environmental microbiology

Environmental Microbiology (2013) **15**(1), 211–226 doi:10.1111/j.1462-2920.2012.02845.x

Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure

Manuel Ferrer,[1†] Alicia Ruiz,[2] Francesca Lanza,[2] Sven-Bastiaan Haange,[3,4,5] Andreas Oberbach,[3,4] Holger Till,[3,4] Rafael Bargiela,[1] Cristina Campoy,[6] Maria Teresa Segura,[7] Michael Richter,[8] Martin von Bergen,[5,9] Jana Seifert[5***†] and Antonio Suarez[2*†]**

[1]*CSIC, Institute of Catalysis, 28049 Madrid, Spain.*
[2]*Department of Biochemistry and Molecular Biology, Centro de Investigación Biomédica, University of Granada, Granada, Spain.*
[3]*Department of Pediatric Surgery, University Hospital of Leipzig, Leipzig, Germany.*
[4]*Integrated Research and Treatment Center (IFB) Adiposity Diseases, University of Leipzig, Leipzig, Germany.*
[5]*Department of Proteomics, UFZ – Helmholtz Centre for Environmental Research, Leipzig, Germany.*
[6]*Department of Pediatrics, Centro de Investigación Biomédica, University of Granada, Granada, Spain.*
[7]*Division of Metabolic and Nutritional Medicine, Ludwig Maximilians University of Munich Medical Centre, Munich, Germany.*
[8]*Ribocon GmbH, Bremen, Germany.*
[9]*Department of Metabolomics, UFZ – Helmholtz Centre for Environmental Research, Leipzig, Germany.*

Summary

Recent research has disclosed a tight connection between obesity, metabolic gut microbial activities and host health. Obtaining a complete understanding of this relationship remains a major goal. Here, we conducted a comparative metagenomic and metaproteomic investigation of gut microbial communities in faecal samples taken from an obese and a lean adolescent. By analysing the diversity of 16S rDNA amplicons (10% operational phylogenetic units being common), 22 Mbp of consensus metagenome sequences (~ 70% common) and the expression profiles of 613 distinct proteins (82% common), we found that in the obese gut, the total microbiota was more abundant on the phylum *Firmicutes* (94.6%) as compared with *Bacteroidetes* (3.2%), although the metabolically active microbiota clearly behaves in a more homogeneous manner with both contributing equally. The lean gut showed a remarkable shift towards *Bacteroidetes* (18.9% total 16S rDNA), which become the most active fraction (81% proteins). Although the two gut communities maintained largely similar gene repertoires and functional profiles, improved pili- and flagella-mediated host colonization and improved capacity for both complementary aerobic and anaerobic *de novo* B_{12} synthesis, 1,2-propanediol catabolism (most likely participating in *de novo* B_{12} synthesis) and butyrate production were observed in the obese gut, whereas bacteria from lean gut seem to be more engaged in vitamin B_6 synthesis. Furthermore, this study provides functional evidence that variable combinations of species from different phyla could 'presumptively' fulfil overlapping and/or complementary functional roles required by the host, a scenario where minor bacterial taxa seem to be significant active contributors.

Introduction

The human distal gut harbours a vast ensemble of microbes that perform vital processes for human physiology and nutrition, although these are not yet completely understood (O'Hara and Shanahan, 2006; Hattori and Taylor, 2009). It is increasingly evident that gut microbes play a pivotal role in host carbohydrate, lipid and amino acid metabolisms and in the production of vitamins (Hooper and Gordon, 2001), indicating the existence of an extensive trans-genomic, trans-mural co-metabolism of multiple substrates, including those involved in host metabolic regulation (Nicholson *et al.*, 2005; Marchesi, 2011). The distal human intestine can be considered an anaerobic bioreactor with trillions of bacteria that add a vast gene catalogue to host genetics, providing complementary metabolic pathways for energy harvesting, food digestion, detoxification, production of bioactive compounds and

Received 8 May, 2012; revised 6 July, 2012; accepted 15 July, 2012. For correspondence. *E-mail asuarez@ugr.es; Tel. (+34) 958 241000-20318; Fax (+34) 958 819132; **E-mail mferrer@icp.csic.es; Tel. (+34) 915 854872; Fax (+34) 915 854760; ***E-mail jana.seifert@ufz.de; Tel. (+49) 341 2351352; Fax (+49) 341 2351787. †These authors contributed equally to this work.

assimilation of otherwise inaccessible nutrients from our diet (Qin *et al*., 2010). Recent studies showed that each human has a unique and relatively stable gut microbiota dominated by the *Bacteroidetes* and *Firmicutes* divisions that constitute over 90% of the phylogenetic categories, together with *Actinobacteria*, *Proteobacteria* and *Verrucomicrobia* as minority members, most of which have not yet been isolated or characterized (Marchesi, 2010; Arumugam *et al*., 2011). It should be noted that these patterns were apparent only in data averaged over many specimens. Increasing evidence suggests that microbiota status is linked to inflammatory disorders, human metabolism and immune system activity (Turnbaugh *et al*., 2006; Peterson *et al*., 2008) and that its composition can be disrupted by external factors, such as antibiotic treatment (Jernberg *et al*., 2010) and dietary modulation (Hildebrandt *et al*., 2009). Furthermore, specific strains of the gut microbiota and/or supplied probiotics have been shown to decrease intestinal inflammation and to normalize dysfunctions of the gut mucosa (Cani and Delzenne, 2011).

Obesity is an enormous public health problem that arises as a consequence of complex socioeconomic and genetic factors that favour increased dietary intake and reduced physical activity, ultimately disrupting energy homeostasis (Turnbaugh and Gordon, 2009; Swinburn *et al*., 2011). Recent experimental evidence revealed that the *ob/ob* leptin-deficient obese mouse has a microbiome with a 50% reduction in *Bacteriodetes* and a concurrent increase in *Firmicutes* (Turnbaugh *et al*., 2006). Moreover, comparative metagenomics revealed enrichment in genes that encode enzymes involved in the initial steps of breaking down indigestible dietary polysaccharides, starch/sucrose metabolism, galactose metabolism and butanoate metabolism, which might increase the capacity to harvest energy from the diet (Turnbaugh *et al*., 2006). Approximately 400 genes in metabolic pathways have been shown to be enriched or depleted in the human gut microbiome of obese individuals compared with lean controls (Turnbaugh *et al*., 2009). An interesting example of these is an enrichment of phosphotransferase systems responsible for microbial processing of carbohydrates (Booijink *et al*., 2010). Another recent study identified H_2-producing *Prevotellaceae* and the H_2-using *Methanobacteriales* in the guts of obese individuals (Zhang *et al*., 2009). Methanogens are reported to increase the extraction of energy by the host from otherwise indigestible polysaccharides. In obese humans, the pattern of microbial diversity and gene-encoded functions is more intricate (Turnbaugh *et al*., 2009; Musso *et al*., 2011), and contradictory reports on the composition of microbial communities complicate the identification of functional and molecular hotspots associated with obesity (Tschöp *et al*., 2009; Elli *et al*., 2010; Ley, 2010; Schwiertz *et al*., 2010).

Overall, the understanding of the underlying microbial metabolic contribution to obesity is sparse. The main reason for this is that the majority of investigations rely on indirect evidence of DNA-based approaches, which only provide the blueprint, or an inventory of genes of human-associated microbial communities, but lack the proof of functionality that might be derived from studies of gene expression and protein synthesis. To solve this problem, techniques such as two-dimensional electrophoresis (2-DE) coupled to high-throughput mass spectrometry-based analytical platforms and direct shotgun proteomic approaches have been used to separate and identify thousands of proteins from at least 14 human faeces-associated bacterial phylotypes (Klaassens *et al*., 2007; Verberkmoes *et al*., 2009; Cantarel *et al*., 2011; Rooijers *et al*., 2011; Kolmeder *et al*., 2012) among the estimated 5000 bacterial species in the human microbiome. Additionally, apart from whole-gut metaproteome analyses, defined proteomes from prominent human intestinal symbionts have been reported, the results of which have revealed that specific microbes, such as *Bacteroides thetaiotaomicron*, contain an elaborate environment-sensing apparatus (Xu *et al*., 2004). Accordingly, the identification of highly expressed proteins from human faecal samples has been aided by creating extensive databases containing the genome sequences of human-associated microbes and human-associated microbiome sequences (Rooijers *et al*., 2011).

Taking all of the above information into consideration, we performed a thorough and holistic (or eco-systems biology) phylogenetic and functional analysis of the gut microbial communities of one obese and one lean individual (for characteristics see *Experimental procedures*). Instead of performing an extensive analysis of the gene repertoire, special emphasis is given to the identification and analysis of active bacterial members and their expressed proteins present in lean and obese microbiomes and how they might form metabolically active networks operating in each community.

Results and discussion

Bacterial diversity and composition blueprints of lean and obese microbiomes

DNA isolated from each microbial community was assayed using a PCR-based 16S rDNA gene diversity survey. More than 300 clones were generated in each library and were screened by restriction fragment length polymorphism (RFLP) analysis using the enzyme MspI. Additionally, we used the 16S rDNA gene partial sequences obtained in the metagenome survey. For this purpose, we just used those sequences with a length > 600 nucleotides. All shorter sequences were discarded.

Table 1. Diversity data and indices for the 16S rDNA libraries generated from 'Obese' and 'Lean' individuals.

