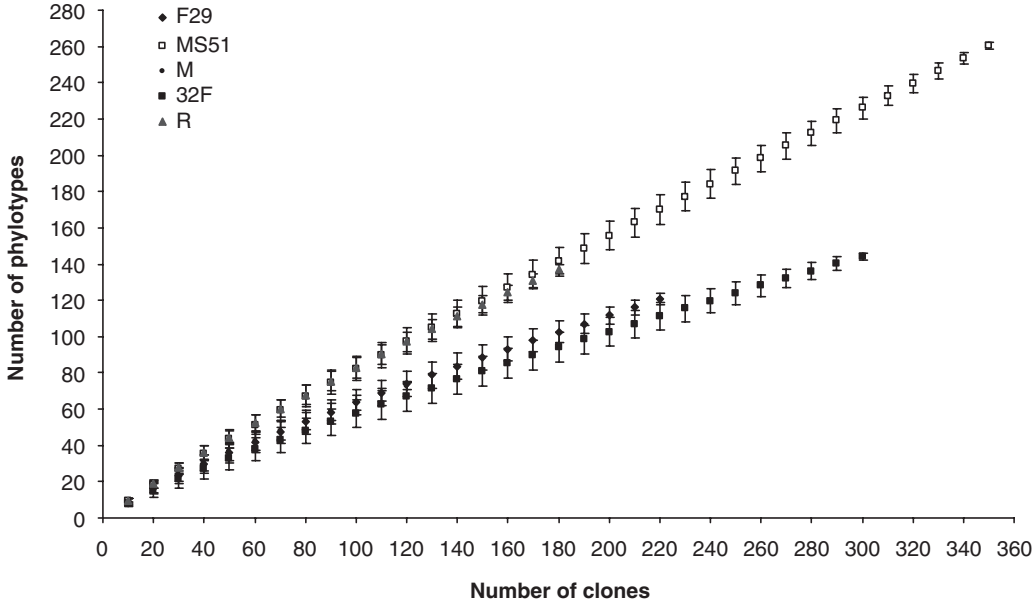

Figure 1 Rarefaction curves generated for 16S rRNA gene sequences obtained from different clone libraries from five individual mucosa samples (see Table 1). Clones were grouped into phylotypes at a level of sequence similarity $\geqslant$ 97%. Error bars indicate 95% confidence intervals.

obtained by plotting the number of different phylotypes identified against the number of clones sequenced (Figure 1). The decrease in the rate of phylotype detection depicted in many of the curves showed that a large part of the diversity in these libraries had been detected. The large number of different phylotypes obtained was analyzed by calculating the Shannon and Chao-1 indexes, both of which suggested a high level of diversity among the mucosa samples/libraries processed (Table 1).

Analysis of the inter-participant variability

Earlier investigations have not rigorously addressed possible differences in the intestinal bifidobacteria between individuals (Eckburg *et al.*, 2005; Wang *et al.*, 2005). We applied techniques that are based on the relative abundance of sequences within communities and the extent of genetic divergence between sequences. To determine whether bifodobacterial communities were different between individuals, the significance test in UniFrac (Lozupone *et al.*, 2006) was used and the raw *P*-values are reported (P-value $\leqslant 0.05$ was considered to be statistically significant). This method was used to determine whether the cluster distribution of the sequences in the different mucosa samples differs from random expectations. Principal coordinate analysis was undertaken using the UniFrac program, axis 1 explained 34% of the variability and principal coordinate analysis axis 2 explained 28%, thus suggesting that this analysis explains a large part of the detected variability. In the analysis, the bifidobacterial communities from each of the five mucosa sample were found to be in different quadrants of the plot, which supports the notion that most of the observed variability in the bifidobacterial population was explained by inter-participant differences. In addition, the raw *P*-values generated using the Unifrac significance test were all less than 0.001 and further re-enforce the view that these are significantly different communities.

To determine whether differences between samples were because of underlying variability in the bifidobacterial communities as opposed to artefacts of sub-sampling, colonic clone libraries were compared with the LIBSHUFF program (Singleton *et al.*, 2001). Pairwise comparisons of each clone library to every other library showed that the majority of the libraries were distinct from each other. As predicted by the principal coordinate analysis results, no pattern of library relatedness between participant groups was evident in this analysis.

So all together, these findings suggest that each individual harbors a specific population of colonic bifidobacteria (Figure 2). The large variability of the bifidobacterial population associated with the intestinal mucosa of each participant is in agreement with the large inter-participant variability of the overall intestinal microbiota that was identified earlier from whole gut microbiota biodiversity investigations (Eckburg *et al.*, 2005; Palmer *et al.*, 2007).

