Revealing the Bacterial Butyrate Synthesis Pathways by Analyzing (Meta)genomic Data

Marius Vital,[a] Adina Chuang Howe,[a,b] James M. Tiedje[a]

Center for Microbial Ecology, Michigan State University, East Lansing, Michigan, USA[a]; Argonne National Laboratory, Lemont, Illinois, USA[b]

ABSTRACT **Butyrate-producing bacteria have recently gained attention, since they are important for a healthy colon and when altered contribute to emerging diseases, such as ulcerative colitis and type II diabetes. This guild is polyphyletic and cannot be accurately detected by 16S rRNA gene sequencing. Consequently, approaches targeting the terminal genes of the main butyrate-producing pathway have been developed. However, since additional pathways exist and alternative, newly recognized enzymes catalyzing the terminal reaction have been described, previous investigations are often incomplete. We undertook a broad analysis of butyrate-producing pathways and individual genes by screening 3,184 sequenced bacterial genomes from the Integrated Microbial Genome database. Genomes of 225 bacteria with a potential to produce butyrate were identified, including many previously unknown candidates. The majority of candidates belong to distinct families within the *Firmicutes*, but members of nine other phyla, especially from *Actinobacteria*, *Bacteroidetes*, *Fusobacteria*, *Proteobacteria*, *Spirochaetes*, and *Thermotogae*, were also identified as potential butyrate producers. The established gene catalogue (3,055 entries) was used to screen for butyrate synthesis pathways in 15 metagenomes derived from stool samples of healthy individuals provided by the HMP (Human Microbiome Project) consortium. A high percentage of total genomes exhibited a butyrate-producing pathway (mean, 19.1%; range, 3.2% to 39.4%), where the acetyl-coenzyme A (CoA) pathway was the most prevalent (mean, 79.7% of all pathways), followed by the lysine pathway (mean, 11.2%). Diversity analysis for the acetyl-CoA pathway showed that the same few firmicute groups associated with several *Lachnospiraceae* and *Ruminococcaceae* were dominating in most individuals, whereas the other pathways were associated primarily with *Bacteroidetes*.**

IMPORTANCE **Microbiome research has revealed new, important roles of our gut microbiota for maintaining health, but an understanding of effects of specific microbial functions on the host is in its infancy, partly because in-depth functional microbial analyses are rare and publicly available databases are often incomplete/misannotated. In this study, we focused on production of butyrate, the main energy source for colonocytes, which plays a critical role in health and disease. We have provided a complete database of genes from major known butyrate-producing pathways, using in-depth genomic analysis of publicly available genomes, filling an important gap to accurately assess the butyrate-producing potential of complex microbial communities from "-omics"-derived data. Furthermore, a reference data set containing the abundance and diversity of butyrate synthesis pathways from the healthy gut microbiota was established through a metagenomics-based assessment. This study will help in understanding the role of butyrate producers in health and disease and may assist the development of treatments for functional dysbiosis.**

Received 5 February 2014 **Accepted** 7 March 2014 **Published** 22 April 2014

Citation Vital M, Howe AC, Tiedje JM. 2014. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. mBio 5(2):e00889-14. doi:10.1128/mBio.00889-14.

Editor Mary Ann Moran, University of Georgia

Address correspondence to James M. Tiedje, tiedjej@msu.edu.

Butyrate-producing bacteria are widespread and can be found in many environments (1) but especially in host-associated sites, including the rumen (2), the mouth (3), and the large intestine (4). Recently, butyrate gained attention, because of its proposed key role in maintaining gut homeostasis and epithelial integrity, since it serves as the main energy source for colonocytes, directly influences host gene expression by inhibiting histone deacetylases, and interferes with proinflammatory signals, such as NF-κB (5, 6). A breakdown of epithelial integrity is associated with emerging diseases such as inflammatory bowel diseases and type II diabetes (7, 8), and butyrate-producing members specifically are reduced in such patients (9, 10).

Butyrate producers form a functional cohort rather than a monophyletic group, and members of *Lachnospiraceae* and *Ruminococcaceae* have received the most attention because they are very abundant in the human colon, comprising 10 to 20% of the total bacteria. Butyrate is synthesized via pyruvate and acetyl-coenzyme A (CoA), mostly by the breakdown of complex polysaccharides (e.g., starch and xylan) that escape digestion in the upper gastrointestinal tract and reach the colon (11). Alternative substrates, particularly those derived from cross-feeding with other primary degraders and lactate-synthesizing bacteria, are described as well (12). Acetyl-CoA is then converted to the intermediate butyryl-CoA in a four-step pathway closely related to the β-oxidation of fatty acids in prokaryotes and eukaryotes (13, 14). It is postulated that butyrate producers can conserve energy dur-

ing the conversion from crotonyl-CoA to butyryl-CoA, which creates a proton motive force via ferredoxin reduction by the butyryl-CoA dehydrogenase electron-transferring flavoprotein complex (15). The final step from butyryl-CoA to butyrate is either catalyzed by butyryl-CoA:acetate CoA transferase (encoded by *but*) or butyrate kinase (encoded by *buk*; after phosphorylation of butyryl-CoA). Typically, these two genes are used as biomarkers for the identification/detection of butyrate-producing communities (16, 17). However, direct functional predictions based on gene homology alone can commonly result in misannotations if genes with distinct function share regions of high similarity, as specifically described for both *but* and *buk* (17). Furthermore, CoA transferases show activity with several different substrate combinations *in vitro* (18), and alternative terminal CoA transferases were proposed for this pathway (19). Targeting the whole pathway for functional predictions is hence a robust way to circumvent difficulties associated with the analysis based on specific genes only. Additionally, there are other known butyrate-producing pathways, namely, the lysine, glutarate, and 4-aminobutyrate pathways, where amino acids serve as major substrates. These pathways are found in *Firmicutes* as well as other phyla, such as *Fusobacteria* and *Bacteroidetes* (20–22), but are traditionally neglected as potential butyrate-producing routes in enteric environments.

The availability of complete databases, including diverse candidates and pathways, is essential to investigate specific microbial functions in complex microbial communities, to assess their effects on the host, and to ultimately develop treatment strategies for functional dysbiosis. The aim of this study was to screen available genomes, many from the Human Microbiome Project (HMP) framework, for potential butyrate producers and to characterize their phylogeny, gene arrangements, and gene phylogeny. The resulting gene catalogue was then used to screen for butyrate synthesis pathways in metagenomic HMP data to reveal this important functional community within the healthy microbiota.