	'Lean'	'Obese'
No. of clones	165	185
No. of taxa	52	32
Dominance (*D*)	0.05	0.13
Shannon-Wiener index (*H*)	3.39	2.59
Simpson's index (1-*D*)	0.95	0.87
Equitability index (*J*)	0.86	0.75
Evenness (*E*)	0.57	0.43
Good's coverage (%)	69.4	82.7

A total of 22 sequences were obtained, and analysed. Based on sequence similarity, 77 operational phylogenetic units (OPUs) were detected (Table 1): 8 conformed to the common set and 45 and 24 were only found in 'Lean' and 'Obese' respectively. Higher values for Shannon-Wiener's, Simpson's and equitability indices indicated that the 'Lean' library was more diverse and less dominated by a few abundant populations. Representatives of the OPUs were selected for sequencing and phylogenetic affiliation. The overall phylogenetic composition of the libraries is shown in Figs 1 and 2, in which one OPU was derived from *Archaea* and the rest branched to four phyla of the domain *Bacteria* in RDP: *Firmicutes*, *Bacteroidetes*, *Actinobacteria* and *Proteobacteria*. Most of the OPUs ('Lean': 32; 'Obese': 22) were closely related (97% identity cut-off) to uncultured gut bacteria found in other diversity surveys (Eckburg *et al*., 2005). The biodiversity in the 'Obese' and 'Lean' libraries was dominated by sequences belonging to *Firmicutes* (94.6% and 78.8% respectively); however, the 'Lean' library had more OPUs affiliated with *Bacteroidetes* (18.9% in 'Lean' versus 3.2% in 'Obese'), and the 'Lean' OPUs were more heterogeneous than those in the 'Obese' library (Fig. 1). The differences in the compositions of our libraries agree with the diversity profile of the EVASYON study (Santacruz *et al*., 2009) that reported lower *Bacteroides-Prevotella* abundance and greater frequencies of *Clostridium* cluster XIVa members in obese subjects. Although there are contradictory reports on the patterns of gut microbial communities in obese humans (Ley, 2010), and the 77 identified OPUs described here only capture a portion of the predicted microbial diversity of the intestinal microbiota, the compositions of our samples are consistent with the 'low *Bacteroidetes*/high *Firmicutes*' hypothesis initially proposed by Turnbaugh and colleagues (2006) that associated obesity with decreased bacterial diversity and reduced representation of *Bacteroidetes*.

In the 'Obese' library, *Firmicutes* OPUs were dominated by the class *Clostridia* (Fig. 2), with 30.8% of 16S gene sequences branching to the *Lachnospiraceae* lineage (*Clostridium* cluster XIVa), followed by *Ruminococcaceae* (*Clostridium* cluster IV, 16.2%), *Clostridiaceae* (9.2%), and unclassified *Clostridiales* (1.6%). Members of the phylum *Firmicutes* branched to the classes *Erysipelotri-*

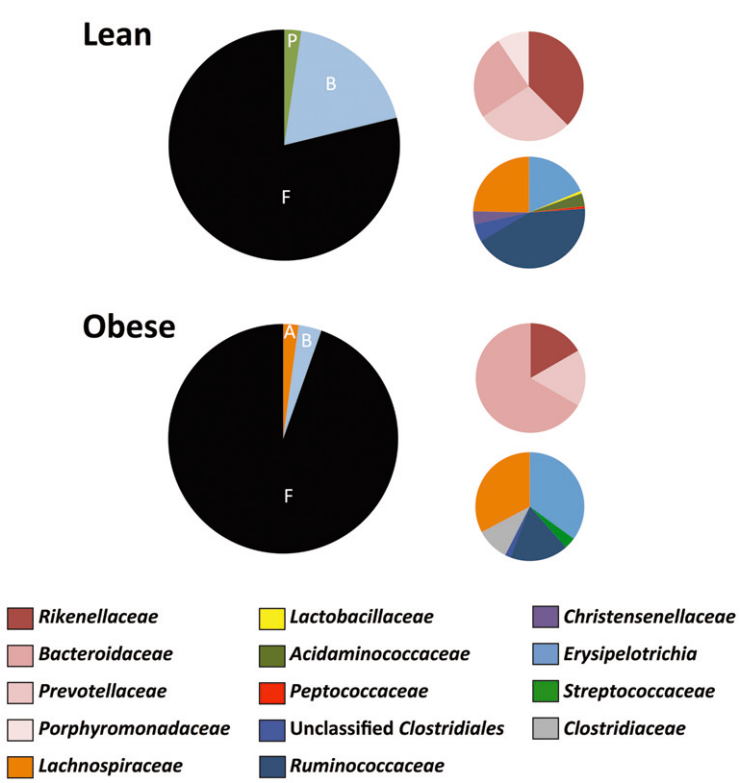

Fig. 1. Phylogenetic compositions of the clone libraries retrieved from 'Obese' and 'Lean' individuals. Clones belonging to the same RFLP group were affiliated according to the sequences of the representative clones. At the level of bacterial division or family, sectors represent the percentages of each group of the total number of clones in each library. Smaller circles represent the abundances of clones belonging to different families within the divisions *Bacteroidetes* (up) and *Firmicutes* (down), with respect to the total of clones from these divisions. Labels are as follows: A, *Actinobacteria*; B, *Bacteroidetes*; F, *Firmicutes*; P, *Proteobacteria*.

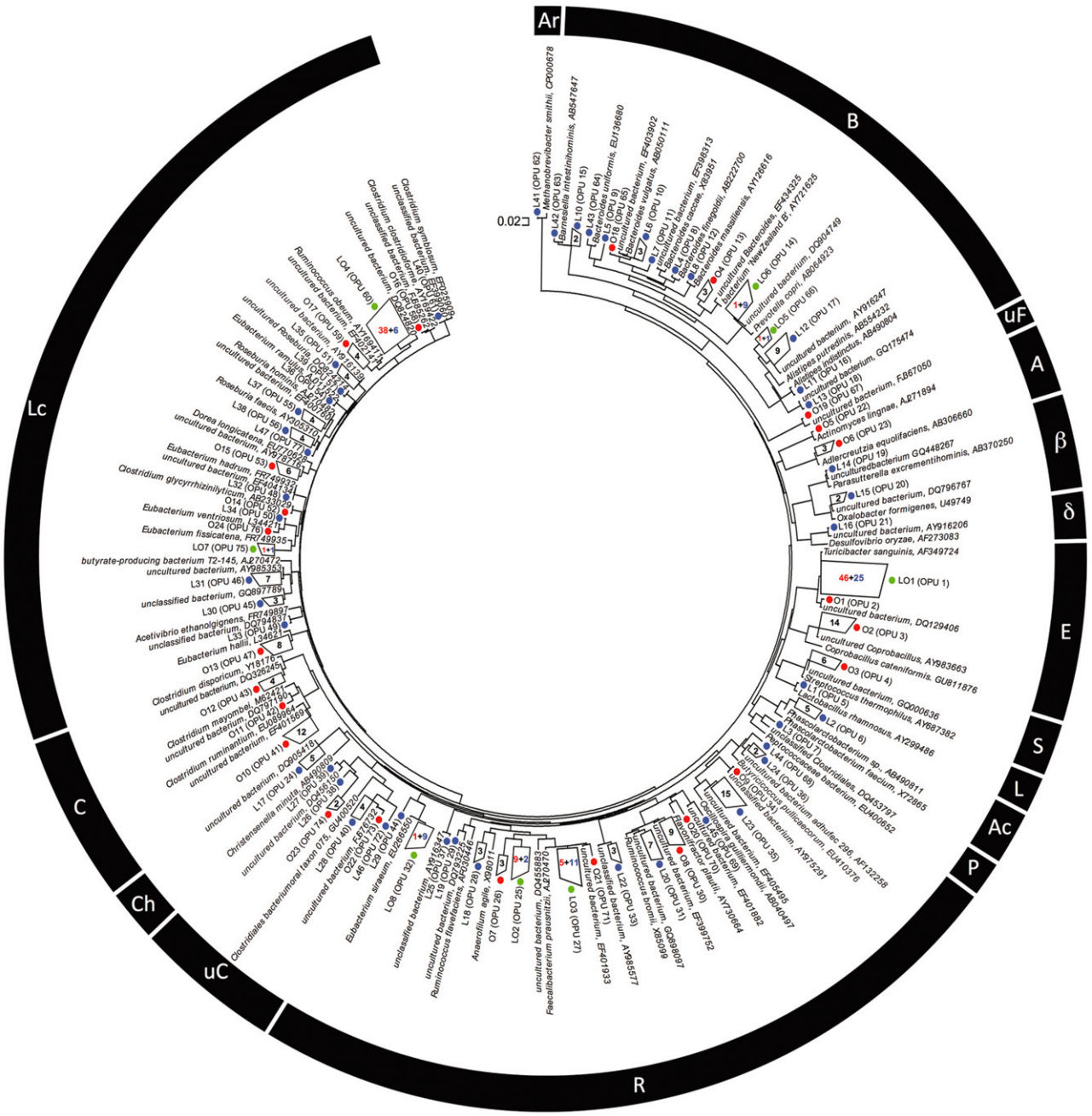

Fig. 2. Dendrogram of the 16S rDNA gene clones (OPUs). Clones are designated as 'O' (red), 'L' (blue), or 'LO' (green) for 'Obese', 'Lean', or 'Lean'/'Obese' microbiota respectively. The tree was calculated with nearly complete reference sequence data obtained from EMBL and RDP databases. Closest type strains or isolates of each OPU were added. The number of sequences contained in each OPU is specified. Bacterial lineages are as follows: A, *Actinobacteria*; Ac, *Acidaminococcaceae*; Ar, *Archaea*; B, *Bacteroidetes*; β, *Betaproteobacteria*; C, *Clostridiaceae*; Ch, *Christensenellaceae*; δ, *Delta-proteobacteria*; E, *Erysipelotrichaceae*; L, *Lactobacillaceae*; Lc, *Lachnospiraceae*; P, *Peptococcaceae*; R, *Ruminococcaceae*; S, *Streptococcaceae*; uC, unclassified *Clostridiales*; uF, unclassified *Firmicutes*.

chia (33%) and *Bacilli* (family *Streptococcaceae*, 3.2%). Bacterial phylogeny in the 'Obese' library included the phylum *Actinobacteria* (2.2%) and *Bacteroidetes* members (3.2%). The 'Lean' microbial community was more diverse, and a shift to *Clostridium* cluster IV members in the class *Clostridia* was observed. Representatives of the *Ruminococcaceae* lineage (33.5%) were more abundant than were *Lachnospiraceae* (19.4%), followed by unclassified *Clostridiales* (4.1%), *Christensenellaceae* (2.9%) and *Peptococcaceae* (0.8%). The phylum *Firmicutes* was completed with representatives of the class *Bacilli* (*Lactobacillaceae*, 0.8%) and *Acidaminococ-*

Table 2. General features of the metagenomes of the human faeces from 'Lean' and 'Obese' subjects.

