nature microbiology

ARTICLES

PUBLISHED: 13 JUNE 2016 | ARTICLE NUMBER: 16088 | DOI: 10.1038/NMICROBIOL.2016.88

Species–function relationships shape ecological properties of the human gut microbiome

Sara Vieira-Silva[1,2†], Gwen Falony[1,2†], Youssef Darzi[1,2,3], Gipsi Lima-Mendez[1,2,3], Roberto Garcia Yunta[2,3], Shujiro Okuda[4], Doris Vandeputte[1,2,3], Mireia Valles-Colomer[1,2], Falk Hildebrand[2,3], Samuel Chaffron[1,2,3] and Jeroen Raes[1,2,3]*

Despite recent progress, the organization and ecological properties of the intestinal microbial ecosystem remain under-investigated. Here, using a manually curated metabolic module framework for (meta-)genomic data analysis, we studied species–function relationships in gut microbial genomes and microbiomes. Half of gut-associated species were found to be generalists regarding overall substrate preference, but we observed significant genus-level metabolic diversification linked to bacterial life strategies. Within each genus, metabolic consistency varied significantly, being low in Firmicutes genera and higher in *Bacteroides*. Differentiation of fermentable substrate degradation potential contributed to metagenomic functional repertoire variation between individuals, with different enterotypes showing distinct saccharolytic/proteolytic/lipolytic profiles. Finally, we found that module-derived functional redundancy was reduced in the low-richness *Bacteroides* enterotype, potentially indicating a decreased resilience to perturbation, in line with its frequent association to dysbiosis. These results provide insights into the complex structure of gut microbiome-encoded metabolic properties and emphasize the importance of functional and ecological assessment of gut microbiome variation in clinical studies.

The gut microbiota contributes numerous metabolic processes to its human host, ranging from the breakdown of complex carbohydrates and proteins to the production of essential micronutrients. Within the microbiota, functional repartition is believed to be driven both by microbial competition and niche specialization and by host selective pressure towards commensalism and functional redundancy[1]. However, because the functional characterization of individual colon bacteria is hampered by isolation and cultivation limitations[2], the current understanding of ecosystem redundancy and complementarity patterns and the resulting intricate food webs remains limited[3–5]. The recent data explosion in the microbiome field provides opportunities for functional gut microbiology research to assess microbiota covariation with respect to host and environmental parameters, niche diversification and microbiota assembly through genome and metagenome mining[6–8]. Here, by applying a combination of a gut-specific metabolic framework and trait-driven (meta-)genome analytical methods, we have (1) constituted a comprehensive map of species-function relationships in the colon ecosystem, (2) investigated functional specialization in gut species, (3) studied the phylogenetic consistency of metabolic capacity, (4) mapped functional microbiome diversification throughout the ecosystem diversity landscape, and (5) explored functional redundancy as a microbiome resilience indicator.

Results and discussion

Gut-specific metabolic pathway reconstruction. The outcome of fermentation processes in the gut ecosystem has important consequences for host health[9,10]. To assess the diversity in metabolic potential and anaerobic fermentation capacity encoded in (meta-)genomic sequences, we developed a targeted gut metabolic analysis framework based on a set of manually curated reference modules (gut metabolic modules (GMMs); for an extended description see Methods, and for the source file see Supplementary Data 1). Each GMM represents a cellular enzymatic process, defined as a set of orthologue groups and delimited by input and output metabolites. GMMs capture prokaryotic, anaerobe catabolism of carbohydrates, amino acids and lipids, cross-feeding interactions, and the production of fermentation end products, with additional modules covering archaeal methanogenesis and enzymatic conversions reported to be involved in mucus degradation. The modules set is refined from fermentation processes as described in the MetaCyc (ref. 11) and KEGG (ref. 12) databases, through manual curation and extensive literature review (for references, see Supplementary Data 1). In addition, compared to a more generic database approach, the targeted nature of the GMM set presented avoids the detection of ecosystem-extraneous metabolic processes, reduces the weight of multiple testing correction, and facilitates interpretation[13].

Metabolic potential of gut bacteria. Applying the GMM analysis framework to a data set of 532 publicly available gut reference genomes, covering 260 species (Supplementary Table 1), we assessed the genomic investment of gut species in the exploitation of the three major nutrient sources in the intestine[9]: (poly-)saccharides, proteins and lipids (Fig. 1a and Supplementary Table 2). Among the reference genomes available, 53% of gut-associated microbes encoded quasi-equal relative amounts of saccharolytic, proteolytic and lipolytic modules (Fig. 1a and Supplementary Fig. 1). These lineages, predominantly Firmicutes, thus represented generalist species with regard to the metabolism of major nutrients provided through the hosts' diet. This observation was further

[1]Department of Microbiology and Immunology, KU Leuven–University of Leuven, Rega Institute, Herestraat 49, B-3000 Leuven, Belgium. [2]VIB, Center for the Biology of Disease, Herestraat 49, B-3000 Leuven, Belgium. [3]Microbiology Unit, Faculty of Sciences and Bioengineering Sciences, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. [4]Niigata University Graduate School of Medical and Dental Sciences 1-757 Asahimachi-dori, Chuo-ku, Niigata 951-8510, Japan. †These authors contributed equally to this work. *e-mail: jeroen.raes@kuleuven.be