Phylogenetic affiliation of clones

Seqmatch analysis of the 16S rRNA gene sequences on the basis of RDP database identified 1312 clones,

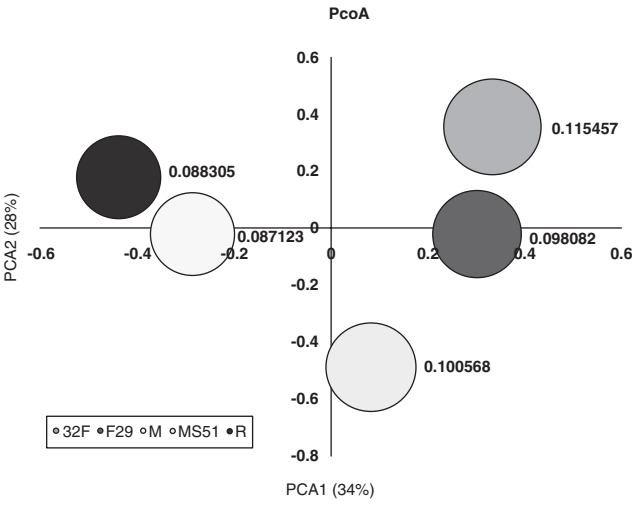

Figure 2 Principal coordinate analyses (PCA) based on the phylotypes identified from different subjects. Percentages shown along the axes represent the proportion of dissimilarities captured by the axes. Each circle represents the clone sequences from each clone library, which have different colors according to the origin. The PCA was performed using the UniFrac associated software and can be found at http://bmf2.colorado.edu/unifrac/index.psp. A full colour version of this figure is available at *The ISME Journal* online.

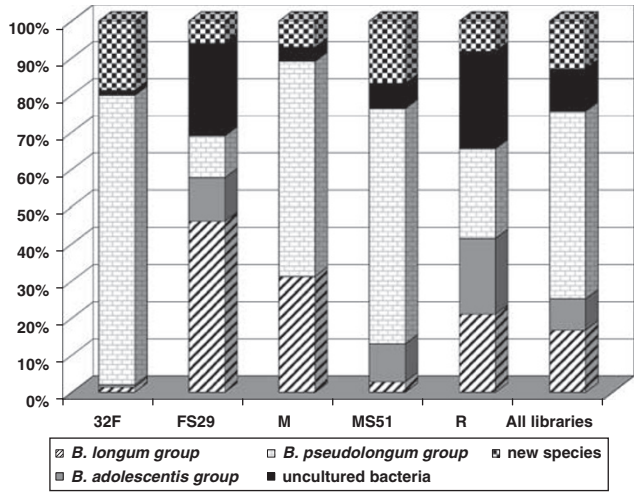

Figure 3 Relative phylotype frequencies of bifidobacterial phylogenetic groups and uncultured bifidobacteria in each clone library or in all the clone sequences analysed.

which contained the expected insert, and which were grouped into phylotypes. Based on earlier studies, which compared partial 16S rRNA gene sequences with new clones (Woese, 1987; Gill *et al.*, 2006; Ley *et al.*, 2008), we considered the following assignments: when the obtained S_{ab} score of a cloned sequence, relative to that of a strain of known species was $\geqslant 0.85$ (which is equivalent of more than 97% sequence identity) in relation to a strain of a known species, the cloned sequences was assigned to that phylogenetic group. When the S_{ab} score of a cloned sequences was less than 0.85 in relation to any known sequences, that clone was labeled an unidentified phylotype (Holben *et al.*, 2004).

A neighbour-joining tree containing the obtained 1312 bifidobacteria-specific sequences and other bacterial phyla as well as 16S rRNA gene sequences from all bifidobacteria neotypes and closely related bifidobacterial taxa (for example, *Scardovia* ssp., *Parascardovia* ssp., *Turicella* ssp.) was constructed (Supplementary Figure 1).