Parameter	'Lean'	'Obese'
Size (bp)	11 889 690	10 731 170
Contigs	14 607	12 074
Average contig size (bp)	814	889
Average GC content (%)	48.59	44.48
Protein-coding genes (CDS)	21 230	18 099
Average CDS size (bp)	470	456
tRNAs	164	186
rRNAs	80	74
Total RNAs	244	260
ORFs with predicted function	10 614	9 122
Hypothetical	4 347	3 525
Conserved hypothetical	6 269	5 452
Assigned to COGs	10 176	8 714
KEGG	14 684	12 322

caceae family (2.9%). Relatively few representatives of *Proteobacteria* were detected (2.4%), and these branched into *Betaproteobacteria* and *Deltaproteobacteria*. Finally, 'Lean' and 'Obese' libraries shared 8 OPUs populated by the *Prevotella copri*, *Alistipes putredinis*, *Eubacterium siraeum*, *Eubacterium fissicatena*, *Turicibacter sanguinis*, *Ruminoccous obeum* and *Faecalibacterium prausnitzii* clades.

Due to the clear shifts in community structure detected among the faecal samples taken from obese and lean individuals, it is important to compare the metabolically active bacteria and to establish the 'presumptive' metabolic capabilities that characterized them, which was further evidenced by using a metaproteomic survey.

Active faecal microbiota in lean and obese guts

Using a complementary approach, DNA was sequenced using a Roche GS FLX DNA sequencer. Pyrosequencing of the metagenomes produced raw DNA sequences that were assembled into 11.9 Mbp (15 319 contigs) and 10.7 Mbp (13 206 contigs) for 'Lean' and 'Obese' microbial metagenomes respectively. Further information regarding the metagenome features is given in Table 2. With a threshold of greater than 95% identity and an aligned length of more than 150 bp, the majority of predicted genes in the metagenomes could be phylogenetically assigned, the analysis of which returned results comparable to those found in the 16S rDNA assignments at the phylum level, i.e. most fragments were assigned to *Bacteroidetes* and *Firmicutes* in 'Lean' and 'Obese' respectively (data not shown). A total of 21 230 (for 'Lean') and 18 099 (for 'Obese') protein-coding sequences (CDS) were predicted. Approximately half of the predicted genes ('Lean': 10 176; 'Obese': 8714) were assigned to clusters of orthologous groups (COG) protein families, and approximately 70% were assigned to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Table 2). Neither the COG nor the KEGG profile exhibited prominent overall differences (data not shown), and among the CDS predicted, 7831 (or 37%) and 6054 (or 33%) were unique for the 'Lean' and 'Obese', respectively, whereas 25 444 formed a common set. With a maximum *E*-value criterion of 10^{-5} and an alignment length of 75% minimum, 21% and 22% of the 'Lean' and 'Obese' sequences, respectively, had no sequence similarity; these numbers are in consonance with previous studies, in which almost 20–25% of all recognized proteins in human gut microbiomes were classified as 'hypothetical proteins' as they did not belong to well-characterized protein families (Kurokawa *et al*., 2007; Verberkmoes *et al*., 2009; Ellrott *et al*., 2010; Qin *et al*., 2010). Additionally, *c.* 30% and 48% were similar to proteins of unknown function ('conserved hypothetical'), which is typical in metagenomic projects (Sleator *et al*., 2008). Thus, an important fraction of the microbiomes remains unknown, and their metabolisms are difficult to unravel.

Using a complementary approach, the metaproteomes of 'Lean' and 'Obese' were investigated via one-dimensional (1-DE) gel-based pre-separation of proteins and subsequent tryptic digestion, fractionation and identification of the resulting peptides by a nano-UPLC system coupled to an LTQ-Orbitrap mass spectrometer. The DNA metagenome sequences and a database containing sequence information of human-associated microbes (Rooijers *et al*., 2011) were used as templates. A total of 613 proteins were unambiguously quantified (Table S1) following the criteria described in *Experimental procedures*, of which 71 were 'Lean' specific, 37 were 'Obese' specific and 505 formed a common set. The number of identified proteins is within common ranges that have been observed for other human gut-associated communities and is only three times lower than that observed for cultivable organisms. For example, more than 200 protein spots were visualized in 2D gels of proteins collected from faecal samples of infants (Klaassens *et al*., 2007), and by using shotgun proteomic approaches 600–900 proteins were identified in two faecal samples collected from two healthy female identical twins (Verberkmoes *et al*., 2009) and 1790 in three healthy, omnivorous females (Kolmeder *et al*., 2012). The identified proteins only capture a portion of the predicted coding capacity of the intestinal microbiome; however, as the majority of the spectra could be assigned to a taxonomic and functional annotation based on highly similar homologues, the metaproteomic approach applied here allowed us to compare taxonomic annotations to evaluate the differences between the contributions of particular groups of organisms to the global community, that is, to identify metabolically active members, and to predict the importance of particular sets of proteins and the microbes containing them for the

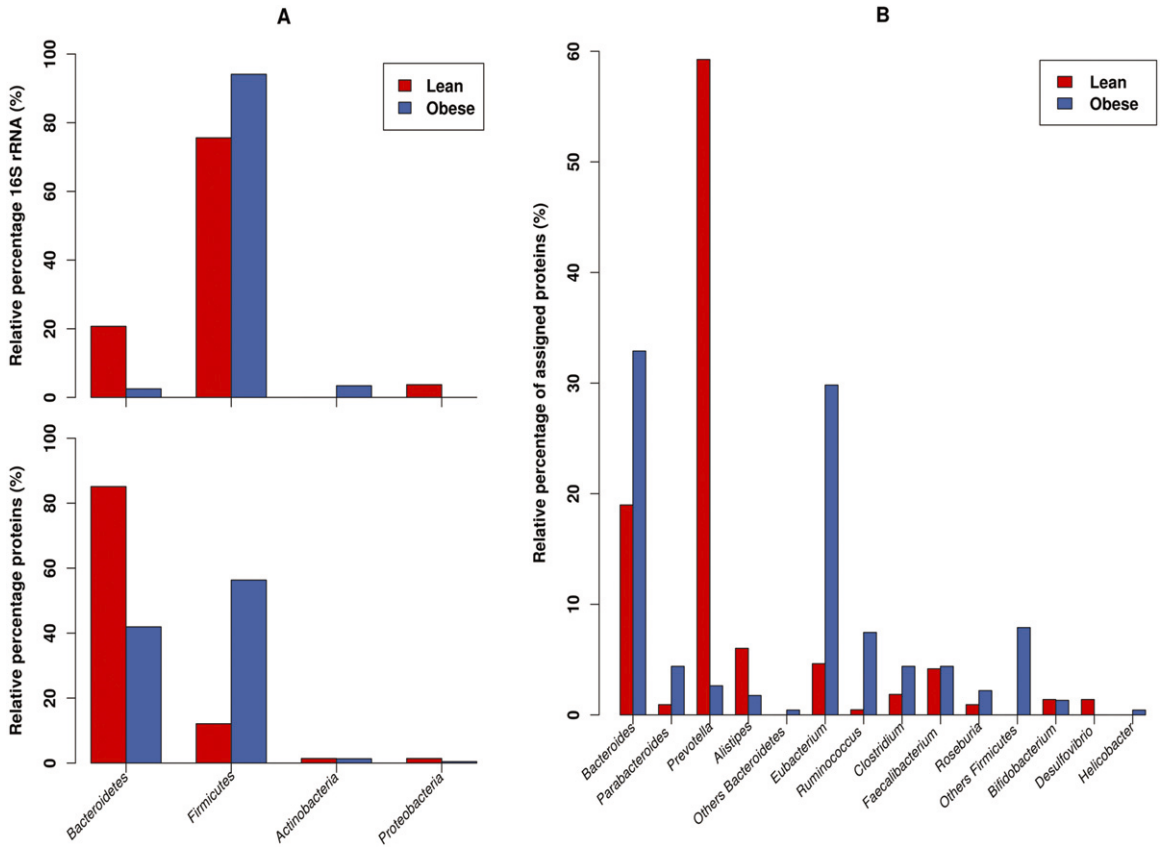

Fig. 3. Phylogenetic classification based on phylogenetic and metaproteome analysis.
A. Taxonomic bins for bacterial 16S rDNA gene sequences (OPUs) and phylogenetic categories at the phylum level of proteins that were identified in the metaproteomes (as shown in Table S1).
B. Phylogenetic categories at the genus level of proteins that were identified in the metaproteomes (as shown in Table S1). Assignments were performed as described in *Experimental procedures*.

overall functioning of the community. Moreover, a label-free quantification based on protein intensities has been adopted to quantify the relative abundance of proteins in analysed shotgun proteomic data.

To ascertain which bacteria were responsible for the observed protein expression, we determined their taxonomic assignments as described in *Experimental procedures*. As illustrated in Fig. 3A, the compositions of active bacteria for the 'Lean' and 'Obese' samples were clearly different. In the 'Lean' sample there are mainly proteins assigned to *Bacteroidetes* (81%), albeit this phylum represents only about 20% (based on 16S rDNA) of the total community (Fig. 3A), followed by those assigned to *Firmicutes* (12%) and, to a minor extent, *Actinobacteria* and *Proteobacteria* (approximately 3% in total). In the 'Obese' sample the major taxa responsible for protein expression was *Firmicutes* phylum (56%) in agreement with its dominance in the total bacteria (94% 16S rDNA). Active *Bacteroidetes* (42%) are almost equally abundant, with *Actinobacteria* and *Proteobacteria* accounting for only 1.7% of the total hits; the shift towards active *Bacteroidetes* members in 'Obese' was noteworthy as they account only 3.2% of the total community (Fig. 3A). This suggests that minor bacterial taxa may play a significant active role in overall 'Lean' and 'Obese' gut metabolisms, with *Bacteroides* members possibly playing a major active role in both 'Lean' (rDNA/protein abundance ratio of 1:4) and 'Obese' (rDNA/protein abundance ratio of 1:13) guts. At the genus level, proteins attributed to *Prevotella* and *Bacteroides* predominated (59% and 20% respectively) in the 'Lean' gut, whereas the 'Obese' gut was mainly dominated by active *Bacteroides* and *Eubacterium* members (33% and 30% respectively) (Fig. 3B).