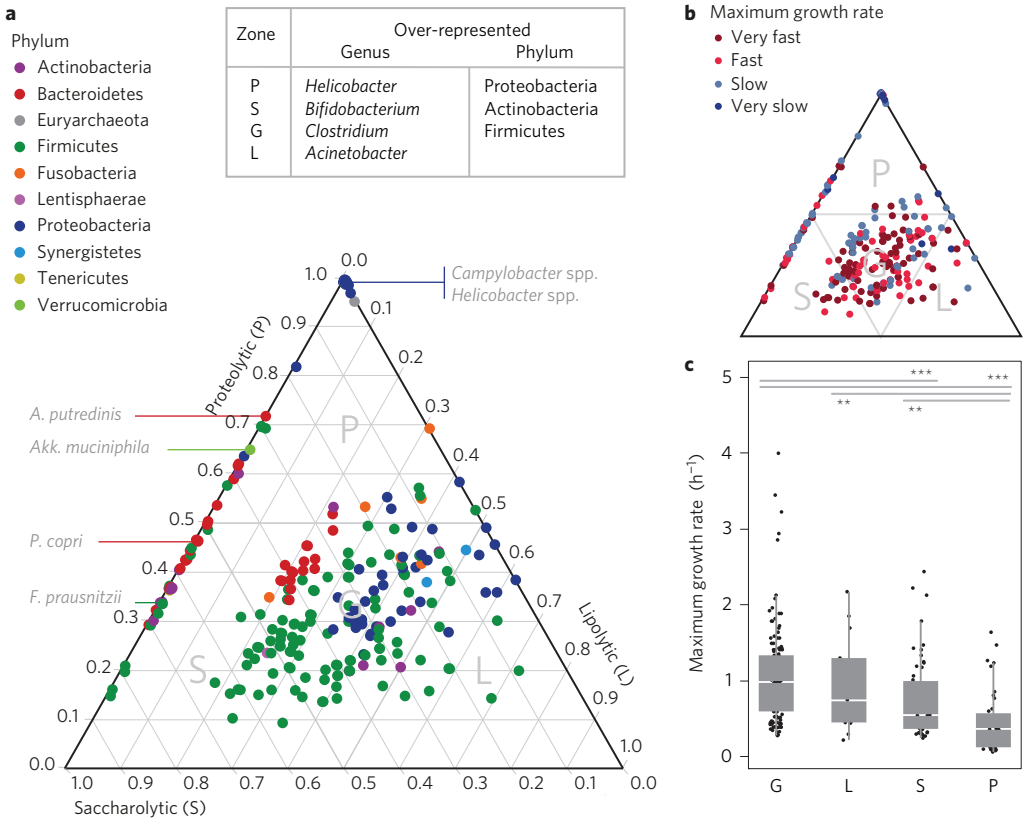

Figure 1 | Gut reference species' saccharolytic, proteolytic and lipolytic fermentation potential. a, Triplot representation of the genomic investment in each of the three fermentation types, defined as average copy number of GMMs. The centre represents equal investment in all three fermentation types (G, generalism), while departure towards a corner corresponds to increased investment in that particular fermentation type (P, proteolytic; S, saccharolytic; L, lipolytic). The table provides over-represented genera/phyla (Fisher exact test) for each area (G, P, L, S), and a few examples are listed. **b**, Same triplot coloured according to bacterial maximum growth rate categories. **c**, Distribution of maximum growth rate (h^{-1}) in each metabolic specialization zone (MWU, significance levels: FDR < 0.01**, <0.001***). The inner line of the boxplot is the median, the edges of the box are the third (Q3) and first (Q1) quartiles, and the whiskers extend to Q3(Q1) +(–) 1.5 × interquartile range, or the last data point when below the computed ranges.

supported by a significantly higher number of transcription factors ($N = 260$ [non-redundant genomes – one random genome per species]; Mann–Whitney U (MWU) $r = 0.56$, $P = 2 \times 10^{-19}$; Supplementary Table 3) and transporters ($r = 0.52$, $P = 5 \times 10^{-17}$, $N = 260$) encoded by these generalists, suggesting an increased genomic investment in sensing and responding to various stimuli to exploit a broad range of substrates[14]. Besides this predominant metabolic generalism among colon bacteria, we identified multiple specific examples of niche specialization (Fig. 1a and Supplementary Table 2). Our analyses confirmed the predominantly saccharolytic nature of Actinobacteria (including (probiotic) *Bifidobacterium* spp.[15]). A total of 40% of Bacteroidetes spp. were identified as generalists (mostly belonging to the *Bacteroides* and *Parabacteroides* genera), while 25% of them invested predominantly in the metabolization of proteins (for example, *Alistipes putredinis*), in line with previous, experimental results of specific strains in these groups[16,17]. Intestinal isolates with increased lipolytic potential were rather uncommon and comprised Proteobacteria (with 6% being classified as lipolytic) and a limited number of Firmicutes (4%), including several *Eubacterium* spp. (Fig. 1a).