Based on the BLAST results, all sequences were assigned to three phylogenetic phyla of the domain bacteria: *Actinobacteria* (*Bifidobacterium*, *Propionibacterium*), *Bacteroidetes* (*Bacteroides*) and *Firmicutes* (*Lachnospiraceae Incertae Sedis*). As expected, the large majority of the recovered sequences (90%) belongs to members of the genus *Bifidobacterium*. The presence of a large variety of clone sequences corresponding to bifidobacteria in the 16S rRNA gene libraries clearly contrasts with what was described earlier in the majority of published metagenomic investigations exploring the diversity of the human gut microbiota (Eckburg *et al.*, 2005; Wang *et al.*, 2005; Palmer *et al.*, 2007). However, it should be pointed out that the results achieved in these studies may have been skewed by the efficacy of the protocols used for extracting DNA directly from the environmental samples, as well as by the accuracy of PCR primers and the bias of the applied PCR conditions. In fact, the low level of detection of phylotypes belonging to the genus *Bifidobacterium* may be explained by the fact that bifidobacteria, like other Gram-positive bacteria, are less susceptible to lysis and hence it is more difficult to isolate chromosomal DNA from these microorganisms. Another possible explanation may be that in these earlier published metagenomic studies of human gut microbiota, the sampling size was too small to detect bifidobacteria, which are present at lower levels compared with other major components of the microbiota, such as *Bacteroides* and *Clostridium* species. It should also be noted that of the 1312 clone sequences, 136 clones were shown to belong to other bacterial taxa (*Propionibacterium*, *Bacteroides* and *Firmicutes*), which suggests that the LM3 and LM26 *Bifidobacterium* genus-specific PCR primers are not as efficient in exclusively identifying bifidobacteria as described earlier by Kaufmann *et al.*, 1997.

The relative abundances of different bifidobacterial phylogenetic groups presented in each clone library are shown in Figure 3. Such phylogenetic groups were designed in accordance with the earlier described clusters of bifidobacterial taxa, which are based on 16S rRNA gene similarities (Ventura *et al.*, 2006). Phylogenetic analysis of sequences showed that the bifidobacterial composition was dominated by sequences closely related to the *B. pseudolongum* (including *B. pseudolongum* species), *B. longum* (harboring *B. longum* subsp. *longum* and *B. breve* species) and *B. adolescentis* (including *B. adolescentis*, *B. catenulatum* and *B. pseudocatenulatum* species) phylogenetic groups (Figure 3). Fifty

percent of obtained sequences were assigned to the *B. pseudolongum* group, whereas 197 clones fell into the *B. longum* group, which includes *B. breve* (36 clones) and *B. longum* subsp. *longum* (161 clones). It is noted that no *B. longum* subsp. *infantis* or *B. bifidum* sequences were found in these libraries, which could be explained by assuming that these bifidobacterial species are ecologically restricted to particular niches that may be different from those investigated in this study (for example, the infant intestine). One hundred and thirty-six clones (11.5% of the total bifidobacterial clone population) were closely related ($\geqslant$97% sequence identity) to uncultured bifidobacterial clones derived from human colonic material (NCBI source). The remaining 151 clones, which represent 13% of the total number of bifidobacterial sequences recovered from mucosal biopsies had not been identified earlier. In fact, all these cloned bifidobacterial 16S rDNA sequences possess a S_{AB} score less than 0.85 when compared with publicly available bifidobacterial 16S rRNA sequences. These clone sequences fell into 75 different operational taxonomic units, which, based on the criteria described above, represent novel bifidobacterial phylotypes. These novel phylotypes were submitted to GenBank database (accession numbers from FJ518625 to FJ518699). A phylogenetic tree was constructed using the 151 sequences of these earlier unknown phylotypes along with the 16S rRNA gene sequences of their closest relatives (Figure 4). It is noted that a large part of these novel phylotypes are closely related to the *B. pseudolongum* or *B. longum* phylogenetic groups. It is also worth noting that the unknown phylotypes were recovered from all five clone libraries suggesting wide occurrence.

Although it is tempting to declare these new bifidobacterial phylotypes identified in this study as new species, bacterial classification must employ a polyphasic approach as concluded by the *ad hoc*