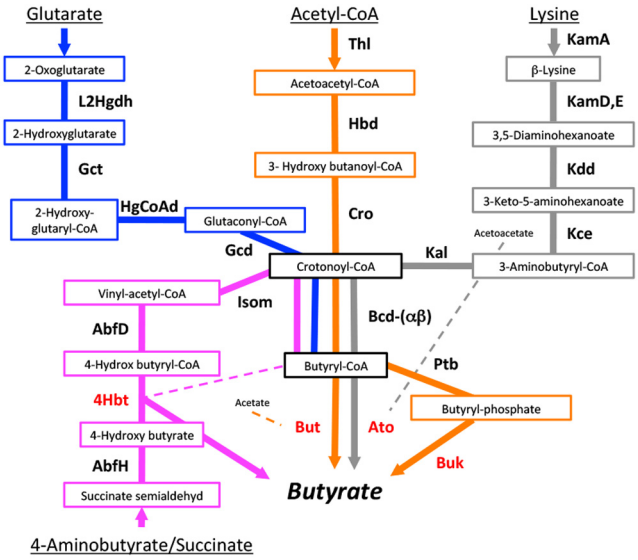

FIG 1 Four different pathways for butyrate synthesis and corresponding genes (protein names) are displayed. Major substrates are shown. Terminal genes are highlighted in red. L2Hgdh, 2-hydroxyglutarate dehydrogenase; Gct, glutaconate CoA transferase (α, β subunits); HgCoAd, 2-hydroxy-glutaryl-CoA dehydrogenase (α, β, γ subunits); Gcd, glutaconyl-CoA decarboxylase (α, β subunits); Thl, thiolase; *hbd*, β-hydroxybutyryl-CoA dehydrogenase; Cro, crotonase; Bcd, butyryl-CoA dehydrogenase (including electron transfer protein α, β subunits); KamA, lysine-2,3-aminomutase; KamD,E, β-lysine-5,6-aminomutase (α, β subunits); Kdd, 3,5-diaminohexanoate dehydrogenase; Kce, 3-keto-5-aminohexanoate cleavage enzyme; Kal, 3-aminobutyryl-CoA ammonia lyase; AbfH, 4-hydroxybutyrate dehydrogenase; AbfD, 4-hydroxybutyryl-CoA dehydratase; Isom, vinylacetyl-CoA 3,2-isomerase (same protein as AbfD): 4Hbt, butyryl-CoA:4-hydroxybutyrate CoA transferase; But, butyryl-CoA:acetate CoA transferase; Ato, butyryl-CoA:acetoacetate CoA transferase (α, β subunits); Ptb, phosphate butyryltransferase; Buk, butyrate kinase. Cosubstrates for individual butyryl-CoA transferases are shown.

RESULTS

Overview of butyrate synthesis pathways. There are four main pathways known for butyrate production, the acetyl-CoA, glutarate, 4-aminobutyrate, and lysine pathways (Fig. 1). All pathways merge at a central energy-generating step where crotonyl-CoA is transformed to butyryl-CoA, catalyzed by the butyryl-CoA dehydrogenase electron-transferring flavoprotein complex (Bcd-Etf$\alpha\beta$). The final conversion to butyrate is performed by various butyryl-CoA transferases that use cosubstrates either formed earlier in the individual pathways, namely, acetoacetate and 4-hydroxybutyrate for the lysine and 4-aminobutyrate pathways, respectively, or from external sources, as shown for butyryl-CoA:acetate CoA transferase (But) (23). Other transferases not shown in Fig. 1 have been proposed as final enzymes as well (19), and our data support those suggestions (see below) (Fig. 2). Alternatively, butyryl-CoA is phosphorylated and transformed to butyrate via butyrate kinase (Buk), leading to the formation of ATP. A small number of strains contain both But and Buk (see below) (Fig. 2). Since no possible cosubstrate for butyryl-CoA transferase is formed in the glutarate pathway, we considered But and Buk as the final enzymes for that pathway.

Potential butyrate producers detected. Potential microbial functions are commonly inferred from isolates/sequenced genomes and whole communities by targeting specific key genes that characterize the function. However, as mentioned above, such an approach can be problematic in the case of butyrate synthesis, and targeting complete pathways together with several downstream analyses is a more robust way to predict potential function and additionally can provide insights into potential substrate requirements for functional performance. A detailed outline of the screening procedure is presented in Fig. S1 and Text S1 in the supplemental material. Briefly, hidden Markov models (HMM) together with EC number searches on the Integrated Microbial Genome (IMG) platform were used to detect potential genes among genomes, and results were subsequently evaluated based on their synteny among all pathway genes. A gene catalogue containing 3,055 entries from 225 organisms was established (see Data Set S1). We found the acetyl-CoA pathway to be present in the majority of potential butyrate producers. The lysine pathway was represented in many phyla as well, whereas the 4-aminobutyrate- and glutarate-based pathways were the least abundant and were found in only four phyla (namely, *Firmicutes*, *Fusobacteria*, *Spirochaetaceae*, and *Bacteroidetes*). Several isolates exhibit genes for two or three pathways, indicating butyrate synthesis as having a central role in energy conservation. Figure 2 displays all potential butyrate producers obtained, including 124 strains with confirmed functional activity (based on species level). Candidate butyrate produces were isolated from distinct environ-

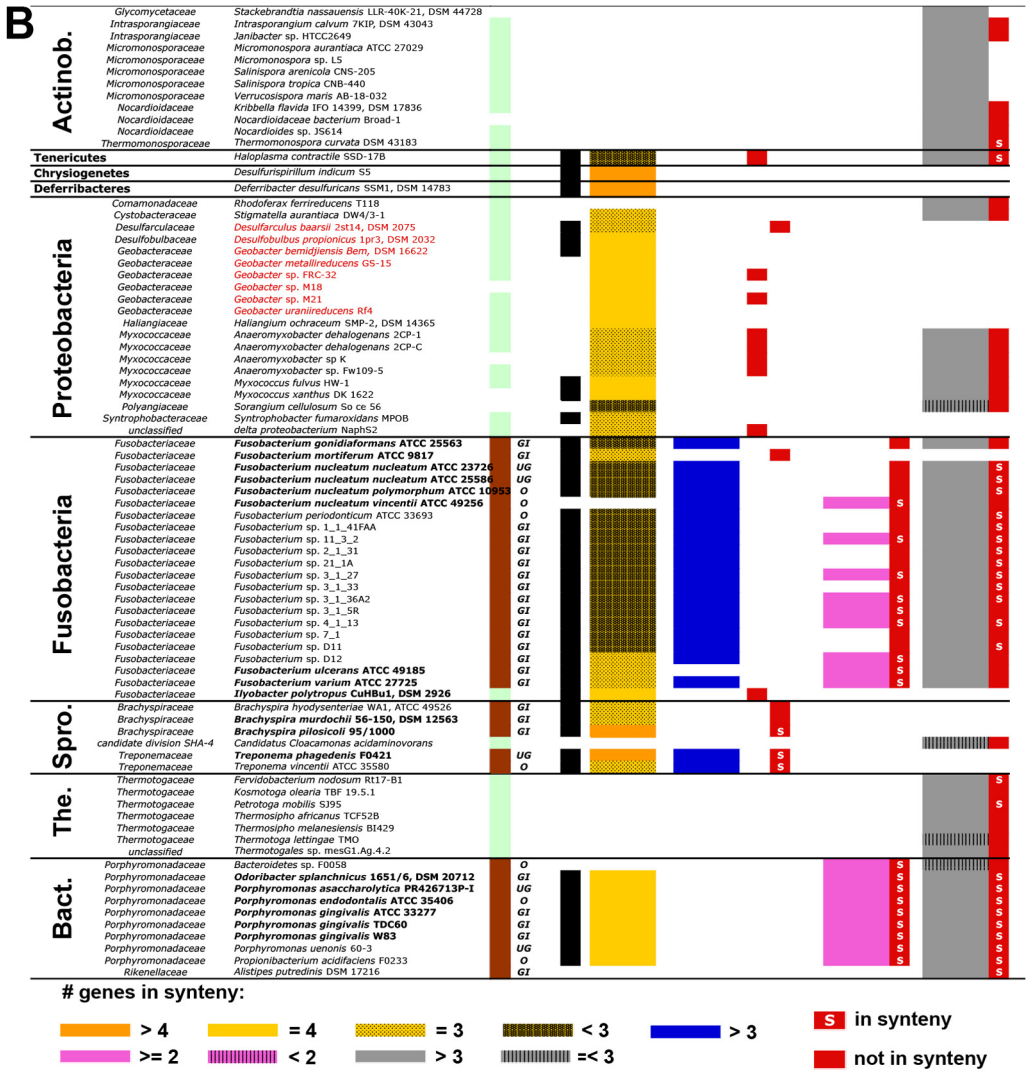

FIG 2 (Continued)

ments and represent a broad taxonomic range associated with 10 different phyla. An additional literature search for nonsequenced butyrate producers revealed that almost all families exhibiting butyrate-producing members, except some strains from *Clostridiales incertae sedis* XIII and the *Synergistaceae* (see Fig. S2), are included in our genome-based study. Hence, we consider our database to be a good representation of the known diversity of butyrate-producing bacteria. Pathway analysis of all strains from individual families confirmed earlier observations that the ability for butyrate production is not consistent within families. Not all members of the same family exhibit butyrate-producing pathways (see Fig. S3), demonstrating that phylogenetic analysis (on the family level) does not enable functional predictions.