Expanding our analysis of genomic investment to the module level, we inferred specific GMM–phylum associations (Supplementary Table 4). Polysaccharide (pectin and starch; both FDR < 1×10^{-3}) degradation was strongly associated with Bacteroidetes, while Firmicutes appeared limited to more readily fermentable substrates such as mono- and disaccharides. Proteobacteria were not only strongly associated with lipid degradation modules (including anaerobic fatty acid beta-oxidation and the glyoxylate bypass; Fisher FDR < 1×10^{-17}), but also with 15 amino-acid degradation modules (all FDR < 0.05). Additionally, our analyses confirmed some previously described specializations[18,19], including mucin degradation harboured by Bacteroidetes (FDR < 1×10^{-26}) and methanogenesis (FDR < 1×10^{-3}) being only encoded by the gut archaea *Methanobrevibacter smithii* and *Methanosphaera stadtmanae*, stressing the importance of these niche-specific species in gut microbial ecology. Together, our results represent a key step towards the drawing of a 'who-does-what' map of the metabolic potential of the colon microbiota.

Metabolic diversification as part of ecological strategy. Nutrient-rich environments such as the gut typically favour fast-growing bacteria, which are rapid, albeit inefficient, consumers of resources[20]. However, a particular aspect of the intestinal habitat is the continuous removal of nutrients along the digestive tract through digestion and absorption as well as fermentation. In the large intestine, progressive depletion of readily fermentable substrates (mostly saccharides) has been shown to affect bacterial ecology, causing a shift from saccharolytic to proteolytic fermentation from the proximal to distal colon[21], associated with a gradual drop in bacterial activity and turnover[22,23]. Here, when comparing the maximum growth rates of gut species predicted from genomic traits[20] (Supplementary Table 2) to their metabolic potential, we found that proteolytic specialists had significantly

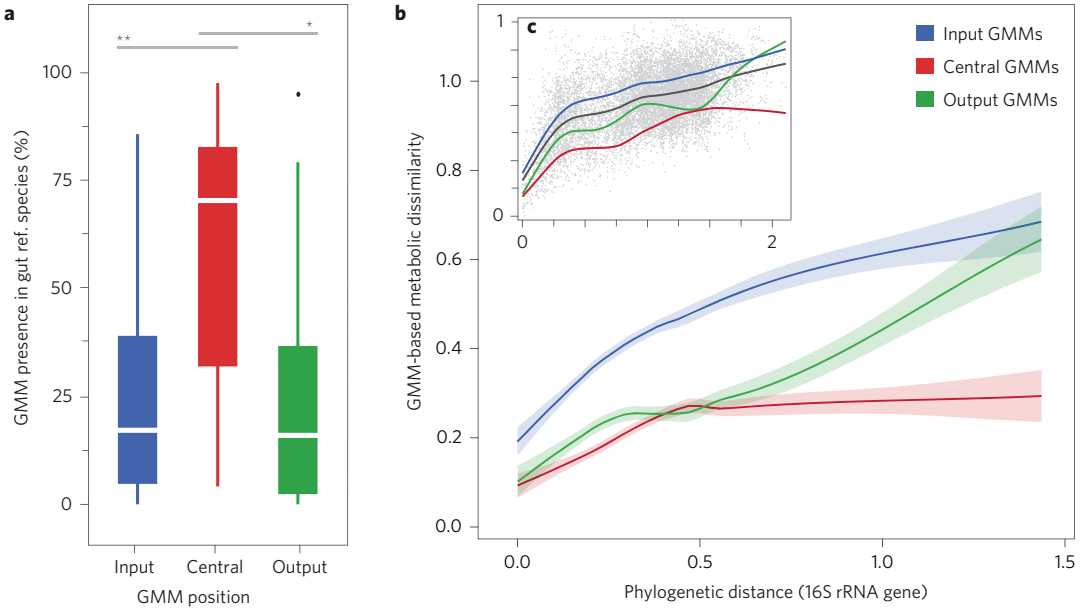

Figure 2 | Gut metabolic diversification of input, central and ouput functions. a, GMMs distribution in gut reference genomes (one representative per species; $N = 260$) according to their position in the GMM map. The inner line of the boxplots represents the median, the edges of the box are the third (Q3) and first (Q1) quartiles, and whiskers extend to Q3(Q1) +(−) 1.5 × interquartile range, or the last data point when below the computed ranges. **b**,**c**, Local polynomial regression fitting with 95% confidence interval of metabolic distance (Bray–Curtis on GMM abundances) by phylogenetic distance (16S rDNA) in gut reference genomes ($N = 260$) restricted to intra-genus comparisons (**b**) and expanded up to intra-phylum comparisons (**c**), where the black line in **c** represents fitting for all GMMs.

lower growth potentials than generalists and saccharolytic/lipolytic specialists (Fig. 1b,c; MWU $r = -0.37$, $P < 1 \times 10^{-8}$). Moreover, the inferred maximum growth rate correlated negatively with the proteolytic potential (Spearman's $\rho = -0.42$, $P = 1 \times 10^{-12}$). These results indicate that both the progressive depletion of readily fermentable carbohydrates along the intestinal tract and the related decrease in ecological pressure for fast growth create a niche for proteolytic bacteria. As end products of proteolytic fermentation have potentially detrimental effects on human health[10], the moderation of proteolysis through increased consumption of non-digestible, fermentable fibres has been targeted by prebiotic research[24]. Our analyses provide an ecological basis for such gut microbiota modulation strategies.