The identification of the different active members in both individuals may be of important significance as, in the gut ecosystem, few studies have focused on metatranscriptomics and metaproteomics to provide information on metabolically active bacteria and associated proteins in faecal samples, any of them related to obese-lean comparisons. Turnbaugh and colleagues (2010) focused on the gene expression analysis of faecal samples from a monozygotic twin pair, whereas Booijink and col-

leagues (2010) and Gosalbes and colleagues (2011) studied the faecal metatranscriptome of healthy volunteers using cDNA-AFLP or pyrosequencing respectively. It was found that the two bacterial phyla, *Firmicutes* (*c.* 50%) and *Bacteroidetes* (*c.* 30%), provided the largest number of 16S rRNA transcripts. More recently, a metaproteomic analysis of three healthy, omnivorous female subjects, revealed *Firmicutes* phylum as the most active one (60–86% of assigned spectra) as compared with *Bacteroidetes* (6–11%) (Kolmeder *et al*., 2012). Therefore, the taxonomic classification of the active microbiota in faeces from the 'Lean' sample herein investigated seems to differ from that reported for other human faecal samples.

Functional activities of the microbial communities in lean and obese guts

Among the 613 proteins identified and quantified, only 86 were characterized as hypothetical, predicted or putative in the annotation. A total of 505 proteins were identified in both samples and exhibited an exponential distribution (Fig. 4), with the relative intensities of 144 (from 619- to 2.8-fold) and 191 (from 852- to 3.1-fold) proteins being significantly higher in 'Lean' and 'Obese' respectively. This clearly indicated that both communities displayed considerable heterogeneity in terms of protein expression, in accord with their distinct phylogenetic compositions (Figs 1 and 2). As shown in Table S1, the ten most abundant proteins in 'Lean' were five outer membrane proteins most likely involved in siderophore/ion/sugar binding and transport, one pectate lyase yielding oligosaccharides from galacturonan, and enzymes involved in glycolysis such as aldose 1-epimerase. Among the proteins found only in 'Lean', we identified 5 rubrerythrin/rubredoxins (ferritin-like diiron-binding domain proteins). The physiological role of rubrerythrins, which have also been identified in faecal samples from healthy female identical twins (Verberkmoes *et al*., 2009), has not been identified; however, they have been shown to be protective against oxidative stress. A superoxide dismutase, also involved in the oxidative stress response, was also found among the proteins exclusive to 'Lean'. In contrast, the 10 most abundant microbial proteins in 'Obese' included an α-L-arabinofuranosidase involved in yielding oligosaccharides from arabinans and xylans, the large (α) subunit of B_{12}-dependent diol dehydratase, possibly involved in 1,2-propanediol metabolism, two outer membrane transport proteins, most likely involved in siderophore/ion/sugar binding and transport, a β-subunit DNA-directed RNA polymerase, a fibronectin type 3 domain-containing

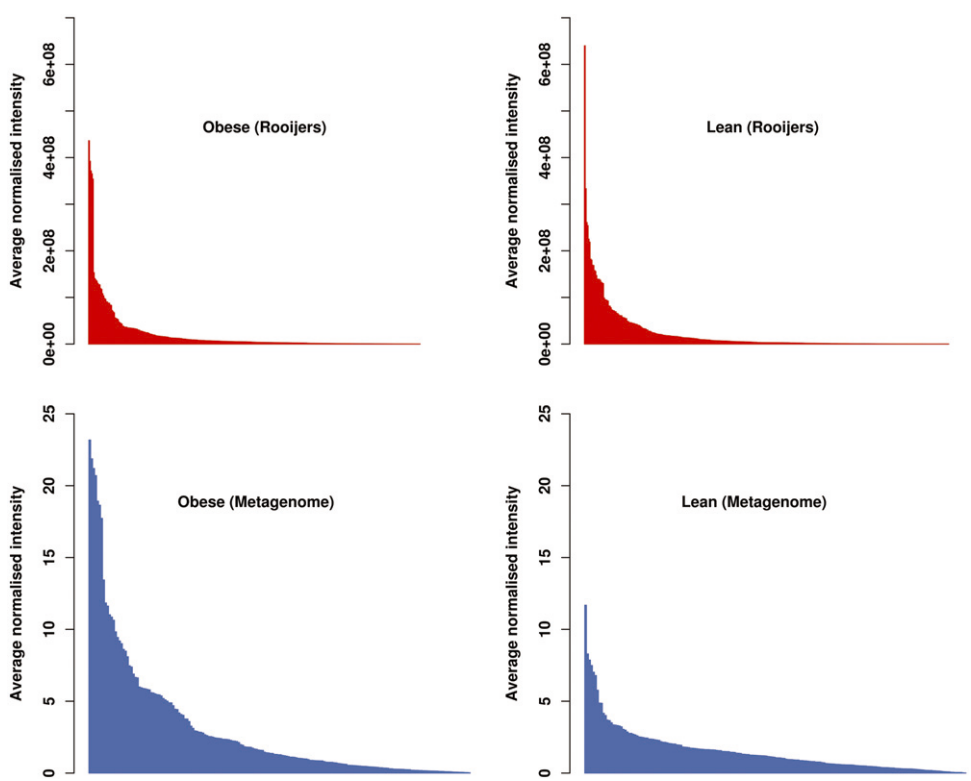

Fig. 4. Normalized intensity distribution of proteins that were identified in 'Obese' (left) and 'Lean' (right) human distal gut metaproteomes. Graphs represent the normalized intensities per each of the identified proteins calculated as described in *Experimental procedures*, using an artificial metagenome based on a set of bacterial and archaeal genomes listed by Rooijers and colleagues (2011) (upper panels) or the metagenomic data from the 'Obese' and 'Lean' samples (lower panels). For details see Table S1.

protein possibly involved in cell adhesion (Jee *et al*., 2002), an IM30-like phage shock protein A that has been demonstrated to maintain the energetic state of the membrane under stress conditions (Kleerebezem *et al*., 1996; Kobayashi *et al*., 2007), and an orotidine 5-phosphate decarboxylase, an enzyme essential to the *de novo* biosynthesis of the pyrimidine nucleotides. Among those exclusively found in 'Obese', the identifications of a CbiK anaerobic cobalt chelatase and a flagellin superfamily protein were noteworthy, and the biological significance of these proteins will be discussed below.

To predict the metabolic potential and to look for significantly over- and under-represented COGs and proteins, a comparative analysis was further used. COG profiling analysis (Fig. 5 and Table S1) demonstrated similarities between both communities, as it has also been reported previously when trying to define a common core microbiota between different subjects (Gill *et al*., 2006; Kurokawa *et al*., 2007). In both cases, the overall COG distribution exhibited a uniform pattern with a high representation of COGs classified into the 'Carbohydrate Metabolism', 'Amino Acid Transport and Metabolism', 'Nucleotide Transport and Metabolism' and 'Energy Production and Conversion' groups; however, clear differences were observed that characterized the 'Obese' gut, that will be discussed below. Significant over-representations of proteins from COGs for 'Cell Motility' and, to a lesser extent, 'Lipid transport and metabolism' and 'Translation, ribosomal structure and biogenesis' were remarkable. Only COG proteins assigned to 'Nucleotide Transport and Metabolism' were more abundant in the 'Lean' gut.

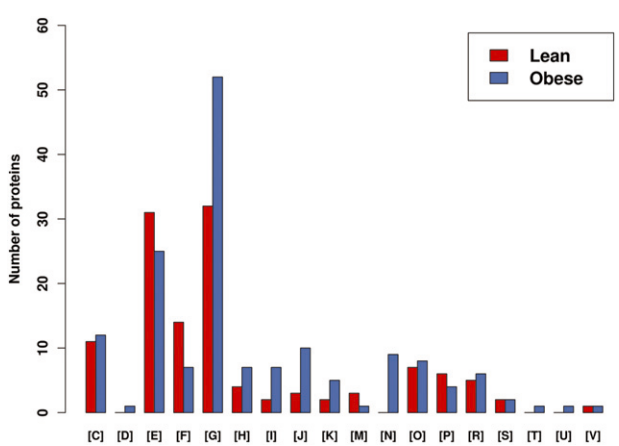

Fig. 5. COG distribution of the proteins detected in the metaproteomes of 'Lean' and 'Obese' human distal gut samples. The COG functional category distribution, based on the identified proteins assigned to particular groups (see details in Table S1), is visualized. COG categories as follows: C, energy production and conversion; D, cell division and chromosome partitioning; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; H, coenzyme metabolism; I, lipid metabolism; J, translation, ribosomal structure and biogenesis; K, transcription; M, cell envelope biogenesis, outer membrane; N, cell motility and secretion; O, post-translational modification, protein turnover, chaperones; P, inorganic ion transport and metabolism; R, general function prediction only; S, Function unknown; T, signal transduction mechanisms.

Shifts associated to potential microbes–host interacting components such as flagellin and pili

The striking abundance [$\log_2$ of the ratios of average intensities ('Obese'/'Lean') from 3.1 to 5.5) of 3 pili-formation proteins and 5 proteins containing the flagellin domain within the 'Cell motility' category was noteworthy in the 'Obese' microbiome, and a flagellin (Tad-like) superfamily protein was also determined to be 'Obese' specific (Fig. 6). Three of these proteins were assigned to *Bacteroides* and 6 to *Firmicutes*. Additionally, a glycoprotein containing the fibronectin type 3 domain, most likely involved in binding extracellular matrix components, such as collagen, fibrin and proteoglycans, was also detected at a significantly higher level in the 'Obese' gut [$\log_2$('Obese'/'Lean') of 8.1]. No representative of proteins belonging to the 'Cell Motility' COG category was identified in the 'Lean' metaproteome. The intestinal flagellin may provide microbes the capacity to be motile, allowing them to better reach their food sources, namely carbohydrates (O'Connell Motherway *et al*., 2011; Kolmeder *et al*., 2012). The fact that the 'Obese' gut may contain a higher amount of carbohydrates may explain flagellin genes being expressed; however, the possibility that abnegation of motility may be an adaptation mechanism of gut microbes to persist in the intestinal environment, as flagella are highly immunogenic, cannot be ruled out. In this context, it has been recently reported that the abundances of genes for the biosynthesis of flagella and chemotaxis in wean-type microbiota are greater than the abundances in adult microbiota (Kurokawa *et al*., 2007). Therefore, this study provided experimental evidence that suggests that not only age but also obesity may be driving forces for lowering or improving cell motility and adhesion and, thus, host colonization and persistence in the human gut. Additional large-scale comparative studies of faecal samples from different individuals may be required to further confirm this hypothesis.