As expected, strains belonging to *Firmicutes* were identified as the major butyrate-producing group, exhibiting both demonstrated producers and potential candidates that span 18 different families. These strains were isolated from many environments and different host-associated sites. In this phylum, the acetyl-CoA

FIG 2 A list of all obtained candidate bacteria and their taxonomic classifications. *Firmicutes* are shown in panel A, whereas candidates associated with other phyla are displayed in panel B. Names in bold represent known butyrate-producing strains. Origins of isolates (Isol.), where brown refers to human/animal-associated strains (individual body sites of isolation are as follows: GI, gastrointestinal tract; UG, urogenital tract; O, oral tract) and green to environmental isolates, are given. Individual pathways with corresponding final genes are shown, namely, the acetyl-CoA pathway (AceCoA; orange-yellow) and the glutarate pathway (Gltr; blue) with *but* (encoding butyryl-CoA:acetate CoA transferase; red; light pink represents "atypical" transferases) and *buk* (butyrate kinase; red), as well as the 4-aminobutyrate pathway (4-Amin; pink) with the 4Hbt gene (butyryl-CoA:4-hydroxybutyrate CoA transferase; red) and the lysine pathway (Lys; grey) with *ato* (encoding butyryl-CoA:acetoacetate CoA transferase). Results of synteny analysis for genes of individual pathways are indicated (see key to color patterns at the bottom). Black cells in the column "*Bcd-αβ*" represent the presence of the butyryl-CoA dehydrogenase electron transfer protein complex, i.e., *bcd* is in synteny with the *etf* genes. Names in red indicate isolates that are reported to oxidize butyrate for growth. Actinob., *Actinobacteria*; Spro., *Spirochaetes*; The., *Thermotogae*; Bact., *Bacteroidetes*; *C. Incertae* Sedis, *Clostridiales incertae sedis*. For more explanation, see the text.

pathway is dominant, genes are in good synteny, and the Bcd-ETF$\alpha\beta$ complex is well conserved (Fig. 2; see also Fig. S4 in the supplemental material). Whereas *but* and *buk* were identified as terminal genes in most candidates, some strains, especially the *Erysipelotrichaceae*, contain atypical transferases (Fig. 2). Only a few firmicute isolates exhibit other pathways. Notably, bacteria linked primarily to nonfermentative growth styles, namely, syntrophic growth of *Syntrophomonadaceae* (24) and anaerobic respiration, especially for the *Peptococcaceae* (25), were also detected where the acetyl-CoA pathway is used in a reverse direction to oxidize butyrate. Their gene sequences and arrangements are closely related to those of known butyrate producers (see below; see also Fig. S4), and all exhibit true terminal enzymes.

The *Fusobacteria* display an interesting diverse pattern, where two strains, namely, *Fusobacterium mortiferum* and *Ilyobacter polytropus*, exhibit only the acetyl-CoA pathway (with *but* as the terminal gene), whereas the amino acid-fed pathways, glutarate and lysine, which are the only known route for butyrate production in *Fusobacteria* (20, 26), are most prominent in other strains. We detected genes from the acetyl-CoA pathway in those strains as well, but without synteny and absence of the terminal genes (Fig. 2). However, butyryl-CoA:4-hydroxybutyrate CoA transferase (encoded by the 4-Hbt gene) was found in all strains, while additional genes from the 4-aminobutyrate pathway were often completely lacking. If the acetyl-CoA pathway is indeed performing in those isolates, 4Hbt might take the role as the terminal transferase.

Bacteroidetes, mainly represented by *Porphyromonadaceae*, exhibit three pathways with genes in good synteny. It was surprising to find the acetyl-CoA pathway in *Porphyromonas* species, since this taxon is considered asaccharolytic (27). Notably, this is in accordance with the observed gene arrangements, where this pathway is colocated with the lysine pathway in the same operon (see Fig. S4 in the supplemental material), and acetoacetate-CoA, formed during lysine fermentation, can be directly used as the substrate at the second step (Fig. 2). Accordingly, thiolase (Thl), the enzyme catalyzing the first reaction in the acetyl-CoA pathway, could not be detected in *Porphyromonadaceae*. This "cross-feeding" is probably occurring in all strains exhibiting these two pathways, since it allows for increased energy production via the Bcd-Etf$\alpha\beta$ complex and ferredoxin reduction. It should be noted that a final enzyme for that pathway is missing in this taxon, but terminal transferases linked to other pathways were detected.

Our analysis suggests that several members of *Actinobacteria* and *Thermotogae* contain the lysine pathway for butyrate production. However, we are not aware of any described butyrate-producing member of those phyla, and culture-based experiments containing lysine as a nutrient source need to be performed to confirm them as real butyrate producers.

Our gene and pathway analysis also revealed isolates of the phyla *Chrysiogenetes*, *Deferribacteres*, *Proteobacteria*, *Spirochaetes*, and *Tenericutes* as potential butyrate producers. However, only *Spirochaetes* contain confirmed butyrate-producing members, and genes linked to the acetyl-CoA pathway with *but* as the terminal gene were found in this taxon. Members of the family *Treponemaceae* additionally exhibit the glutarate pathway.

Detailed gene analysis. Detailed sequence analysis of aligned gene products from all candidates revealed several conserved sites for each gene (see Data Set S2 in the supplemental material). Simplified presentations of neighbor-joining trees of the individual genes (protein sequences) are displayed in Fig. 3. Based on the trees, horizontal gene transfer (HGT) signatures were detected, especially for the acetyl-CoA pathway, where genes from individual *Firmicutes* families do not form homogenous groups but interrupt each other. Additionally, phylum-level HGT for members of *Fusobacteria* and *Spirochaetes* was observed (Fig. 3). Trees can be split into four major sections for this pathway; the first contains *Eubacteriaceae*, *Lachnospiraceae*, and *Ruminococcaceae*, disrupted by *Fusobacteria* and *Spirochaetes*, followed by *Erysipelotrichaceae* and members of *Clostridiaceae*. The second part consists of *Clostridiales incertae sedis* XI, *Peptostreptococcaceae*, and all members of *Bacteroidetes*. Strains belonging to *Thermoanaerobacteriaceae* and *Clostridiaceae*, mainly *Clostridium botulinum*, form the third cluster, whereas the bottom section consists of *Proteobacteria* and paralogous genes of *Syntrophomonadaceae* and *Peptococcaceae*. However, some exceptions to this overall trend exist where only one single *thl* gene cluster for all *Clostridiaceae* strains was detected and an additional tight group of crotonase genes linked to certain *Lachnospiraceae* is located outside the taxon's first section, indicating that they have evolved from different precursors than in other strains of *Lachnospiraceae* (Fig. 2). Interestingly, those genes are not in synteny with other genes from that pathway (see Fig. S4). Genes belonging to additional families of *Firmicutes*, such as the *Veillonellaceae*, did not display consistent patterns for this pathway. *Peptococcaceae* and *Syntrophomonadaceae* are clustering together close to known butyrate producers (except for β-hydroxybutyryl–CoA dehydrogenase [encoded by the *hbd* gene]). With only a few exceptions, all individual phyla form tight clusters within the other three pathways analyzed, indicating little HGT (at phylum level). The genes shared by all pathways, i.e., *bcd* and the Etf$\alpha\beta$ genes, did not display consistent patterns associated with a specific pathway (data not shown).