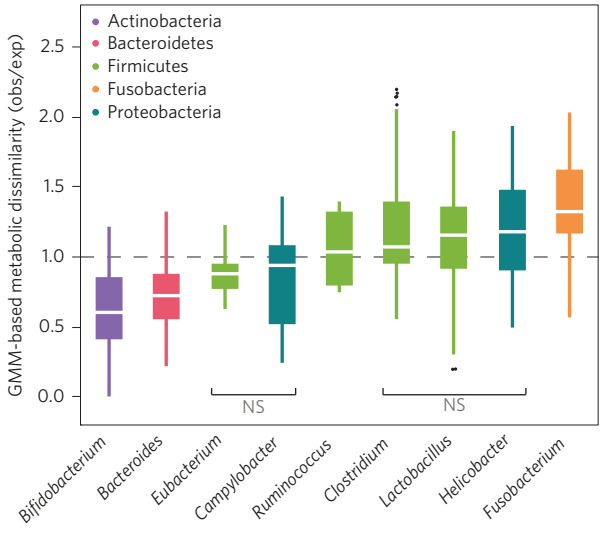

Figure 3 | Metabolic consistency between species belonging to the same genus. Dissimilarity in GMM profiles of pairs of conspecies of the human gut microbiota (deviation from expected from phylogenetic distance), in the nine genera with more than five sequenced species representatives (NS, non-significant MWU FDR > 0.05). The expected functional dissimilarity was determined by least-square linear fitting of the relation between phylogenetic distance (16S rRNA gene tree) and GMM dissimilarity. The inner line of the boxplots represents the median, the edges of the box are the third (Q3) and first (Q1) quartiles, and whiskers extend to Q3(Q1) +(−) 1.5 × interquartile range, or the last data point when below the computed ranges.

Input diversification drives gut metabolic network differentiation. Very few GMMs were ubiquitous in gut reference species. Fewer than 12% of GMMs were present in more than 70% of gut species, with a given module being present in 30% of genomes on average (Supplementary Table 5). Thus, although host-associated habitats are believed to display higher metabolic convergence than free-living habitats[25], these results indicated significant genomic diversification of the metabolic capacities of gut species.

The GMM distribution breadth across genomes correlated both with module position in the gut-specific metabolic map—central GMMs being more ubiquitous than input (MWU $r = 0.32$, FDR = 0.008) and output ($r = 0.46$, FDR = 0.022) modules (Fig. 2a)—and with community-level metabolic network centrality[26] (betweenness centrality index; Spearman's $\rho = 0.31$, $P = 0.004$; Supplementary Table 5). The former centrality definition (categorical) is part of the GMM description, splitting the first conversions steps of substrates (input) from central reactions and those leading to the end-products of bacterial fermentation processes (output), while the latter is a topology measure (betweenness centrality) of the previously published gut community-level network reconstruction[26]. Both these observations indicated that gut metabolic networks mainly differentiate through peripheral unit diversification. In particular, input GMMs were diversifying most at short phylogenetic distances (Fig. 2b), suggesting that substrate uptake specialization is the favoured method for functional diversification in closely related gut species.