Shifts associated to microbes-mediated biosynthesis of vitamins

In the COG category 'Coenzyme transport and metabolism', which was shown to be twofold enriched in the 'Lean' gut, distinct 'presumptive' capacities for the productions of vitamin B_6 and B_{12} were observed that differentiate the microbial communities residing in 'Lean' and 'Obese' guts respectively (Fig. 6). This was shown by the identifi-

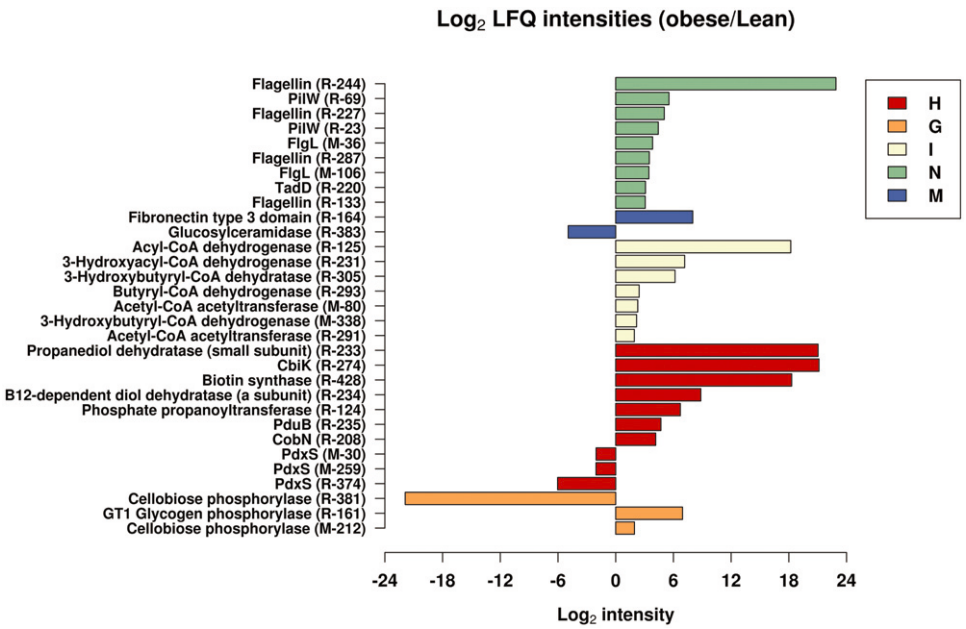

Fig. 6. Relevant proteins detected in the metaproteomes of 'Lean' and 'Obese' human distal gut samples with significant abundance changes. Full details of identified and quantified proteins are provided in Table S1. Graph represent the logarithms of the ratios to the base of 2 of the normalized intensities per each of the identified proteins calculated as described in *Experimental procedures*. For upregulation of protein expression in the 'Obese' environment, a threshold of at least 1.5 in the ratio value was set; for downregulation, a maximum value of –1.5 was set. Protein ID as described in Table S1: R refers to protein identified using an artificial metagenome based on a set of bacterial and archaeal genomes listed by Rooijers and colleagues (2011), and M refers to proteins identified using the metagenomic data. Protein names as follows: PilW, type IV pilus biogenesis/stability protein; FlgL, flagellar hook-associated protein; TadD, Flp pilus assembly protein; CbiK, anaerobic cobalt chelatase; PduB, 1,2-Propanediol utilization protein B; CobN, cobaltochelatase subunit CobN; PdxS, Pyridoxal biosynthesis lyase. H, coenzyme metabolism; G, carbohydrate transport and metabolism; I, lipid metabolism; N, cell motility and secretion; M, cell envelope biogenesis, outer membrane.

cation of three pyridoxal biosynthesis lyases (PdxS) that were expressed at higher levels in the 'Lean' as compared with the 'Obese' gut [$\log_2$('Obese'/'Lean') from –2.0 to –6.0); additionally, a functional partner of PdxS, phosphoribosylaminoimidazole carboxylase (catalytic subunit), was found only in the 'Lean' gut. These proteins most likely all binned to members of *Prevotella*. Both types of proteins are essential in pyridoxine (vitamin B_6) synthesis. In the case of vitamin B_{12} biosynthesis, cobalt is inserted into a macrocycle by a cobaltochelatase, which is specific to either the aerobic (oxygen-dependent) or the anaerobic pathway for cobalamin biosynthesis (Lobo *et al.*, 2008). The identification of two cobaltochelatases in the 'Obese' gut metaproteome, including one aerobic CobN-like binned to a member of the *Bacteroides* genus [$\log_2$('Obese'/'Lean') of 4.2] and one anaerobic CbiK-like binned to *Eubacterium* (only found in 'Obese'), is noteworthy. This suggests that in the 'Obese' gut, vitamin B_{12} most likely is biosynthesized by two distinct active pathways. Possibly, different members of the community are able to complement each other in an aerobic-anaerobic inter alia association for the synthesis of host-important molecules, such as cofactors and vitamins. Although a study with greater statistical power may be required for further interpretation, our results suggest that the biological production of vitamins and the uptake of key metals such as Co^{2+} through microbes in the human gut may be affected positively or negatively in lean and obese individuals. This is important because animals are not capable of synthesizing vitamin B_{12} (among other vitamins), and only bacteria have the enzymes required for its synthesis. Altered capacity for its production may have additional biological implications, including host health and the activation or deactivation of B_{12}-dependent enzymes (Rodionov *et al.*, 2003). This is in agreement with previous observations that reported higher levels of propionate in obese subjects and that propionate fermentation is mediated by a B_{12}-dependent methylmalonyl-CoA mutase (Schwiertz *et al.*, 2010). Our results also agree with previous observations that demonstrated that the distal gut microbiomes of healthy adults are enriched with a variety of COGs involved in the synthesis of essential vitamins, including B_1 and B_6 (Gill *et al.*, 2006); however, our study provides further experimental evidences that suggest that in lean and obese individuals the biosynthesis of multiple vitamins and, in turn, their contributions to the overall gut metabolism, may be altered at different levels through the enrichment or depletion of set of

enzymes. Taken together, these results agree with a recent study by Greenblum and colleagues (2012), which through a 'metagenomic system biology' approach has revealed gain or loss of certain peripheral metabolic enzymes and that lean and obese microbiomes differ primarily in their interface with the host in the way they interact with host metabolisms, as has been also evidenced here (i.e. vitamin metabolic network between gut microbes–host).

In relation to the synthesis of vitamin B_{12}, the identification of a *Firmicutes*-binned 1,2-propanediol utilization protein, a PduB protein, in the 'Obese' should also be highlighted (Fig. 6). The *pdu* genes have been shown to be contiguous and co-regulated with the cobalamin (*cob*) (B_{12}) biosynthetic genes, indicating that propanediol catabolism is the primary reason for *de novo* B_{12} synthesis in some organisms (for details see Bobik *et al*., 1999 and references therein). The enzymes of the propanediol utilization (*pdu*) operon may work during either aerobic or anaerobic growth (Krooneman *et al*., 2002), which also agrees with the aerobic and anaerobic features of the two cobalt chelatases identified in the obese microbiome (see above) and suggests the presence of at least two distinct active bacteria operating in vitamin B_{12} synthesis, which is also important for 1,2-propanediol utilization. The PduB proteins have been shown to be involved in the formation of polyhedral bodies that contain genes (i.e. dehydrates) devoted to 1,2-propanediol degradation, and their roles in protecting essential proteins from oxygen and in bacterial survival and niche establishment have been reported (Krooneman *et al*., 2002). This agrees with the further identification of the small and large subunits of 1,2-propanediol dehydrates and a phosphate propanoyltransferase in the 'Obese' metaproteome (Fig. 6) and suggests the presence of 1,2-propanediol-fermenting, facultative anaerobic and hetero-fermentative bacteria in the 'Obese' gut.

Shifts associated to microbes-mediated hydrolysis and fermentation of dietary-fibre or host-derived glycans

Butyrate kinase was the most highly enriched COG in a human gut metagenomic study by Gill and colleagues (2006) after examining DNA sequences obtained from the faecal DNA of two healthy adults. Although we did not identify butyrate kinases in the shotgun metaproteomes, we did find that three 3-hydroxybutyryl-CoA dehydratases, two butyryl-CoA dehydrogenases and two thiolases (acetyl-CoA acetyltransferases), all possibly involved in butyrate metabolism, had relatively higher expression levels in the 'Obese' microbiome [$\log_2$('Obese'/'Lean') ranging from 1.9 to 7.2 and one specific; Fig. 6]. These proteins contributed to the approximate fourfold overabundance of proteins within the 'Lipid transport and metabolism' category (Fig. 5). The high abundance of butyryl-CoA dehydrogenases, also reported in healthy female identical twins (Verberkmoes *et al*., 2009), is noteworthy. Butyrate producers that colonize the human gut are strict *Firmicutes* (anaerobes), with the two most abundant groups being related to *Eubacterium rectale*/*Roseburia* spp. and to *Faecalibacterium prausnitzii*. Accordingly, all proteins possibly involved in butyrate metabolism herein identified in the 'Obese' gut most likely binned to *Firmicutes* members, such as *Eubacterium*. The activation of butyrate production in the 'Obese' microbiome, which may be linked to dietary polysaccharide fermentation, may have practical beneficial implications, as butyrate is thought to play an important role in maintaining colonic health in humans as the major source of energy to the colonic mucosa (Louis and Flint, 2009).