Terminal genes displayed patterns similar to those of other genes of the corresponding pathways, indicating that all genes of an individual pathway coevolved in many of the strains analyzed. Thus, overall, our results suggest that specific types of transferases are indeed associated with a certain pathway. However, the acetyl-CoA pathway, especially, shows exceptions, where alternative transferases were found in several isolates (Fig. 2) (19), and gene arrangement analysis indicated that transferases linked to other pathways might catalyze the final step to butyrate in certain isolates (e.g., see *Fusobacteria* or *Porphyromonas*). Several paralogous genes were detected for both *buk*, associated with *C. botulinum* and *Clostridium difficile*, and *but*, derived mainly from *Lachnospiraceae*, *Syntrophomonadaceae*, and *Veillonellaceae*, at the bottoms of individual trees (Fig. 3).

Metagenomic analysis. Figure 4 displays the overall butyrogenic potential of 15 stool samples of healthy individuals provided by the HMP. High percentages of genomes were calculated to exhibit a pathway (median, 19.1%; range, 3.2% to 39.4%), where the acetyl-CoA pathway was dominating for almost all individuals (mean, 79.7%; range, 46% to 97.5% of all pathways), followed by the lysine pathway, which showed large variations between samples (mean, 11.2%;, range, 0.5% to 49.7% of all pathways) and was especially highly abundant (>35%) for four individuals. The glutarate and 4-aminobutyrate pathways were consistently detected at low abundances (mean, 2.5%; range, 0.8% to 9.6%; and mean, 2.4%; range, 0.3% to 10.5% of all pathways, respectively). The overall butyrogenic potential was estimated as the sum of all detected pathways. Notably, amino acid-fed pathways did not of-

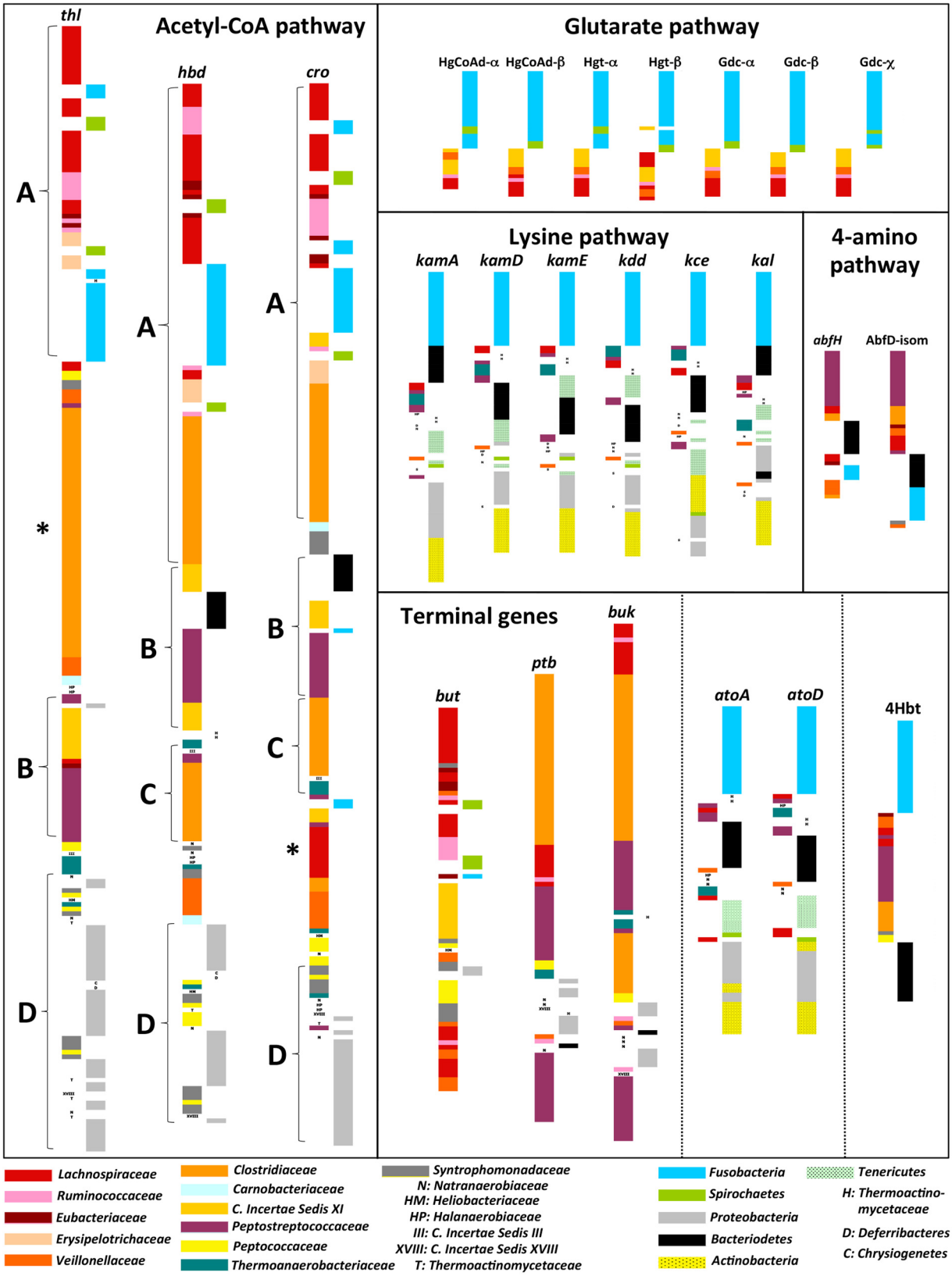

FIG 3 Simplified representations of neighbor-joining trees of individual genes (protein sequences) are shown. The left column in each tree shows arrangement of different genes associated with different families within the phylum *Firmicutes*, whereas gene entries linked to other phyla are given in the right column. For a key to colors see the bottom. Letters (A to D) represent the four distinct regions of individual trees based on genes of the acetyl-CoA pathway, and "*" marks deviations from the overall trend. For an explanation, see the text.

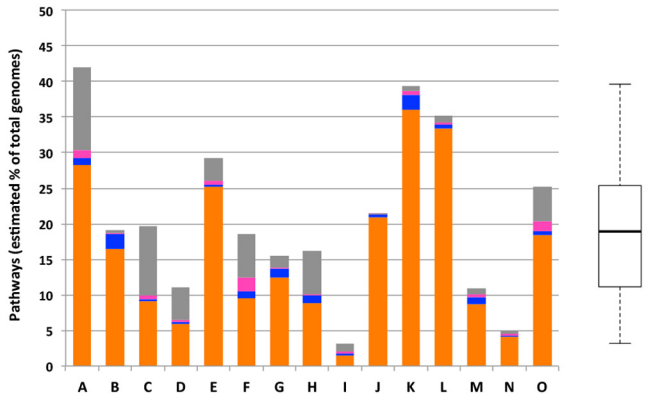

FIG 4 Abundance of butyrate-producing pathways (calculated as a percentage of total bacterial genomes theoretically exhibiting a pathway) in metagenomic data from stool samples of 15 healthy humans is shown. Different colors represent individual pathways (acetyl-CoA pathway, orange; glutarate pathway, blue; 4-aminobutyrate pathway, pink; lysine pathway, grey). The box plot displays the data distribution for all 15 samples analyzed (A to O).

ten occur in genomes alone but usually occurred together with the acetyl-CoA pathway (Fig. 2), and the summarized cumulative results presented in Fig. 4 are hence likely an overestimate.