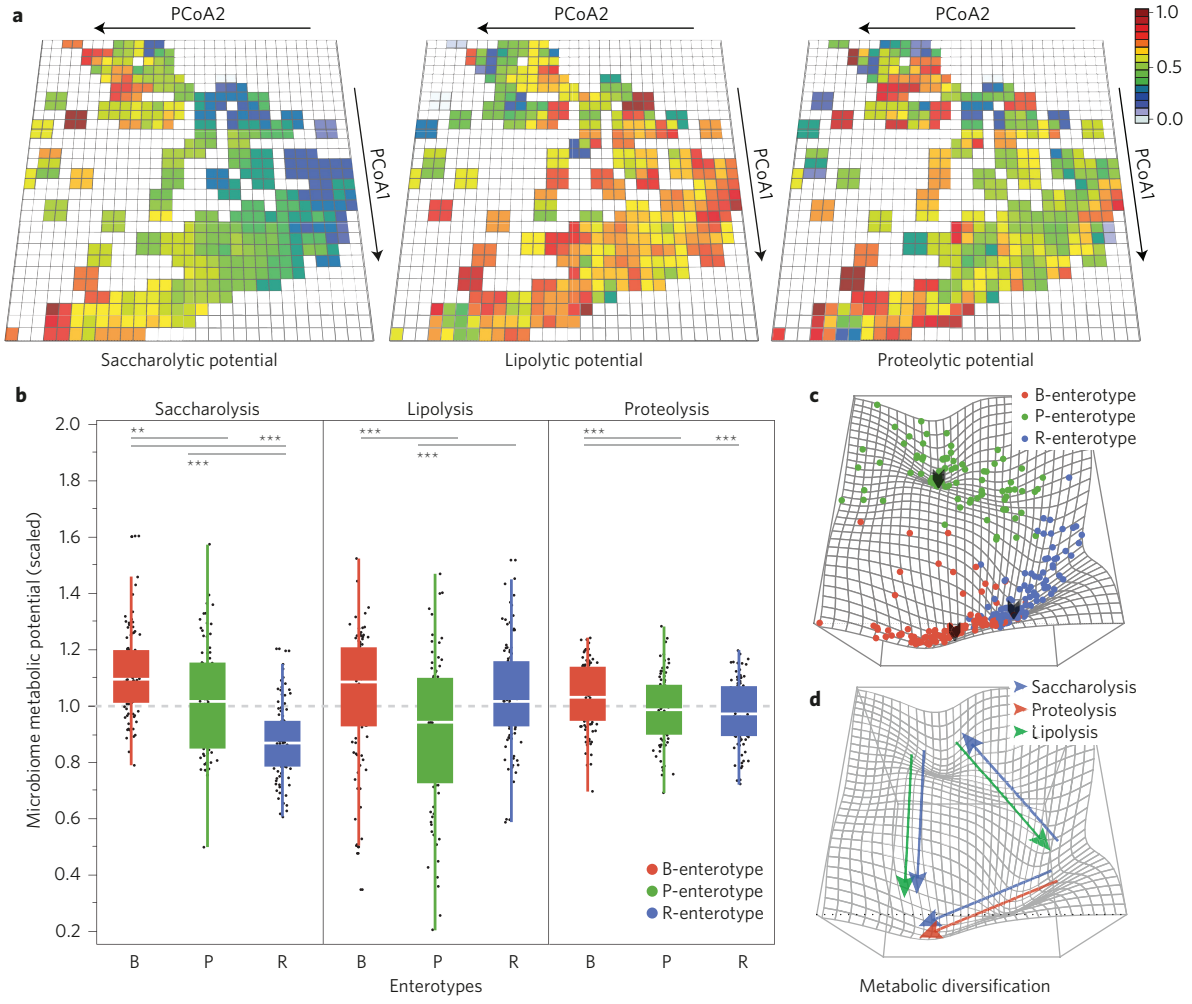

Figure 4 | Saccharolytic, lipolytic and proteolytic potential diversification across the enteroscape (277 MetaHIT samples). **a**, Enteroscape visualization of the three major fermentation potentials (scaled). **b**, Saccharolytic, lipolytic and proteolytic potential variation between the different enterotypes (*Bacteroides*, B; Ruminococcaceae, R; *Prevotella*, P)[27] (MWU FDR <0.01**, <0.001***; Supplementary Table 10). The inner line of the boxplots represents the median, the edges of the box are the third (Q3) and first (Q1) quartiles, and whiskers extend to Q3(Q1) +(−) 1.5 × interquartile range, or the last data point when below the computed ranges. **c**, Samples coloured by enterotype on the enteroscape, with enterotype centroids marked with black arrowheads. **d**, Summary of significant fermentation potential differences between the enterotypes, using a gradient approach (*Bacteroides*:Firmicutes, *Prevotella*:*Bacteroides* and *Prevotella*:Firmicutes ratios) with arrows representing the direction of increase of saccharolytic, lipolytic or proteolytic potential (Spearman's correlation with ratios: FDR < 0.05; Supplementary Table 10).

Genus-level metabolic consistency is not guaranteed. GMM diversification at short phylogenetic distances raises the question of to what extent human gut conspecies—that is, species belonging to the same genus—are metabolically similar and whether differences exist in the degree of diversification between genera. Here, we inferred important genus-level metabolic variation (Bray–Curtis dissimilarity), showing that metabolic consistency was not guaranteed among conspecies, even after correction for within-genus phylogenetic distances (Fig. 3). Functionally speaking, *Bacteroides* and *Bifidobacterium* were the most consistent genera, and Firmicutes genera had the lowest metabolic pairwise similarity. Hence, these results emphasize that extrapolations of metabolic potential based on phylogenetic assignments are not without risk, and relevant information regarding gut niche exploitation can be missed. This is especially important in the context of 16S rDNA-based community profiling, and underlines the necessity of uncertainty quantification when performing community functional predictions[8].

Saccharolytic, proteolytic and lipolytic potential diversify the gut ecological landscape. Using the public MetaHIT metagenomic data set of healthy lean, overweight and obese individuals[6] (Supplementary Table 6), we assessed nutrient trichotomy in microbiomes by exploring saccharolytic, proteolytic and lipolytic gradients in the gut ecological landscape (Fig. 4a,b and Supplementary Table 7). Faecal microbiome compositional variation was depicted as an 'enteroscape', the first plane of a normalized genus-level principal coordinates analysis (PCoA, Bray–Curtis dissimilarity; Fig. 4a). Next, we used this enteroscape backbone to vizualize functional diversity by colouring facets by the median value of a variable of interest (here saccharolytic, proteolytic and lipolytic potential) of the samples falling within a cell of the grid (Fig. 4a). To adress the debate over the discrete[27,28] or continuous[29] nature of human gut microbiota diversity, we performed subsequent analyses with both a stratification (Fig. 4b,c) and a gradients approach (Fig. 4d).

When visualizing the GMM-derived functional potential on the enteroscape, we observed that microbiome diversification across individuals appeared driven by fermentation substrate preference (Fig. 4a). Both lipolytic and proteolytic fermentation genomic potential were lower in the *Prevotella* (P) enterotype zone (Table 1 and

Table 1 | | Saccharolytic, proteolytic and lipolytic potential variation between enterotypes.