In both microbiomes, there were over-representations of proteins belonging to the COG classified as 'Carbohydrate Transport and Metabolism', with 32 (for 'Lean') and 48 (for 'Obese') proteins identified (Fig. 5 and Table S1). In both cases, abundances and high expression levels of proteins involved in xylose, arabinose, glucose, galactose and mannose metabolism were observed, in agreement with the fact that xylan-, pectin- and arabinose-containing carbohydrate structures are among those commonly consumed by humans (Gill *et al*., 2006). Among these, 6 and 12 distinct glycosyl hydrolases (and a number of enzymes acting against the produced oligosaccharides) were identified in 'Lean' and 'Obese' metaproteomes respectively (Table S1). All 'Lean' proteins binned to *Prevotella* members of the *Bacteroides* phylum, whereas among the proteins identified in the 'Obese' (see Table S1), the majority (~ 90%) binned to *Firmicutes* members, especially to *Ruminococcus*. This shift is also observed when examining the proteomic signatures in relation to the uptake and hydrolysis of cellobiose. Two cellobiose phosphorylases were identified, including one 'Lean' specific and one that was more abundant in the 'Obese' gut [$\log_2$('Obese'/'Lean') of 1.95] (Fig. 6). Whereas the first enzyme binned to a *Prevotella* member of the *Bacteroides* phylum, the second binned to a *Faecalibacterium* member of the *Firmicutes* phylum. Taken together, the results provided evidences that suggest that different metabolically active bacterial members in 'Lean' and 'Obese' microbiomes mediate similar functions, in agreement with previous investigations that associated cellulolytic activity to different bacterial groups (Turnbaugh *et al*., 2010). It is also worth mentioning the identification of a putative GT1-glycogen phosphorylase and an HPr kinase/phosphorylase that were more abundant in the 'Obese' gut [$\log_2$('Obese'/'Lean') from ~ 5.0 to 7.0]. These enzymes, together with other oligosaccharide phosphorylases, catalyse the breakdown of oligosaccharides [i.e. glycogen (from host tissue), starch or maltodextrins]

into glucose-1-phosphate units and are thus important enzymes in carbohydrate metabolism. Additionally, several COGs responsible for fucose (most likely originating from host mucus) utilization are enriched in the human gut microbiome (Gill *et al*., 2006). This is also supported by this investigation: 4 distinct L-fucose isomerases were identified in both 'Lean' and 'Obese' guts (2 hits each) (Table S1), thus providing experimental evidence that fucose is an attractive and accessible source of energy for members of the microbiome – in particular, *Bacteroides* in the 'Lean' gut and *Firmicutes* in the 'Obese' gut (as demonstrated by binning analysis of the corresponding protein sequences) (Table S1).

Presence of other proteins in lean and obese guts

Additionally, a number of proteins were identified in both 'Lean' and 'Obese' guts (Table S1), including glutamate dehydrogenases (12), aminotransferases such as ornithine/acetylornithine (9) and phosphoserine aminotransferases (3), lactate dehydrogenase (1) and phosphoenolpyruvate carboxykinases (4), most likely playing roles as electron sinks and in gluconeogenesis, pyruvate-formate lyases (4), which are spread in anaerobes and convert pyruvate to acetyl-CoA and formate, and GroEL chaperonins (9) and a NifU protein with chaperone function (1). This set of proteins has also been reported as being expressed in the guts of healthy adults (Kolmeder *et al*., 2012), which corroborates their relevance in human gut microbiomes in both 'Lean' and 'Obese' individuals.

Conclusions

The human distal gut is a metabolically active organ that consumes a considerable amount of energy; when the food supply, and thereby the energy supply, is modified, structural and functional changes occur rapidly with consequences that have been suggested to modulate pathways controlling the overall metabolism. Global changes in microbial diversity have been visualized by a number of approaches, but a complete understanding of this process from a functional point of view remains to be achieved. Here, cultivation-independent metagenomic, shotgun metaproteomic and 16S rDNA assessments were employed to deduce correlations between metabolic potential and phylogenetic, genomic and proteomic blueprints of the microbial communities active in faecal samples from one 'Lean' and one 'Obese' adolescent. Our metagenomic approach revealed that the putative origins (at the phylum level) of genes encoded by the recovered DNA fragments were in accordance with the observed 16S rDNA gene diversity and reflected the expected range of organisms that thrive in gut environments. A few ecological and metabolic clues can be correlated to differences in bacterial lineages defined by full-length 16S rDNA sequences and proteins identified by shotgun proteomics. Thus, there was a drastic change in total and active microbial community. For active bacteria, *Firmicutes* and *Bacteroidetes* represent both the most active phyla in 'Obese' gut, whereas a clear shift towards active *Bacteroidetes* was observed in 'Lean'. Furthermore, our data suggest that the 'Obese' gut contains a subset of specialized primary *Firmicutes* degraders that use pili and fibronectin domain-containing proteins to adhere tightly to substrates and/or host tissues. Accordingly, recent *in vitro* studies showed that the colonization of starch by human microbiota was mainly performed by *Eubacterium rectale* and *Ruminococcus bromii* (Leitch *et al*., 2007), whose clade members account for 20% and 4% in the 'Obese' and 'Lean' libraries respectively. Moreover, many of these bacteria are hydrogen producers that, in association with an acetogen, use hydrogen to produce acetate (Chassard and Bernalier-Donadille, 2006). The production of fermentation products from released oligosaccharides is distributed among many clostridial clusters; however, an abundant group of butyrate producers belongs to *Roseburia* and *Faecalibacterium* genera (acetate converters) and to *Eubacterium halii* and *Anaerostipes caccae* (acetate and lactate converters) that are also hydrogen producers (Flint *et al*., 2007). Although no differences in the abundance of butyrate producers were found between our samples at the level of 16S rDNA, the net functional activation of butyrate metabolism and acetogenesis (OPU 41, *Clostridium ruminantium* clade) was only observed in the 'Obese' library. These data suggest that the 'Obese' faecal ecosystem provides a more efficient cross-feeding interaction between consortia members for dietary polysaccharide fermentation. Finally, this study should be of special interest in defining the roles of individual functional microbial pathways in human metabolism, as differences in vitamin biosynthetic pathways were observed on several levels in 'Lean' and 'Obese' gut microbiomes. Furthermore, this study provides experimental evidence that the functional roles of bacterial members can be interchanged in 'Lean' and 'Obese' gut microbiomes to produce stable functional communities depending on the food (or energy) supply. Accordingly, gut microbial communities may display a higher metabolic plasticity than previously assumed. The study herein reported for two individuals highlights the importance of integrated metagenomic and metaproteomic approaches, towards unravelling the dynamics and mechanisms underlying the response of intestinal microbiota to obesity. This is of special significance as we know that most research relied on indirect evidence from DNA-based approaches that fail to provide information on actual protein synthesis. Further studies investigating different (un)related individuals and alternative integrated meta-omic approaches, such as

metatranscriptomic and metametabolomic, should be required to ascertain the metabolic consequences in relation to obesity.

Experimental procedures

Subjects, anthropometric measures and biochemical analyses

Subjects for the study were selected according to their body mass index [BMI; weight (kg) height $(m)^{-2}$] from the panel of adolescents collected during the course of the EVASION study (Nadal *et al.*, 2009; Santacruz *et al.*, 2009). The study was designed to develop a multidisciplinary obesity treatment programme that was assessed and controlled by Paediatric services in Spain. The study was approved by the local ethics committees and was conducted in accordance with the ethical rules of the Helsinki Declaration (Hong Kong revision, September 1989), following the EEC Good Clinical Practice guidelines (document 111/3976/88 of July 1990) and current Spanish law, which regulates clinical research in Humans (Royal Decree 561/1993 regarding clinical trials). Written informed consent was obtained from all participants and their parents. The subjects had not been subjected to any antibiotic treatments, dietary interventions, or specific diets for at least 1 month prior to sampling and considered themselves healthy. Overweight (including obesity) in participants was determined according to the International Obesity Task Force criteria defined by Cole and colleagues (2000). Fasting plasma glucose and insulin, uric acid, total cholesterol, triglycerides, and HDL and LDL cholesterol levels were measured in obese subjects as described in Nadal and colleagues (2009). Individuals designated 'Lean' (lean adolescent) and 'Obese' (obese subject) were the focus of our study. The anthropometric and clinical characteristics of both adolescents are as follows: sex ('Lean': female; 'Obese': male), Age ('Lean': 15 years; 'Obese': 15 years), weight ('Lean': 63.1 kg; 'Obese': 102.7 kg), Height ('Lean': 165 cm; 'Obese': 171 cm), BMI (kg m^{-2}) ('Lean': 23.18; 'Obese': 35.12), fasting glucose (mg dl^{-1}) ('Lean': n.d. – not determined; 'Obese': 80), fasting insulin (pg ml^{-1}) ('Lean': n.d.; 'Obese': 517), uric acid (mg dl^{-1}) ('Lean': n.d.; 'Obese': 8.6), total cholesterol (mg dl^{-1}) ('Lean': n.d.; 'Obese': 198), HDL cholesterol (mg dl^{-1}) ('Lean': n.d.; 'Obese': 46), LDL cholesterol (mg dl^{-1}) ('Lean': n.d.; 'Obese': 118), and tryglicerides (mg dl^{-1}) ('Lean': n.d.; 'Obese': 172).

Faecal sample collection and preparation for protein extraction

Fresh faecal samples were collected, immediately frozen at –20°C and stored at –80°C until they were analysed. Faeces were thawed, diluted in 0.05% L-cysteine PBS, pH 7.5, solution and vigorously mixed at 37°C for 10 min. After low-speed centrifugation (2000 *g*, 15 min), the supernatant was collected and centrifuged at 16 000 *g* for 5 min to precipitate faecal bacteria. Faecal bacteria were further washed with 0.05% L-cysteine PBS, pH 7.5, solution and then centrifuged at 16 000 *g* for 5 min. The resultant pellet of bacteria was lysed in BugBuster Protein Extraction Reagent (Novagen, Darmstadt, Germany) for 30 min at room temperature. Faecal bacteria were treated with Lysonase Bioprocessing Reagent (Novagen) for 1 h at 4°C, with further disruption by sonication for 2.5 min on ice. The extract was then centrifuged for 10 min at 12 000 *g* to separate cell debris and intact cells. The supernatant was carefully aspirated (to avoid disturbing the pellet) and transferred to a new tube, and proteins were precipitated by methanol-chloroform precipitation and dry pellet stored at –80°C until they were analysed.

DNA extraction

DNA was extracted from faecal bacteria as described by Zoetendal and colleagues (1998). Faecal samples of 100 mg each were homogenized in 1 ml of TN150 buffer (10 mM Tris-HCl pH 8.0 and 150 mM NaCl) with 0.3 g of glass beads (0.1 mm diameter) and 150 ml of buffered phenol. Bacteria were disrupted with a mini bead beater at 5000 r.p.m. for 1 min at 4°C (Biospec Products, Bartlesville, OK, USA). After centrifugation, genomic DNA was purified from the supernatant with a GNOME DNA isolation kit (MP Biomedicals, USA), quality-checked by agarose electrophoresis and spectrophotometrically quantified (NanoDrop, Fisher Scientific, Spain).