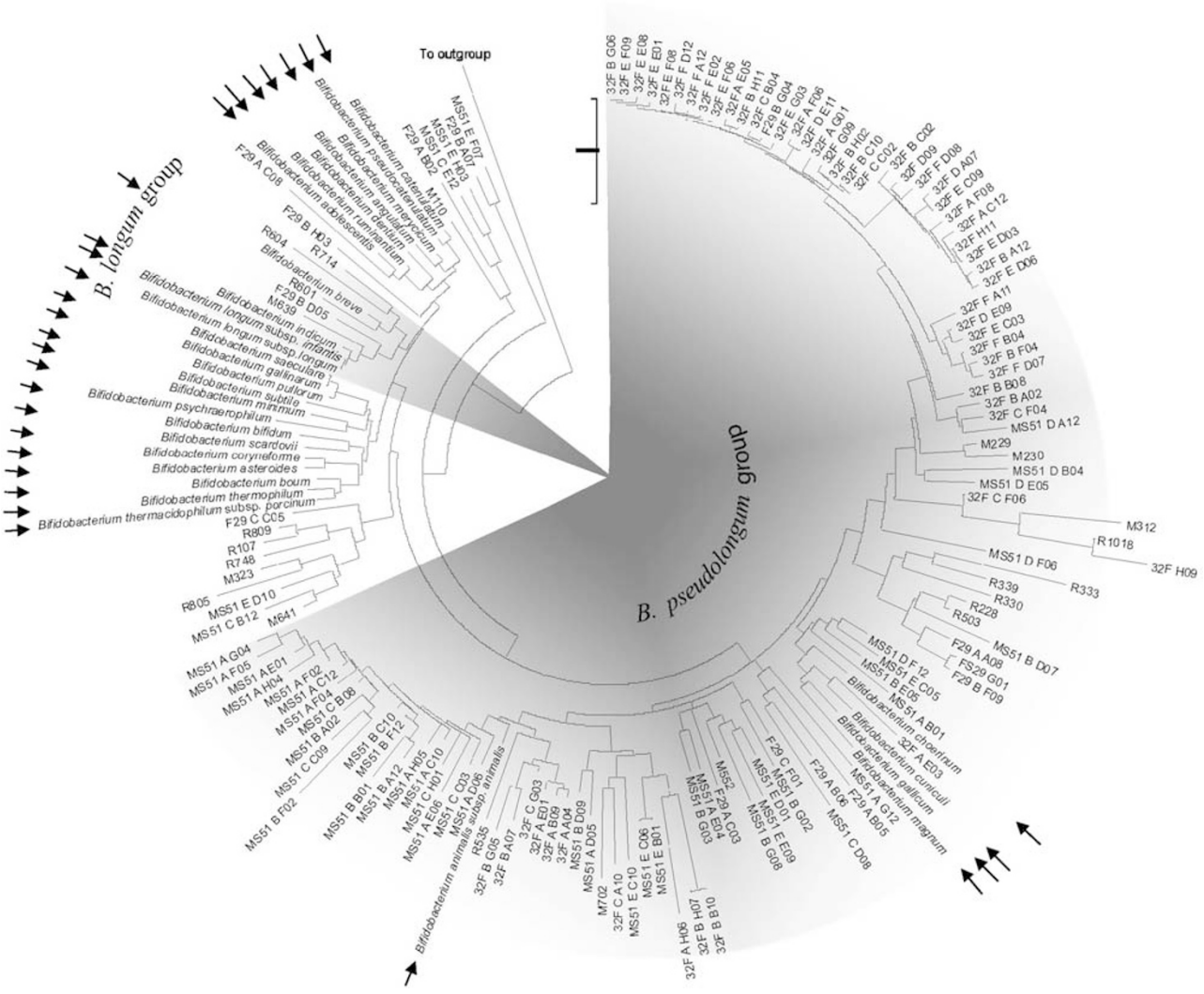

Figure 4 Phylogenetic tree showing the relationships between 151 16S rRNA gene sequences, which were identified as novel bifidobacterial phylotypes and 16S rRNA gene sequences of the currently recognized bifidobacterial species. Highlighted clusters indicate the main phylogenetic groups to which the novel bifidobacteria phylotype belong. The bifidobacteria phylogenetic groups are indicated. Known species of bifidobacteria are highlighted with arrows.

Committee for the Re-evaluation of the Species Definition in Bacteriology (Stackebrandt *et al.*, 2002). Furthermore, as more than half of the phylotypes detected in this survey were represented only once it can be assumed that there may have been many more novel phylotypes present in low abundance that went undetected.

These observations may indicate that the *Bifidobacteriaceae* family is even more diverse than we have shown here.