Detailed analysis revealed a broad diversity of butyrate-producing-pathway genes for individual samples, where almost all detected groups are associated with known butyrate producers (Fig. 5). Interestingly, a few key groups dominated for most individuals, suggesting a butyrate-producing taxonomic core in healthy colons. This core consisted of groups associated with known butyrate producers linked to specific *Lachnospiraceae* and *Ruminococcaceae* for the acetyl-CoA and glutarate pathways. Groups linked to *Odoribacter splanchnicus* and *Alistipes putredinis* (both members of the *Bacteroidetes*) dominate the lysine pathway, whereas groups similar to *O. splanchnicus* and *Clostridium symbiosum* prevailed in the 4-aminobutyrate pathway. These results indicate that butyrate production is not associated solely with members of the phylum *Firmicutes* and suggest that the *Bacteroidetes* are often contributing to the overall butyrogenic potential as well. However, current knowledge of the *Bacteroidetes* suggests that most carbon consumed does not result in butyrate production; hence, metabolic flux studies, under various nutritional conditions, are needed to quantify the contribution of this taxon to the butyrate pool. Obtained read abundances were relatively consistent for all genes of a pathway in an individual group (see Fig. S5 in the supplemental material). Furthermore, the degree of explanation was high, i.e., the amount of reads that matched any gene in our database, which were subsequently also included in diversity analysis, where all genes of a pathway of an individual group had to be detected in order to be considered (see Materials and Methods). However, especially for the lysine pathway, the detected genes of the entire pathway were occasionally split between different groups, i.e., no group was positive for all genes of that pathway, which inhibited diversity analysis for some samples (not for those exhibiting an overall high abundance of this pathway) (Fig. 5).

but and *buk* were the dominating terminal genes in most samples for the acetyl-CoA pathway, with median abundances of 77.2% and 21.8%, respectively (see Fig. S6 in the supplemental material). Alternative transferases were detected only at very low abundances, suggesting that those enzymes do not play an important role for butyrate synthesis in healthy humans. Although *but* is the most prevalent terminal gene in our metagenomic data (median, 61.8%; range: 24.7% to 85.1% [considering all pathways]), it represents only one terminal point of the butyrate-producing pathways, and studies targeting only *but* for total functional analysis should be aware of this limitation.

DISCUSSION

The established gene catalogue together with our metagenomic analysis allowed us to reveal microbial butyrate-producing communities in the healthy microbiota and their associated metabolic pathways. This metabolic framework is a critical step in investigating the role of this function in host health and disease. Although targeting complete pathways is a more robust way to predict function than single-gene analysis, their detection in genomes does not automatically imply functionality, since that must be done by specific biochemical testing. For several isolates, such as members of *Peptococcaceae* and *Syntrophomonaceae*, the detected ability to produce butyrate is doubtful, since they are known rather to oxidize butyrate for growth (see reference 28). This is also true for the majority of the *Proteobacteria* shown in Fig. 2, which belong to the delta class, that use anaerobic respiration for energy conservation, and butyrate consumption is documented for several isolates (e.g., see reference 29). In these taxa, pathway genes are often not in synteny and only distantly related to genes of confirmed butyrate producers (Fig. 3), and terminal genes are missing in many strains. However, it cannot be excluded that certain environmental conditions, such as the absence of H_2-consuming bacteria or lack of appropriate inorganic electron acceptors, might trigger fermentative growth and the synthesis of butyrate in certain isolates. Furthermore, a few strains are known to generate butyrate as building blocks for secondary metabolites, such as salinosporamide B, produced by the actinobacterium *Salinispora tropica* (30).

Neighbor-joining trees revealed very consistent patterns for all genes of an individual pathway, indicating a high degree of coevolution. Nevertheless, clear HGT signatures were detected in isolates, especially for the acetyl-CoA pathway, confirming earlier findings (31). However, our results indicate transfer of entire pathways rather than of single genes. The fast microbial turnover and enormous selective pressures in the colonic environment promote large-scale HGT (32). Since the acetyl-CoA pathway was detected to be the dominant pathway, displaying the greatest diversity, observations of HGT signatures specifically for this pathway make sense. Furthermore, our metagenomic results also did not detect unknown "disconnected HGT" events, i.e., bacteria that acquired genes of the acetyl-CoA pathway from distinct precursors (representing unknown gene combinations). This supports the observed coevolutionary behavior of all genes in this pathway. However, for the lysine pathway, the presence of gene combinations that have not yet been captured in sequenced isolates was indicated.

Diet is a major external force shaping gut communities (33). Good reviews of studies investigating the influence of diet on butyrate-producing bacteria exist (11 and 34) and suggest that plant-derived polysaccharides such as starch and xylan, as well as cross-feeding mechanisms with lactate-producing bacteria, are the main factors governing their growth. Our metagenomic analysis supports the acetyl-CoA pathway as the main pathway for butyrate production in healthy individuals (Fig. 4), implying that