DNA sequencing, assembly, gene prediction, annotation and gene taxonomic assignments

Sequencing was performed with a Roche 454 GS FLX Ti sequencer (454 Life Sciences, Branford, CT, USA) at Lifesequencing S.L. (Valencia, Spain), with one picotiterplate each. Assembly was performed with Roche's Newbler assembler v. 2.5.3 with default parameters. For the identification of 16S rDNA gene partial sequences (with a length $>$ 600 nucleotides), raw sequences were used.

Potential protein-coding genes were identified by using MetaGene (Noguchi *et al.*, 2006). Additionally, contigs without ORF predictions by MetaGene were translated into artificial reading frames and were then BLAST-searched against the NCBI-nr database for similar sequences. Artificial translations with similarities were further processed in the same way as the predicted ORFs from MetaGene. Transfer RNA genes were identified using tRNAScan-SE (Lowe and Eddy, 1997), and ribosomal RNA genes were identified with Meta-RNA 1.0 (Huang *et al.*, 2009). Annotation was performed with GenDB, version 2.2 (Meyer *et al.*, 2003), supplemented by the tool JCoast, version 1.7 (Richter *et al.*, 2008). For each predicted ORF, observations were collected from similarity searches against the sequence databases NCBI-nr, Swiss-Prot, KEGG and genomesDB (Richter *et al.*, 2008) and the Pfam (Bateman *et al.*, 2004) and InterPro (Mulder *et al.*, 2005) protein family databases. SignalP was used for signal peptide predictions (Bendtsen *et al.*, 2004), and TMHMM was used for transmembrane helix-analysis (Krogh *et al.*, 2001). Predicted protein coding sequences were automatically annotated by the in-house software MicHanThi (Quast, 2006). MicHanThi predicts gene functions based on similarity searches using the NCBI-nr (including Swiss-Prot) and InterPro databases. The annotation of proteins highlighted within the scope of this study was subject to manual inspection. For all observations regard-

ing putative protein functions, an *E*-value cut-off of 10^{-4} was standard. To evaluate the taxonomic consistency of the conserved ORFs, all ORFs were tested by BLAST analysis for the taxonomic distribution of best hits against a local genome database (genomesDB; M. Richter, unpubl. data). Only hits with an *e*-value below e^{-05} were considered significant. The local genome database (genomesDB) provides a computationally well-defined environment of ~ 3000 published whole/draft genome sequences of bacterial and archaeal origin, with all ORFs of each genome carrying a unique ID. To allow genome comparisons between specific user-defined groups, all ORFs are assigned to the respective organism and metabolic group. In contrast to the general purpose database NCBI-nr, which contains every sequence ever submitted, the focus of genomesDB is the association of every protein to their taxonomic affiliation in a refined environment.

To identify potential metabolic pathways, genes were searched for similarity against the KEGG database. A match was counted if the similarity search resulted in an *E*-value below $1e^{-5}$. All occurring KO (KEGG Orthology) numbers were mapped against KEGG pathway functional hierarchies and were statistically analysed. All predicted ORFs were also searched for similarity against the cluster of orthologous groups (COG) database (Tatusov *et al*., 2003). A match was counted if the similarity search resulted in an *E*-value below $1e^{-5}$.

PCR amplification of 16S rDNA

Clone libraries were generated from total genomic DNA extracted from faecal samples. Amplified PCR products of 16S rDNA obtained with universal primers 27f and 1492r were purified and cloned using the pGEM-T cloning kit (Promega, USA) followed by electroporation into *Escherichia coli* XL1 Blue cells; the resulting clones were selected according to the manufacturer's recommendations. Initially, the cloned inserts were amplified by PCR with vector primers, digested with AluI, HaeIII and MspI restriction endonucleases (New England BioLabs, Ipswich, MA) and separated by electrophoresis in 3% (w/v) SeaKem LE agarose (Cambrex, Wiesbaden, Germany). The RFLP analysis showed that MspI restriction was the best classifier for our samples; therefore, it was used to group the selected clones. Clones exhibiting the same RFLP patterns were grouped, and representatives from each group were selected for sequencing. Full-length 16S rDNA sequences of the selected clones were determined from plasmid DNA preparations obtained using a Nucleospin plasmid kit (Macherey-Nagel, Germany) and a BigDye terminator cycle sequencing v3.1 kit (Applied Biosystems, Foster City, USA) according to the manufacturers' instructions.

Phylogenetic analysis of 16S rDNA

Full-length 16S rDNA sequences were compared with reference sequences from the EMBL Nucleotide Sequence Database (Cochrane *et al*., 2006) using the FASTA algorithm (Pearson and Lipman, 1988) and were subsequently aligned with reference sequences using CLUSTALW. Sequences were checked for chimeric artefacts using the CHECK_CHIMERA programme (Cole *et al*., 2003). Phylogenetic trees were constructed with the MEGA programme (Tamura *et al*., 2011) using the neighbour-joining distance method with reference 16S rDNA sequences and full-length sequences obtained in this study.

1D-gel electrophoresis, mass spectrometry and data analysis

Technical replicates of samples (herein referred as 'Obese'-3, 'Obese'-5, 'Lean'-3 and 'Lean'-5) were thawed on ice and resuspended in 500 µl 20 mM Tris buffer (pH 7.5; 0.1 mg ml^{-1} $MgCl_2$; 1 mM PMSF). Protein concentrations were determined using the Bradford assay. For 1-DE analysis, two replicates of 50 µg of protein per sample [technical replicates denoted by (a) or (b)] were precipitated with five volumes of ice-cold acetone and were separated using a 12% acrylamide separating gel and the Laemmli buffer system. After electrophoresis, protein bands were stained with colloidal Coomassie Brilliant Blue G-250 (Roth, Kassel, Germany). Entire protein lanes were individually cut into 7 bands, and subsequent in-gel tryptic digestion was performed. Tryptic digests of each band were desalted using the ZipTip C18 before MS analysis.

Peptides were analysed by UPLC-LTQ Orbitrap-MS/MS as described in Bastida and colleagues (2010). The peptides were eluted for 77 min over a 2–60% acetonitrile + 0.1% formic acid gradient. Continuous scanning of eluted peptide ions was performed between 300–1600 *m/z*, automatically switching to MS/MS CID mode on ions exceeding an intensity of 2000. Raw data were further processed with MaxQuant™ (version 1.2.18) (Cox and Mann, 2008). Raw data obtained from peptide samples originating from the same lane on the 1DE gel were combined. The databases searched were the provided metagenomic data from the 'Obese' and 'Lean' samples and an artificial metagenome based on a set of bacterial and archaeal genomes listed by Rooijers and colleagues (2011). Settings for MaxQuant: Peptide modifications given were methionine oxidation as variable and cysteine carbamidomethylation as fixed. Further settings were first search p.p.m. of 20, main search p.p.m. of 6, maximum number of modifications per peptide 5, maximum missed cleavages 2 and a maximum charge for the peptide of 7. Parameters for the identification were a minimum peptide length of 5 amino acids and 1% false discovery rates for peptides, proteins and levels of modification sites. A minimum of 2 unique peptides was required for protein identification. Apart from unmodified peptides, only peptides with oxidized methionine and carbamidomethylized cysteine were used for quantification. Only unique or razor peptides were chosen for use in quantification. Miscellaneous settings switched on were re-quantified; low-scoring versions of identified peptides, match between runs (time window of 2 min), label-free quantification and second peptides were kept.

To analyse the data, intensities attributed to each identified protein were divided by the number of peptides assigned to the protein. Then, normalization was performed by dividing these corrected intensity values by the median of all corrected intensities from the same sample. The ratio of the normalized intensities was calculated for each protein by

dividing the mean of the normalized intensities from the obese samples ('Obese'-3a; 'Obese'-3b; 'Obese'-5a; 'Obese'-5b) by the mean of the normalized intensities from the lean samples ('Lean'-3a; 'Lean'-3b; 'Lean'-5a; 'Lean'-5b). Ratio values were then calculated by taking the logarithms of the ratios to the base of 2. For upregulation of protein expression in the 'Obese' environment, a threshold of at least 1.5 in the ratio value was set; for downregulation, a maximum value of −1.5 was set.

MIxS submission

Consistent contextual data acquisition for MIxS-compliant submission using the environmental package 'host-associated' was performed with JCoast v1.7.

Deposition of sequence data

Project has been registered as umbrella BioProject at NCBI with the ID PRJNA88107 and PRJNA88153 for lean and obese human gut metagenomes respectively. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession numbers AJWY00000000 and AJWZ00000000, in the same order. The version described in this paper is the first version, AJWY01000000 and AJWZ01000000.

Acknowledgements

The investigation was funded by the Spanish Ministry of Economy and Competitiveness (former MICINN) and German BMBF, within the ERA NET PathoGenoMics2 call, grant number 0315441A. This work was further funded by the grants BFU2008-04501-E/BMC, CSD2007-00005, AGL2006-11697/ALI, IFB Adiposity Diseases Programme at the University Leipzig and by FEDER funds.

References

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., *et al*. (2011) Enterotypes of the human gut microbiome. *Nature* **473:** 174–180.

Bastida, F., Rosell, M., Franchini, A.G., Seifert, J., Finsterbusch, S., Jehmlich, N., *et al*. (2010) Elucidating MTBE degradation in a mixed consortium using a multidisciplinary approach. *FEMS Microbiol Ecol* **73:** 370–384.

Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., *et al*. (2004) The Pfam protein families database. *Nucleic Acids Res* **32:** D138–D141.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340:** 783–795.

Bobik, T.A., Havemann, G.D., Busch, R.J., Williams, D.S., and Aldrich, H.C. (1999) The propanediol utilization (pdu) operon of *Salmonella enterica* serovar *Typhimurium* LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B(12)-dependent 1, 2-propanediol degradation. *J Bacteriol* **181:** 5967–5975.

Booijink, C.C., Boekhorst, J., Zoetendal, E.G., Smidt, H., Kleerebezem, M., and de Vos, W.M. (2010) Metatranscriptome analysis of the human fecal microbiota reveals subject-specific expression profiles, with genes encoding proteins involved in carbohydrate metabolism being dominantly expressed. *Appl Environ Microbiol* **76:** 5533–5540.

Cani, P.D., and Delzenne, N.M. (2011) The gut microbiome as therapeutic target. *Pharmacol Ther* **130:** 202–212.

Cantarel, B.L., Erickson, A.R., VerBerkmoes, N.C., Erickson, B.K., Carey, P.A., Pan, C., *et al*. (2011) Strategies for metagenomic-guided whole-community proteomics of complex microbial environments. *PLoS ONE* **6:** e27173.