			Z	r (Z / √N)	FDR	Direction
Saccharolysis	P	B	−3.42	−0.26	0.0006	Higher in B than P
Saccharolysis	R	B	−9.57	−0.68	3.0×10^{-21}	Higher in B than R
Saccharolysis	R	P	−5.00	−0.37	8.50×10^{-7}	Higher in P than R
Proteolysis	P	B	−2.52	−0.19	0.0174	Higher in B than P
Proteolysis	R	B	−3.15	−0.22	0.0049	Higher in B than R
Proteolysis	R	P	−0.41	−0.03	0.6840	NS
Lipolysis	P	B	−3.50	−0.26	0.0014	Higher in B than P
Lipolysis	R	B	−0.78	−0.06	0.4332	NS
Lipolysis	R	P	3.25	0.24	0.0018	Higher in R than P

Data from 277 MetaHIT samples: *Bacteroides* (B), Ruminococcaceae (R) and *Prevotella* (P) enterotypes[27]. MWU test estimates (score, *Z*; effect size, *r*; significance after multiple testing correction, FDR). NS, non-significant.

Fig. 4b,c). The *Bacteroides* (B) enterotype was characterized by increased saccharolytic and proteolytic capacity (Table 1 and Fig. 4b). The distribution of individual modules over enterotypes was broadly consistent with these observations (Supplementary Fig. 3). Similar profiles were detected when applying a gradient approach[29] to microbiome diversity (Firmicutes:*Bacteroides*, *Prevotella*:Firmicutes and *Prevotella*:*Bacteroides* ratios; Supplementary Table 8). Nutrient-based diversification among individuals' gut microbiomes hints to diet or transit-driven ecosystem variation. Indeed, our results confirmed a previously described association of the P-enterotype with an agrarian diet[30]. Although the B-enterotype has been linked with high fat/high protein intake[30], our results also reflected the genomic investment in carbohydrate degradation of the dominating genus, referred to as the *Bacteroides* glycobiome[31]. Indeed, deconvolution of the functional shifts between enterotypes into genus-level contributions (Supplementary Fig. 4)[32] confirmed that the higher saccharolytic potential in the B- compared to R- and P-enterotypes was mainly driven by the *Bacteroides* genus, with an essential part of the community's saccharolytic potential being compensated by *Prevotella* in the P-enterotype. The increase in proteolytic:saccharolytic potential ratio that characterized the R-enterotype compared to the B-enterotype (MWU $r = 0.62$, FDR $= 1 \times 10^{-18}$; Supplementary Fig. 5) and P-enterotype ($r = 0.39$, FDR $= 1 \times 10^{-7}$) is in line with its reported association with slow transit/firm stool consistency[33,34]. Indeed, progressive depletion of readily fermentable substrates during prolonged transit would lead to an increased proteolytic:saccharolytic fermentation ratio[9].

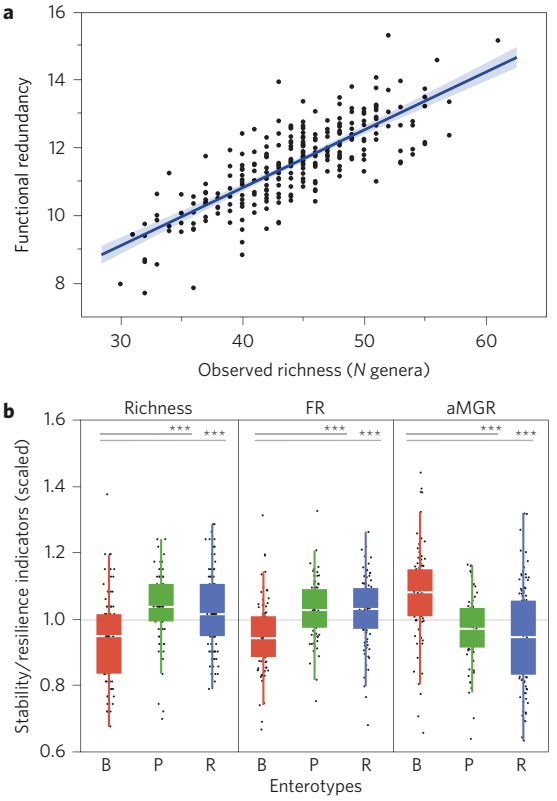

Figure 5 | Enterotype-associated differences in ecosystem resilience and stability indicators (277 MetaHIT samples). a, Linear least-squares fit between functional redundancy and phylogenetic richness. **b**, Distribution of phylogenetic richness, functional redundancy (FR) and sample average maximum growth rate (aMGR) in the enterotypes (*Bacteroides*, B; Ruminococcaceae, R; *Prevotella*, P)[27]. MWU FDR <0.001***. The inner line of the boxplots represents the median, the edges of the box are the third (Q3) and first (Q1) quartiles, and whiskers extend to Q3(Q1) +(−) 1.5 × interquartile range, or the last data point when below the computed ranges.

Functional redundancy gradients across the gut enteroscape. A recent 16S rRNA gene sequencing-based survey of temporal variation in the colon microbiota identified microbial diversity as a strong indicator of gut ecosystem temporal stability[35]. Moreover, lowered gut microbial diversity is frequently associated with dysbiosis[6]. These statements are in line with the diversity–stability hypothesis[36]. Yet, deviations from this ecological model are also observed and explained by the fact that diversity does not inherently reflect underlying resilience mechanisms such as species and function redundancy[36]. Here, we compared functional redundancy (FR; number of microbial taxa encoding a particular GMM present in a single sample) across the enteroscape to stability indicators such as species richness and the promotion of fast growth. The latter should reflect the recent history of perturbation of the ecosystem, which alters the r-versus K-strategists ratio, with r-selected organisms (faster growing but more wasteful consumers of resources) being favoured after disturbances[20]. Recently or frequently disturbed environments are thus expected to harbour communities with more fast growers. We assessed community growth rates by inferring community average maximum growth rates[20] (aMGR) and estimating actual growth rates at the time of sampling[37] (peak-to-trough ratio, PTR).