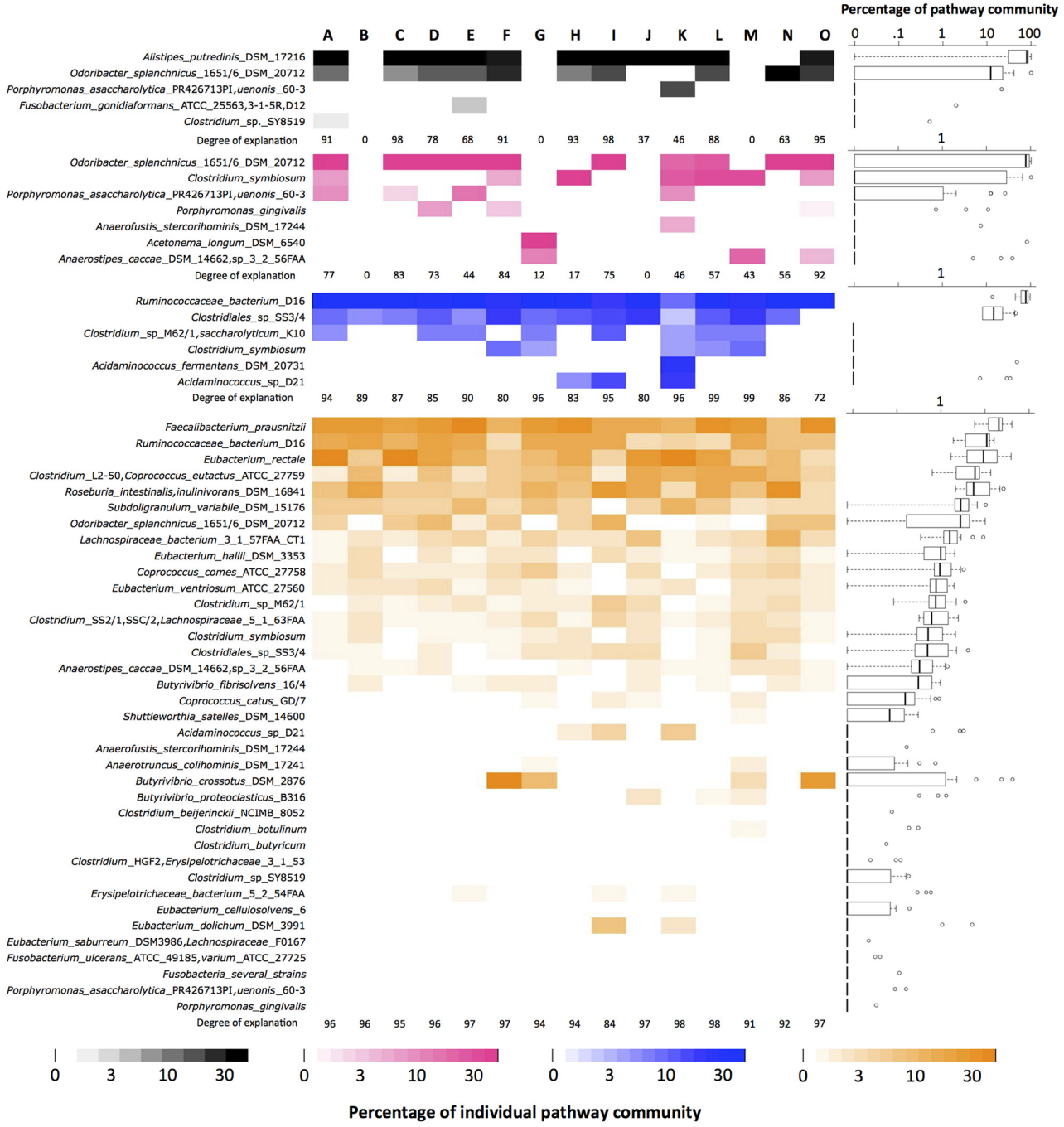

FIG 5 The detected diversity in metagenomic data associated with individual pathways. Colors correspond to different pathways (acetyl-CoA pathway, orange; glutarate pathway, blue; 4-aminobutyrate pathway, pink; lysine pathway, grey). Bacterial names represent members of individual groups (based on 10% complete linkage clustering; for details, see Materials and Methods). Groups consist of the following: (i) only one reference genome (indicated by single strain names), (ii) merged strains of the same species (indicated by species name without strain information), and (iii) merged genomes from distinct species (individual names are given). The group "Fusobacteria several strains" consists of the following strains: *Fusobacterium nucleatum* subsp. *nucleatum*, *Fusobacterium nucleatum* subsp. *polymorphum* ATCC 10953, *Fusobacterium periodonticum* ATCC 33693, sp. 1_1_41FAA, sp. 11_3_2, sp. 2_1_31, sp. 21_1A, sp. 3_1_27, sp. 3_1_33, sp. 3_1_36A2, sp. 4_1_13, sp. 7_1, sp. D11, and sp. D12. For more information on taxon assignment, see Materials and Methods. The box plots display data distributions for each group of all 15 samples analyzed (A to O). The degree of explanation indicates the percentages of reads matching a group, which was included in diversity analysis (this figure). For more explanation, see the text.

a sufficient polysaccharide supply is probably sustaining a well-functioning butyrate-producing community, at least in these North American subjects. However, the detection of additional amino acid-fed pathways, especially the lysine pathway, indicates that proteins could also play an important role in butyrate synthesis and suggests some flexibility of the microbiota to adapt to various nutritional conditions maintaining butyrate synthesis. Whether the prevalence of amino acid-fed pathway is associated with a protein-rich diet still needs to be assessed. It should be noted that those pathways are not restricted to single substrates, as displayed in Fig. 1, i.e., glutarate and lysine, but additional amino acids, such as aspartate, can be converted to butyrate via those routes as well (26). Furthermore, the acetyl-CoA pathway also can be supplied with substrates derived from proteins either by cross-feeding with the lysine pathway (as discussed above) or by direct fermentation of amino acids to acetyl-CoA (35). However, whereas diet-derived proteins are probably important for butyrate synthesis in the ileum, where epithelial cells use butyrate as a main energy source as well (36), it still needs to be assessed whether enough proteins reach the human colon to serve as a major nutrient source for microorganisms. Another possible colonic protein source could originate with lysed bacterial cells. Enormous viral loads have been detected in this environment, suggesting fast cell/nutrient turnover, which might explain the presence of corresponding pathways in both fecal isolates and metagenomic data (Fig. 1, 4, and 5). Detailed investigations of butyrate-producing communities in the colon of carnivorous animals will add additional key information on the role of proteins in butyrate production in that environment. It should be noted that diet provides only a part of the energy/carbon sources for microbial growth in the colon, since host-derived mucus glycans serve as an important nutrient source as well. Several butyrate-producing organisms do specifically colonize mucus (37), and for some, growth on mucus-derived substrates was shown (38).

Systems biology together with metabolic modeling is a promising approach to handle complexities of nutrient fluxes within the gut microbiota and will eventually help in predicting functional performance (39). This study provides an important step forward, since it enabled us to assess the butyrate-producing potential of complex microbial communities, including predictions of basic nutritional requirements for butyrate synthesis. However, next to substrate availability, additional factors, such as pH, were demonstrated to be important factors governing the successful competition of butyrate producers with other intestinal organisms (11). Furthermore, the presence of butyrate-producing pathways alone might not allow optimal predictions of actual butyrate production, since the organisms involved show metabolic flexibility and diverse profiles of fermentation products. Butyrate synthesis was shown to be influenced by several factors, such as type of limiting substrate and growth rate (40), oxygen concentration (41), and growth style (attached versus unattached [42]). Furthermore, both the presence of inorganic electron acceptors promoting anaerobic respiration and aceto-/methanogenesis lowering the H_2 partial pressure can lead to more oxidized fermentation products, especially acetate, at the expense of more reduced substances, such as butyrate (40). Our metagenomic approach, in combination with additional "-omics"-based technologies, will help to improve functional predictions and to assess the resulting effects on the host.

MATERIALS AND METHODS

Establishing the gene catalogue. Individual pathways shown in Fig. 1 are based on KEGG with modifications. Most importantly, the entire lysine pathway and certain steps in the 4-aminobutyrate pathway are not present in KEGG and were included based on references 22 and 43. KEGG additionally displays the conversion from butanol to butyrate, which was not included in this study. Furthermore, a possible route from acetoacetate via poly-β-hydroxybutyrate and crotonoyl-CoA to butyrate is suggested in KEGG. However, this pathway contains an unlikely reverse reaction of extracellular poly-β-hydroxybutyrate degradation enzymes that differ considerably from intracellular depolymerases (44), and this route was hence not considered. The stereospecific separation between *R*-hydroxybutyrate and *S*-hydroxybutyrate in the acetyl-CoA pathway was omitted, and the two routes were merged.

Screening of genomes was divided into two main parts, where the first was based on EC number searches (from KEGG) within the Integrated Microbial Genome (IMG) (http://img.jgi.doe.gov) database and the second part used HMM models (both approaches were applied on a protein level). A detailed schematic representation of the work flow and abundance of obtained candidates (and associated genes) at each step is given in Fig. S1 in the supplemental material. First, all genes matching individual EC numbers were obtained, and the data were queried for all candidates exhibiting all genes of a specific pathway. Since several model butyrate producers failed the query, we allowed for one missing gene in each pathway. Candidates were then subjected to synteny analysis (see Fig. S1 and Text S1 in the supplemental material). Since it was proposed that several different gene products are able to catalyze the final step in the acetyl-CoA pathway and their location is often apart from other genes in this pathway, we excluded the terminal enzymes here and treated them in separate analyses. After these first steps, we harvested genes from model butyrate producers and candidate strains displaying all genes of the individual pathway in close synteny (not considering terminal genes) and used the obtained sequences to construct HMM models to screen genomes again. After applying certain cutoffs based on HMM scores (for details, see Fig. S1 and Text S1), candidates were filtered for exhibiting entire pathways (allowing one missing gene), and terminal genes were treated in separate analyses (for details, see Fig. S1 and Text S1). Finally, candidates from both EC number and HMM searches were combined and subjected to additional filtering based on detailed gene analysis considering synteny and phylogenetic trees (for details, see Fig. S1 and Text S1). Protein sequences were aligned in the software program Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo), and neighbor-joining trees were constructed using the program MEGA (http://www.megasoftware.net). Taxonomy is displayed as provided by IMG with some modifications for the phylum *Firmicutes* based on RDP's classifications.