Chassard, C., and Bernalier-Donadille, A. (2006) H2 and acetate transfers during xylan fermentation between a butyrate-producing xylanolytic species and hydrogenotrophic micro-organisms from the human gut. *FEMS Microbiol Lett* **254:** 116–122.

Cochrane, G., Aldebert, P., Althorpe, N., Andersson, M., Baker, W., Baldwin, A., *et al*. (2006) EMBL nucleotide sequence database: developments in 2005. *Nucleic Acids Res* **34:** D10–D15.

Cole, J.R., Chai, B., Marsh, T.L., Farris, R.J., Wang, Q., Kulam, S.A., *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.

Cole, T.J., Bellizzi, M.C., Flegal, K.M., and Dietz, W.H. (2000) Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* **320:** 1240–1243.

Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26:** 1367–1372.

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

Elli, M., Colombo, O., and Tagliabue, A. (2010) A common core microbiota between obese individuals and their lean relatives? Evaluation of the predisposition to obesity on the basis of the fecal microflora profile. *Med Hypotheses* **75:** 350–352.

Ellrott, K., Jaroszewski, L., Li, W., Wooley, J.C., and Godzik, A. (2010) Expansion of the protein repertoire in newly explored environments: human gut microbiome specific protein families. *PLoS Comput Biol* **6:** e1000798.

Flint, H.J., Duncan, S.H., Scott, K.P., and Louis, P. (2007) Interactions and competition within the microbial community of the human colon: links between diet and health. *Environ Microbiol* **9:** 1101–1111.

Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., *et al*. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* **312:** 1355–1359.

Gosalbes, M.J., Durbán, A., Pignatelli, M., *et al*. (2011) Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS ONE* **6:** e17447.

Greenblum, S., Turnbaugh, P.J., and Borenstein, E. (2012) Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc Natl Acad Sci USA* **109:** 594–599.

Hattori, M., and Taylor, T.D. (2009) The human intestinal microbiome: a new frontier of human biology. *DNA Res* **16:** 1–12.

Hildebrandt, M.A., Hoffmann, C., Sherrill-Mix, S.A., Keilbaugh, S.A., Hamady, M., Chen, Y.Y., *et al*. (2009) High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137:** 1716–1724.

Hooper, L.V., and Gordon, J.I. (2001) Commensal host-bacterial relationships in the gut. *Science* **292:** 1115–1118.

Huang, Y., Gilna, P., and Li, W. (2009) Annotation of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* **25:** *1338–1340*.

Jee, J.G., Ikegami, T., Hashimoto, M., Kawabata, T., Ikeguchi, M., Watanabe, T., and Shirakawa, M. (2002) Solution structure of the fibronectin type III domain from *Bacillus circulans* WL-12 chitinase A1. *J Biol Chem* **277:** 1388–1397.

Jernberg, C., Löfmark, S., Edlund, C., and Jansson, J.K. (2010) Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology* **156:** 3216–3223.

Klaassens, E.S., de Vos, W.M., and Vaughan, E.E. (2007) Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. *Appl Environ Microbiol* **73:** 1388–1392.

Kleerebezem, M., Crielaard, W., and Tommassen, J. (1996) Involvement of stress protein PspA (phage shock protein A) of *Escherichia coli* in maintenance of the proton motive force under stress conditions. *EMBO J* **15:** 162–171.

Kobayashi, R., Suzuki, T., and Yoshida, M. (2007) *Escherichia coli* phage-shock protein A (PspA) binds to membrane phospholipids and repairs proton leakage of the damaged membranes. *Mol Microbiol* **66:** 100–109.

Kolmeder, C.A., de Been, M., Nikkilä, J., Ritamo, I., Mättö, J., Valmu, L., *et al*. (2012) Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PLoS ONE* **7:** e29913.

Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305:** 567–580.

Krooneman, J., Faber, F., Alderkamp, A.C., Elferink, S.J., Driehuis, F., Cleenwerck, I., *et al*. (2002) *Lactobacillus diolivorans* sp. nov., a 1,2-propanediol-degrading bacterium isolated from aerobically stable maize silage. *Int J Syst Evol Microbiol* **52:** 639–646.

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.

Leitch, E.C., Walker, A.W., Duncan, S.H., Holtrop, G., and Flint, H.J. (2007) Selective colonization of insoluble substrates by human faecal bacteria. *Environ Microbiol* **9:** 667–679.

Ley, R.E. (2010) Obesity and the human microbiome. *Curr Opin Gastroenterol* **26:** 5–11.

Lobo, S.A., Brindley, A.A., Romão, C.V., Leech, H.K., Warren, M.J., and Saraiva, L.M. (2008) Two distinct roles for two functional cobaltochelatases (CbiK) in *Desulfovibrio vulgaris* hildenborough. *Biochemistry* **47:** 5851–5857.

Louis, P., and Flint, H.J. (2009) Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol Lett* **294:** 1–8.

Lowe, T.M., and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25:** 955–964.

Marchesi, J.R. (2010) Prokaryotic and eukaryotic diversity of the human gut. *Adv Appl Microbiol* **72:** 43–62.

Marchesi, J.R. (2011) Human distal gut microbiome. *Environ Microbiol* **13:** 3088–3102.

Meyer, F., Goesmann, A., McHardy, A.C., Bartels, D., Bekel, T., Clausen, J., *et al*. (2003) GenDB–an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31:** 2187–2195.

Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., *et al*. (2005) InterPro, progress and status in 2005. *Nucleic Acids Res* **33:** D201–D205.

Musso, G., Gambino, R., and Cassader, M. (2011) Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. *Annu Rev Med* **62:** 361–380.

Nadal, I., Santacruz, A., Marcos, A., Warnberg, J., Garagorri, M., Moreno, L.A., *et al*. (2009) Shifts in clostridia, bacteroides and immunoglobulin-coating fecal bacteria associated with weight loss in obese adolescents. *Int J Obes (Lond)* **33:** 758–767.

Nicholson, J.K., Holmes, E., and Wilson, I.D. (2005) Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol* **3:** 431–438.

Noguchi, H., Park, J., and Takagi, T. (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* **34:** 5623–5630.

O'Connell Motherway, M., Zomer, A., Leahy, S.C., Reunanen, J., Bottacini, F., Claesson, M.J., *et al*. (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 USA* **108:** 11217–11222.

O'Hara, A.M., and Shanahan, F. (2006) The gut flora as a forgotten organ. *EMBO Rep* **7:** 688–693.

Pearson, W.R., and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85:** 2444–2448.

Peterson, D.A., Frank, D.N., Pace, N.R., and Gordon, J.I. (2008) Metagenomic approaches for defining the pathogenesis of inflammatory bowel diseases. *Cell Host Microbe* **3:** 417–427.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., *et al*. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464:** 59–65.

Quast, C. (2006) MicHanThi – design and implementation of a system for the prediction of gene functions in genome annotation projects. Master Thesis. Bremen, Germany: University of Bremen.

Richter, M., Lombardot, T., Kostadinov, I., Kottmann, R., Duhaime, M.B., Peplies, J., and Glöckner, F.O. (2008) JCoast – a biologist-centric software tool for data mining and comparison of prokaryotic metagenomes. *BMC Bioinformatics* **9:** 177.

Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. (2003) Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. *J Biol Chem* **278:** 41148–41159.

Rooijers, K., Kolmeder, C., Juste, C., Doré, J., de Been, M., Boeren, S., *et al*. (2011) An iterative workflow for mining the human intestinal metaproteome. *BMC Genomics* **12:** 6.

Santacruz, A., Marcos, A., Wärnberg, J., Martí, A., Martin-Matillas, M., Campoy, C., *et al*. (2009) Interplay between weight loss and gut microbiota composition in overweight adolescents. *Obesity* **17:** 1906–1915.

Schwiertz, A., Taras, D., Schäfer, K., Beijer, S., Bos, N.A., Donus, C., and Hardt, P.D. (2010) Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* **18:** 190–195.

Sleator, R.D., Shortall, C., and Hill, C. (2008) Metagenomics. *Lett Appl Microbiol* **47:** 361–366.

Swinburn, B.A., Sacks, G., Hall, K.D., McPherson, K., Finegood, D.T., Moodie, M.L., and Gortmaker, S.L. (2011) The global obesity pandemic: shaped by global drivers and local environments. *Lancet* **378:** 804–814.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28:** 2731–2739.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., *et al*. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4:** 41.

Tschöp, M.H., Hugenholtz, P., and Karp, C.L. (2009) Getting to the core of the gut microbiome. *Nat Biotechnol* **27:** 344–346.

Turnbaugh, P.J., and Gordon, J.I. (2009) The core gut microbiome, energy balance and obesity. *J Physiol* **587:** 4153–4158.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444:** 1027–1031.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., and Ley, R.E. (2009) A core gut microbiome in obese and lean twins. *Nature* **457:** 480–484.

Turnbaugh, P.J., Quince, C., Faith, J.J., *et al*. (2010) Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc Natl Acad Sci USA* **107:** 7503–7508.

Verberkmoes, N.C., Russell, A.L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., *et al*. (2009) Shotgun metaproteomics of the human distal gut microbiota. *ISME J* **3:** 179–189.

Xu, J., Chiang, H.C., Bjursell, M.K., Gordon, J.I., Xu, J., Chiang, H.C., *et al*. (2004) Message from a human gut symbiont: sensitivity is a prerequisite for sharing. *Trends Microbiol* **12:** 21–28.

Zhang, H., DiBaise, J.K., Zuccolo, A., Kudrna, D., Braidotti, M., Yu, Y., *et al*. (2009) Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci USA* **106:** 2365–2370.

Zoetendal, E.G., Akkermans, A.D., and De Vos, W.M. (1998) Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. *Appl Environ Microbiol* **64:** 3854–3859.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1. Proteins of 'Lean' and 'Obese' gut communities identified and quantified by metaproteomic approach. Protein annotations, tentative phylogenetic affiliations, COG assignations and relative abundances are specifically shown. Panel (A) 'Proteins identified-quantified' provide the list of proteins identified using Rooijers and colleagues' (2011) genome list database search and the pyrosequences provided in this study; Panel (B) 'Proteins by COG distribution' provides the complete list of proteins distributed per COG and the relative abundance of each in 'Lean' and 'Obese' gut metaproteomes.