Detection of bifidobacteria by quantitative-real time PCR

The quantification of the total *Bifidobacterium* population in mucosal biopsies was performed by quantitative real-time PCR using earlier described genus-specific primers (Gueimonde *et al.*, 2004). Quantitative real-time PCR reactions were performed on MicroAmp optical plates sealed with MicroAmp optical caps (Applied Biosystems, Foster City, CA, USA) and amplifications were carried out in a 7500 Fast Real Time PCR System (Applied Biosystems) using SYBR Green PCR Master Mix (Applied Biosystems). Thermal cycling consisted of an initial cycle of 95 °C for 10 min followed by 40 cycles of 95 °C for 15 sec and 60 °C for 1 min. DNA extracts from cultures of the strain *B. longum* subsp. *longum* NCIMB 8809 were used for standard curves. Samples were analysed in duplicate in at least two independent PCR runs. Total bifidobacteria levels ranged between 240 and 3000 cells per mucosa sample (5–7 mm^2).

Conclusions

This study used a sensitive molecular method to show earlier uncharacterized phylotypes of the bifidobacterial microbiota of the human intestine. Thirteen percent of the sequences acquired in this study had not been identified earlier and some of these sequences display $\geqslant 3\%$ sequence dissimilarity from publicly available bifidobacterial 16S rRNA gene sequences. As bifidobacteria are often employed in several health-promoting foods or probiotic products, further molecular studies are needed to better understand the biological and clinical significance of these putative novel bifidobacteria on the health of the human host, as well as their effect on and their interactions with other mucosal microbial communities.

It is noted that a culture-independent approach such as the one described here provides an in-depth image of the diversity of the bifidobacterial population in the human distal gut, which is different from that obtained through culture-based- or molecular-based (for example, denaturant gel gradient electrophoresis) approaches. These differences occur probably because the latter techniques highlight only those bifidobacteria that can be cultivated on synthetic media or that can be discriminated on the basis of the different melting temperatures of small fragments of their rRNA genes (Turroni *et al.*, 2008, 2009). However, the nutritional requirements of bifidobacteria are still largely unknown, and largely depend on the ecophysiological state of each strain, which is extremely variable in a complex ecosystem like that of the human gut (Ventura *et al.*, 2007). Investigative techniques, such as denaturant gel gradient electrophoresis or temperature gel gradient electrophoresis possibly provide only a partial image of bifidobacterial diversity in the distal gut, because such procedures do not analyze near-complete rRNA gene sequences as was done for the microbiomics approach described here, but, rather, the different melting behavior of relatively small fragments of the rRNA gene (Turroni *et al.*, 2008).

In this investigation, bifidobacterial sequences in the human intestine were not just a random sampling of bacterial sequences from fecal communities. In fact, the use of colonoscopic biopsies provides a more detailed image of the diversity of bifidobacterial population occurring in the autochthonous (bifidobacteria that naturally resides in the human intestine) intestinal microbiota. However, caution should be taken in interpreting these results because the mucosa samples used were not subjected to an extensive washing. Thus, apart from mucosa-adherent bifidobacterial, luminal bifidobacteria populations might have been sampled as well. Furthermore, the biopsies were derived from different participants of different ages. Further studies should include a more detailed spacial (anatomical) and temporal analysis of the intestinal bifidobacterial community structure within participants of different ethnic and racial backgrounds. In addition, a better understanding of indigenous bifidobacterial communities at healthy and diseased sites could shed light on the perceived beneficial effects of these microorganisms.

Acknowledgements

This work was financially supported by Parmalt spa, Italy, by the Italian Award for Outstanding Young Researcher Scheme 'Incentivazione alla mobilità di studiosi stranieri e italiani residente all'estero' 2005–2009 and a Marie Curie Reintegration Grant (MERG-CT-2005-03080) to MV, and by the Science Foundation Ireland CSET award to the Alimentary Pharmabiotic Centre and a DAF/HRB FHRI award to the ELDERMET project to DV and FS. We also thank Dr Carlos Canchaya for bioinformatics support and helpful discussions.

References

Bourget N, Simonet JM, Decaris B. (1993). Analysis of the genome of the five *Bifidobacterium breve* strains: plasmid content, pulsed field gel electrophoresis genome size estimation and *rrn* loci number. *FEMS Microbiol Lett* **110**: 11–20.

Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA *et al.* (2003). The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res* **31**: 442–443.

Comstock LE. (2007). Microbiology: the inside story. *Nature* **448**: 542–544.

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M *et al.* (2005). Diversity of the human intestinal microbial flora. *Science* **308**: 1635–1638.

Fanaro S, Vigi V, Chierici R, Boehm G. (2003). Fecal flora measurements of breastfed infants using an integrated transport and culturing system. *Acta Paediatr* **92**: 634–635.

Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS *et al.* (2006). Metagenomic analysis of the human distal gut microbiome. *Science* **312**: 1355–1359.