Analysis of metagenomic data. Stool samples from 15 different individuals were randomly selected from the HMP Data Analysis and Coordination Center (http://www.hmpdacc.org; parameters defining health can be obtained from the website). Raw nucleotide read sequences were aligned (blastn) against our database, requiring a minimum alignment length of 70 bp and sequence identity of ≥80%. Only the best-scoring alignment (lowest E value) was used for further analysis. The abundance of individual butyrate-producing pathways (Fig. 4) was calculated as follows: (i) $(\#reads_{tot} \times length_{pathway})/4 \times 10^6$ bp $= th_{100\%}$, and (ii) $\#reads_{pathway}/th_{100\%}$ = result (genomes exhibiting pathway [%]), where $\#reads_{tot}$ is the total number of reads for a sample, $length_{pathway}$ stands for the total length (bp) of all unique pathway genes (calculated from the median length of all entries in the database for a specific gene), 4×10^6 bp corresponds to an average genome size, $th_{100\%}$ is the theoretical number of reads if all genomes exhibit the pathway, and $\#reads_{pathway}$ corresponds to the number of reads matching the pathway (BLAST result). Detailed results are presented in Fig. S7 in the supplemental material.

Prior to diversity analysis, individual genes from the database were subjected to multiple complete linkage clustering (using the Pyrosequencing Pipeline provided by the Ribosomal Database Project; http://rdp.cme

.msu.edu) on the nucleotide level, applying a 10% cutoff. All genes of an individual pathway clustered very similarly (clusters for all individual pathway genes were usually associated with the same genomes), allowing us to group individual clusters of all genes of a specific pathway together. Thus, obtained groups contained all genes of a specific pathway. If cluster results varied between genes (e.g., all *thl* genes from three candidates cluster together, whereas two clusters were generated for the *hbd* gene), then clusters were manually merged (e.g., merging of all three *hbd* genes as associated *thl* genes) to achieve consistency, and the most conservative approach was always applied, i.e., clusters were only merged and never split. Genes of the same strain were always merged. For metagenomic analysis, a specific group (e.g., the group *Faecalibacterium prausnitzii* for the acetyl-CoA pathway consists of all pathway genes from all five strains of this taxon) was considered present only if all pathway genes could be identified for that group in the BLAST result (thus, BLAST hits did not have to match all genes from the same strain but only from the same group—an example [sample A] is shown in Fig. S5 in the supplemental material). Results presented in Fig. 5 are a median value for all individual pathway genes (see Fig. S5). The degree of explanation was calculated as the percentage of reads matching groups that were included in the diversity analysis (average from individual genes) from the total number of reads matching any gene in the database.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00889-14/-/DCSupplemental.

Text S1, DOCX file, 0.2 MB.
Figure S1, TIF file, 0.3 MB.
Figure S2, EPS file, 0.4 MB.
Figure S3, EPS file, 0.3 MB.
Figure S4, PDF file, 3.2 MB.
Figure S5, PDF file, 0.2 MB.
Figure S6, EPS file, 0.2 MB.
Figure S7, EPS file, 0.6 MB.
Data Set S1, XLSX file, 0.1 MB.
Data Set S2, PDF file, 0.2 MB.

ACKNOWLEDGMENTS

Financial support was provided by the NIH Human Microbiome Project Demonstration Project (UH3 DK083993).

We thank Mike Rizzo and Kris Opron for assistance in data analysis.

REFERENCES

1. **Sørensen J, Christensen D, Jørgensen BB.** 1981. Volatile fatty acids and hydrogen as substrates for sulfate-reducing bacteria in anaerobic marine sediment. Appl. Environ. Microbiol. **42:**5–11.
2. **Paillard D, McKain N, Chaudhary LC, Walker ND, Pizette F, Koppova I, McEwan NR, Kopecný J, Vercoe PE, Louis P, Wallace RJ.** 2007. Relation between phylogenetic position, lipid metabolism and butyrate production by different *Butyrivibrio*-like bacteria from the rumen. Antonie Van Leeuwenhoek **91:**417–422. http://dx.doi.org/10.1007/s10482-006-9121-7.
3. **Shah HN, Williams RA, Bowden GH, Hardie JM.** 1976. Comparison of the biochemical properties of *Bacteroides melaninogenicus* from human dental plaque and other sites. J. Appl. Bacteriol. **41:**473–495. http://dx.doi.org/10.1111/j.1365-2672.1976.tb00660.x.
4. **Pryde SE, Duncan SH, Hold GL, Stewart CS, Flint HJ.** 2002. The microbiology of butyrate formation in the human colon. FEMS Microbiol. Lett. **217:**133–139. http://dx.doi.org/10.1111/j.1574-6968.2002.tb11467.x.
5. **Hamer HM, Jonkers D, Venema K, Vanhoutvin S, Troost FJ, Brummer RJ.** 2008. Review article: the role of butyrate on colonic function. Aliment. Pharmacol. Ther. **27:**104–119. http://dx.doi.org/10.1111/j.1365-2036.2007.03562.x.
6. **Roediger WE.** 1982. Utilization of nutrients by isolated epithelial cells of the rat colon. Gastroenterology **83:**424–429.
7. **Cani PD, Amar J, Iglesias MA, Poggi M, Knauf C, Bastelica D, Neyrinck AM, Fava F, Tuohy KM, Chabo C, Waget A, Delmée E, Cousin B, Sulpice T, Chamontin B, Ferrières J, Tanti JF, Gibson GR, Casteilla L, Delzenne NM, Alessi MC, Burcelin R.** 2007. Metabolic endotoxemia initiates obesity and insulin resistance. Diabetes **56:**1761–1772. http://dx.doi.org/10.2337/db06-1491.
8. **Macia L, Thorburn AN, Binge LC, Marino E, Rogers KE, Maslowski KM, Vieira AT, Kranich J, Mackay CR.** 2012. Microbial influences on epithelial integrity and immune function as a basis for inflammatory diseases. Immunol. Rev. **245:**164–176. http://dx.doi.org/10.1111/j.1600-065X.2011.01080.x.
9. **Clemente JC, Ursell LK, Parfrey LW, Knight R.** 2012. The impact of the gut microbiota on human health: an integrative view. Cell **148:**1258–1270. http://dx.doi.org/10.1016/j.cell.2012.01.035.
10. **Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R, Pedersen O, Kristiansen K, Wang J.** 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature **490:**55–60. http://dx.doi.org/10.1038/nature11450.
11. **Louis P, Flint HJ.** 2009. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol. Lett. **294:**-1–8. http://dx.doi.org/10.1111/j.1574-6968.2009.01514.x.
12. **Duncan SH, Louis P, Flint HJ.** 2004. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. Appl. Environ. Microbiol. **70:**5810–5817. http://dx.doi.org/10.1128/AEM.70.10.5810-5817.2004.
13. **Bennett G, Rudolph F.** 1995. The central metabolic pathway from acetyl-CoA to butyryl-CoA in *Clostridium acetobutylicum*. FEMS Microbiol. Rev. **17:**241–249. http://dx.doi.org/10.1111/j.1574-6976.1995.tb00208.x.
14. **Boynton ZL, Bennet GN, Rudolph FB.** 1996. Cloning, sequencing, and expression of clustered genes encoding beta-hydroxybutyryl-coenzyme A (CoA) dehydrogenase, crotonase, and butyryl-CoA dehydrogenase from *Clostridium acetobutylicum* ATCC 824. J. Bacteriol. **178:**3015–3024.
15. **Herrmann G, Jayamani E, Mai G, Buckel W.** 2008. Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. J. Bacteriol. **190:**784–791. http://dx.doi.org/10.1128/JB.01422-07.
16. **Louis P, Flint HJ.** 2007. Development of a semiquantitative degenerate real-time PCR-based assay for estimation of numbers of butyryl-coenzyme A (CoA) CoA transferase genes in complex bacterial samples. Appl. Environ. Microbiol. **73:**2009–2012. http://dx.doi.org/10.1128/AEM.02561-06.
17. **Vital M, Penton CR, Wang Q, Young VB, Antonopoulos DA, Sogin ML, Morrison HG, Raffals L, Chang EB, Huffnagle GB, Schmidt TM, Cole JR, Tiedje JM.** 2013. A gene-targeted approach to investigate the intestinal butyrate-producing bacterial community. Microbiome **1:**8. http://dx.doi.org/10.1186/2049-2618-1-8.
18. **Buckel W, Dorn U.** 1981. Glutaconate CoA-transferase from *Acidaminococcus fermentans*. Eur. J. Biochem. **321:**315–321.
19. **Eeckhaut V, Van Immerseel F, Croubels S, De Baere S, Haesebrouck F, Ducatelle R, Louis P, Vandamme P.** 2011. Butyrate production in phylogenetically diverse *Firmicutes* isolated from the chicken caecum. Microb. Biotechnol. **4:**503–512. http://dx.doi.org/10.1111/j.1751-7915.2010.00244.x.
20. **Barker HA, Kahn JM, Hedrick L.** 1982. Pathway of lysine degradation in *Fusobacterium nucleatum*. J. Bacteriol. **152:**201–207.
21. **Buckel W, Barker HA.** 1974. Two pathways of glutamate fermentation by anaerobic bacteria. J. Bacteriol. **117:**1248–1260.
22. **Gerhardt A, Cinkaya I, Linder D, Huisman G, Buckel W.** 2000. Fermentation of 4-aminobutyrate by *Clostridium aminobutyricum*: cloning of two genes involved in the formation and dehydration of 4-hydroxybutyryl-CoA. Arch. Microbiol. **174:**189–199. http://dx.doi.org/10.1007/s002030000195.
23. **Duncan SH, Barcenilla A, Stewart CS, Pryde SE, Flint HJ.** 2002. Acetate utilization and butyryl coenzyme A (CoA):acetate-CoA transferase in butyrate-producing bacteria from the human large intestine. Appl. Environ. Microbiol. **68:**5186–5190. http://dx.doi.org/10.1128/AEM.68.10.5186-5190.2002.
24. **Schink B.** 1997. Energetics of syntrophic cooperation in methanogenic degradation. Microbiol. Mol. Biol. Rev. **61:**262–280.
25. **Widdel F, Pfennig N.** 1977. A new anaerobic, sporing, acetate-oxidizing,