According to the insurance hypothesis in ecology[36,38], functional exclusivity (species having the monopoly of essential functions for the host or ecosystem) implies that environmental disturbances affecting these keystone species will have drastic effects on the ecosystem. Inversely, an increase in the number of different species carrying overlapping functional repertoires (increased FR) increases the habitat's resilience[38,39]. We found that FR was strongly correlated to genus richness (Spearman's $\rho = 0.77$, $P < 1 \times 10^{-55}$) and aMGR ($\rho = -0.22$, $P < 1 \times 10^{-4}$) in the MetaHIT samples[6] (Fig. 5a; Supplementary Fig. 6 and Supplementary Table 9). In addition, richness, aMGR and PTR were associated with enterotype classification (Fig. 5b, Supplementary Fig. 7 and Supplementary

Table 10). B-enterotype samples not only exhibited strikingly lower FR values than the R-enterotype (MWU $r = 0.40$, FDR $< 1 \times 10^{-8}$; Fig. 5b) and P-enterotype (MWU $r = 0.41$, FDR $< 1 \times 10^{-7}$), the FR of B-enterotype samples was also lower than would be expected by their genus richness (t-test on residuals mean < 0: P-value = 0.02). These observations shed an interesting light on the association of low-richness and/or B-enterotype communities with (low-grade) inflammation[6,40]. In these cases, FR could serve as a quantifier of pathology-associated ecosystem dysbiosis[38]. Also here, previously observed reduced passage rates in R-individuals[33] suggest an ecological mechanism underlying this observation: increased transit times not only allow proliferation of slower growing species, but progressive depletion of water availability also induces niche diversification and, hence, phylogenetic diversification and potentially increased functional redundancy.

Conclusions

Microbial niche differentiation and host pressure for functional redundancy have been put forward as the main selective forces shaping the colon ecosystem. Here, we demonstrate that host–microbiota coexistence indeed results in a delicate balance between microbial individual functional generalism and diversification and community redundancy. Analysis of the phylogenetic distribution of functions reveals a duality that is independent of taxonomic boundaries: although some traits are highly associated with (and over-represented within) particular phyla, niche specialization occurs even between closely related taxa. Core metabolic repertoire conservation over distant taxa contributes to gut microbiota resilience by assuring a high degree of ecosystem functional redundancy. The availability of fermentable substrates to the microbiota clearly acts as a diversifier of the enteroscape, with associated differences in functional redundancy. Although substrate availability can reflect diet, our results also suggest a diversifying impact of the depletion of readily fermentable energy sources due to reduced passage rates. In terms of community classification, this summarizes as the R-enterotype corresponding to highly proteolytic, more K-selected and functionally redundant communities, and the B-counterparts being characterized by increased saccharolytic potential, r-selection and a low degree of redundancy, potentially making them more vulnerable to perturbations.

Methods

Module set. GMMs reflect bacterial and archaeal metabolism specific to the human gut environment, with a focus on anaerobic fermentation processes. The current set of 103 GMMs (Supplementary Data 1) was built through an extensive review of the literature and metabolic databases, including MetaCyc (ref. 11) and KEGG (ref. 12), followed by expert curation and delineation of modules and alternative pathways. Modules describe enzymatic processes annotated using exclusively prokaryotic and archaeal KEGG Orthology (KO). We chose to use KO annotation to represent enzymatic functions due to their widespread availability in genome annotation (for example, IMG and Uniprot). As a single enzymatic activity can be encoded by several orthologues, GMMs include variants of a similar metabolic conversion with in most cases identical input/output compounds. To integrate such variants into the module structure, alternative options are tab-separated, while return- and comma-separated KOs are all required for process completeness. Given the nature of the modules and the sometimes closely related metabolic conversions described, overlap between modules is inevitable but limited (Supplementary Table 11). GMMs are classified according to their position in the gut metabolic map as input, central and output modules ($N = 75$, 11 and 17, respectively). Two higher functional hierarchical levels were created, grouping GMMs into 10 metabolic categories and 30 subcategories (Supplementary Data 1). GMM descriptions and metabolic diagrams representing the GMM network and metabolic (sub-)categories are available at http://www.raeslab.org/companion/gmms and are free to download and use as a resource for bioinformatics pipelines. GMMs will be updated as soon as new modules are available.

Reference genomes. All available bacterial and archaeal genomes were downloaded from img/hmp v4.0 (ref. 41), from which 562 KO-annotated genomes categorized as 'Gastrointestinal tract' were used in this study. Gut reference genomes encompass 10 phyla, 91 genera and 260 species (Supplementary Table 3). For most analyses, one randomly picked representative genome per species was used to reduce the culturing and sequencing bias for species of interest, such as *Escherichia coli*. The genomes' full lineage descriptions were retrieved from NCBI Taxonomy[42] (ftp://ftp.ncbi.nih.gov/pub/taxonomy/).