Gueimonde M, Tolkko S, Korpimaki T, Salminen S. (2004). New real-time quantitative PCR procedure for quantification of bifidobacteria in human fecal samples. *Appl Environ Microbiol* **70**: 4165–4169.

Harmsen HJ, Wildeboer-Veloo AC, Raangs GC, Wagendorp AA, Klijn N, Bindels JG *et al.* (2000). Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods. *J Pediatr Gastroenterol Nutr* **30**: 61–67.

Holben WE, Feris KP, Kettunen A, Apajalahti JH. (2004). GC fractionation enhances microbial community diversity assessment and detection of minority populations of bacteria by denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **70**: 2263–2270.

Kaufmann P, Pfefferkorn A, Teuber M, Meile L. (1997). Identification and quantification of Bifidobacterium species isolated from food with genus-specific 16S rRNA-targeted probes by colony hybridization and PCR. *Appl Environ Microbiol* **63**: 1268–1273.

Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A *et al.* (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**: 169–181.

Ley RE, Peterson DA, Gordon JI. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**: 837–848.

Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS *et al.* (2008). Evolution of mammals and their gut microbes. *Science* **320**: 1647–1651.

Lozupone C, Hamady M, Knight R. (2006). UniFrac-an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**: 371.

Ouwehand AC, Salminen S, Isolauri E. (2002). Probiotics: an overview of beneficial effects. *Antonie Van Leeuwenhoek* **82**: 279–289.

Palmer C, Bik EM, Digiulio DB, Relman DA, Brown PO. (2007). Development of the human infant intestinal microbiota. *PLoS Biol* **5**: e177.

Poxton IR, Brown R, Sawyerr A, Ferguson A. (1997). Mucosa-associated bacterial flora of the human colon. *J Med Microbiol* **46**: 85–91.

Singleton DR, Furlong MA, Rathbun SL, Whitman WB. (2001). Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Appl Environ Microbiol* **67**: 4374–4376.

Stach JE, Bull AT. (2005). Estimating and comparing the diversity of marine actinobacteria. *Antonie Van Leeuwenhoek* **87**: 3–9.

Stackebrandt E, Rainey FA, Ward-Rainey NL. (1997). Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. *Int J Syst Bacteriol* **47**: 479–491.

Stackebrandt E, Frederiksen W, Garrity GM, Grimont PA, Kämpfer P, Maiden MC *et al.* (2002). Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**: 1043–1047.

Suau A, Bonnet R, Sutren M, Godon JJ, Gibson GR, Collins MD *et al.* (1999). Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl Environ Microbiol* **65**: 4799–4807.

Swidsinski A, Loening-Baucke V, Theissig F, Engelhardt H, Bengmark S, Koch S *et al.* (2007). Comparative study of the intestinal mucus barrier in normal and inflamed colon. *Gut* **56**: 343–350.

Turroni F, Foroni E, Pizzetti P, Giubellini V, Ribbera A, Merusi P *et al.* (2009). Exploring the diversity of bifidobacterial population in the human intestinal tract. *Appl Environ Microbiol* (in press).

Turroni F, Ribbera A, Foroni E, van Sinderen D, Ventura M. (2008). Human gut microbiota and bifidobacteria: from composition to functionality. *Antonie Van Leeuwenhoek* **94**: 35–50.

van der Waaij LA, Harmsen HJ, Madjipour M, Kroese FG, Zwiers M, van Dullemen HM *et al.* (2005). Bacterial population analysis of human colon and terminal ileum biopsies with 16S rRNA-based fluorescent probes: commensal bacteria live in suspension and have no direct contact with epithelial cells. *Inflamm Bowel Dis* **11**: 865–871.

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.

Ventura M, Canchaya C, Del Casale A, Dellaglio F, Neviani E, Fitzgerald GF *et al.* (2006). Analysis of bifidobacterial evolution using a multilocus approach. *Int J Syst Evol Microbiol* **30**: 734–759.

Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF *et al.* (2007). Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* **71**: 495–548.

Wang M, Ahrné S, Jeppsson B, Molin G. (2005). Comparison of bacterial diversity along the human intestinal tract by direct cloning and sequencing of 16S rRNA genes. *FEMS Microbiol Ecol* **54**: 219–231.

Watve MG, Gangal RM. (1996). Problems in measuring bacterial diversity and a possible solution. *Appl Environ Microbiol* **62**: 4299–4301.

Woese CR. (1987). Bacterial evolution. *Microbiol Rev* **51**: 221–271.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)