sulfate-reducing bacterium, *Desulfotomaculum* (*emend.*) *acetoxidans*. Arch. Microbiol. **112:**119–122. http://dx.doi.org/10.1007/BF00446665.
26. **Gharbia SE, Shah HN.** 1991. Pathways of glutamate catabolism among *Fusobacterium* species. J. Gen. Microbiol. **137:**1201–1206. http://dx.doi.org/10.1099/00221287-137-5-1201.
27. **Shah HN, Collins MD.** 1988. Proposal for reclassification of *Bacteroides asaccharolyticus*, *Bacteroides gingivalis*, and *Bacteroides endodontalis* in a new genus, *Porphyromonas*. Int. J. Syst. Bacteriol. **38:**128–131. http://dx.doi.org/10.1099/00207713-38-1-128.
28. **Widdel F, Pfennig N.** 1981. Sporulation and further nutritional characteristics of *Desulfotomaculum acetoxidans*. Arch. Microbiol. **129:**401–402. http://dx.doi.org/10.1007/BF00406471.
29. **Shelobolina ES, Vrionis HA, Findlay RH, Lovley DR.** 2008. *Geobacter uraniireducens* sp. nov., isolated from subsurface sediment undergoing uranium bioremediation. Int. J. Syst. Evol. Microbiol. **58:**1075–1078. http://dx.doi.org/10.1099/ijs.0.65377-0.
30. **Beer LL, Moore BS.** 2007. Biosynthetic convergence of salinosporamides A and B in the marine actinomycete *Salinispora tropica*. Org. Lett. **9:**845–848. http://dx.doi.org/10.1021/ol063102o.
31. **Louis P, McCrae SI, Charrier C, Flint HJ.** 2007. Organization of butyrate synthetic genes in human colonic bacteria: phylogenetic conservation and horizontal gene transfer. FEMS Microbiol. Lett. **269:**240–247. http://dx.doi.org/10.1111/j.1574-6968.2006.00629.x.
32. **Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ.** 2011. Ecology drives a global network of gene exchange connecting the human microbiome. Nature **480:**241–244. http://dx.doi.org/10.1038/nature10571.
33. **Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD.** 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science **334:**105–108. http://dx.doi.org/10.1126/science.1208344.
34. **Scott K, Duncan S.** 2008. Dietary fibre and the gut microbiota. Nutr. Bull. **33:**-201–211. http://dx.doi.org/10.1111/j.1467-3010.2008.00706.x.
35. **Barker HA.** 1981. Amino acid degradation by anaerobic bacteria. Annu. Rev. Biochem. **50:**23–40. http://dx.doi.org/10.1146/annurev.bi.50.070181.000323.
36. **Chapman MA, Hutton M, Grahn MF, Williams NS.** 1997. Metabolic adaptation of terminal ileal mucosa after construction of an ileoanal pouch. Br. J. Surg. **84:**71–73. http://dx.doi.org/10.1002/bjs.1800840127.
37. **Van den Abbeele P, Belzer C, Goossens M, Kleerebezem M, De Vos WM, Thas O, de Weirdt R, Kerckhof FM, Van de Wiele T.** 2013. Butyrate-producing *Clostridium* cluster XIVa species specifically colonize mucins in an *in vitro* gut model. ISME J. **7:**1–13. http://dx.doi.org/10.1038/ismej.2012.158.
38. **Levine UY, Looft T, Allen HK, Stanton TB.** 2013. Butyrate-producing bacteria, including mucin degraders, from the swine intestinal tract. Appl. Environ. Microbiol. **79:**3879–3881. http://dx.doi.org/10.1128/AEM.00589-13.
39. **Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C.** 2013. Computational meta'omics for microbial community studies. Mol. Syst. Biol. **9:**666. http://dx.doi.org/10.1038/msb.2013.22.
40. **Macfarlane S, Macfarlane GT.** 2003. Regulation of short-chain fatty acid production. Proc. Nutr. Soc. **62:**67–72. http://dx.doi.org/10.1079/PNS2002207.
41. **Khan MT, Duncan SH, Stams AJ, van Dijl JM, Flint HJ, Harmsen HJ.** 2012. The gut anaerobe *Faecalibacterium prausnitzii* uses an extracellular electron shuttle to grow at oxic-anoxic interphases. ISME J. **6:**1–8. http://dx.doi.org/10.1038/ismej.2011.71.
42. **Macfarlane S, Macfarlane GT.** 2006. Composition and metabolic activities of bacterial biofilms colonizing food residues in the human gut. Appl. Environ. Microbiol. **72:**6204–6211. http://dx.doi.org/10.1128/AEM.00754-06.
43. **Kreimeyer A, Perret A, Lechaplais C, Vallenet D, Médigue C, Salanoubat M, Weissenbach J.** 2007. Identification of the last unknown genes in the fermentation pathway of lysine. J. Biol. Chem. **282:**-7191–7197. http://dx.doi.org/10.1074/jbc.M609829200.
44. **Jendrossek D, Handrick R.** 2002. Microbial degradation of polyhydroxyalkanoates. Annu. Rev. Microbiol. **56:**403–432. http://dx.doi.org/10.1146/annurev.micro.56.012302.160838.