Faecal metagenomes. Healthy human faecal metagenomes from the MetaHIT[6] project ($N = 277$), functionally mapped to the reference human gut microbial gene catalogue, with KO annotation[40] and original gene taxonomic assignment (BLASTP e-value of 1×10^{-5} to NCBI-NR), were used for metabolic pathway reconstruction (GMM threshold coverage = 2/3; previously, a threshold of 30% coverage after removal of overly 'promiscuous' enzymes (that is, present in multiple modules) has been suggested[6]), binning genes by genus. Sample enterotyping was performed as described in the original publication[27] (http://enterotype.embl.de/; Supplementary Table 7), using mOTU-derived genus abundance matrix[43,44].

Module detection. The coverage of each metabolic variant encoded in a GMM was calculated as the number of steps for which the enzyme (KO) was found in a genome or metagenome, divided by the total number of steps constituting the variant. The coverage of the GMM was defined as being equal to that of the variant with maximum coverage. GMM presence/absence was identified with a detection threshold of more than 66% coverage to provide tolerance to miss-annotations and missing data in incomplete (draft) genomes or metagenomes (Supplementary Table 1). GMM abundance was calculated as the median of KO abundance in the pathway with maximum coverage. Genomic investment was defined as the sum of the abundances of GMMs associated with a metabolic trait divided by the number of GMMs in that group (Supplementary Table 2). Accordingly, for triplot representation, the genomic investment in proteolytic (modules hierarchically grouped as 'amino acid degradation'), saccharolytic ('carbohydrate degradation') and lipolytic ('lipid degradation') fermentation was divided by the number of GMMs in each category ($N = 36$, 22 and 6, respectively). The number of modules involved in lipolytic fermentation was relatively small, both due to the fact that anaerobic fermentation of lipids is rather uncommon and the (related) observation that the genetic background of oxidation of fatty acids under anaerobic conditions is not well characterized[45].

Metabolic consistency. Functional dissimilarity was defined for a given pair of taxa by the Sørensen index for GMM presences and Bray–Curtis dissimilarity on GMM abundances. Expected functional dissimilarity (Y) was determined by linear least-squares fitting of the relation between pairwise phylogenetic distance (X; calculated from the phylogenetic tree provided in Supplementary Data 2) and pairwise functional dissimilarity (Supplementary Fig. 8) at the genus level: $Y = 0.23 + 0.3105X$. Metabolic consistency at the genus level was only calculated for genera with more than five species sequenced. A similar analysis was performed at the species level for species with a minimum of six strains (Supplementary Fig. 9).

Module centrality index. For each GMM, a centrality index was calculated based on the previously published KO betweenness centrality measures of a published community-level gut metabolic network reconstruction, which were provided by the authors of ref. 26 (Supplementary Table 5). These were calculated by the authors as the proportion of shortest paths in a complex network that pass through a given node, such that high betweenness centrality is associated with nodes located in the core of the network, and low being associated with more peripheral nodes. For each GMM, the centrality of each alternative pathway was calculated as the average betweenness of the KOs involved. The maximum pathway centrality was kept as representative of the GMM. The betweenness of 134 of 442 KOs in the GMMs was uncharacterized, resulting in 19 of the 103 modules with unassigned centrality index. As expected, input/output GMMs tend to be more peripheral (median = 0.004), and central GMMs display a higher median centrality index (median = 0.023).

Ecological strategy. For every reference genome, transporters and transcription factors richness were defined as the total number of genes in KEGG BRITE classes ko02000 (excluding Eukaryotic type – ABC transporters) and ko03000 (excluding Eukaryotic type), respectively (Supplementary Table 3). Maximum growth rates (1/minimum generation time (MGT^{-1})) were inferred from genomes' codon usage bias (Supplementary Table 3), which is associated with minimal generation times ($R^2 = 0.59$ in mesophylic organisms), as described in ref. 20. Similarly, for the MetaHIT data set ($N = 277$), the metagenomic sample aMGR was inferred from codon usage bias. The technique applied to metagenomic samples results in an estimation of the average of the MGRs of the organisms in the sample, weighted by their abundances[20]. As a complement, an estimator of species growth rates from metagenomic data was used (estimations for MetaHIT samples were provided by the authors of the method described in ref. 37; Supplementary Table 7), where the ratio of sequencing coverage between the peak and trough (PTR) provides a quantitative measure of a species' growth rate at the time of sampling, for species with a representative genome sequenced.

Phylogenetic tree. 16S rDNA sequences were retrieved for the 391 reference genomes from the Silva database[46] (www.arb-silva.de/). A total of 735 rDNA sequences were added to increase taxon sampling within each phylum and avoid

long-branch attraction artefacts. Sequences were aligned with MUSCLE[47]. Poorly aligned regions were removed using Gblocks[48] with non-stringent parameters (ten non-conserved contiguous positions/minimum block size of eight/gaps allowed). A deep-phylogenetic tree was reconstructed with RAXML (model GTR+8Γ+I, 100 rapid bootstraps; Supplementary Section 2), from which pairwise phylogenetic distances were retrieved (R package 'ape').

Functional redundancy. Functional redundancy was defined for each sample of the MetaHIT data set ($N = 277$) as the median GMM functional redundancy, being the number of genera encoding each GMM in the sample.

Enteroscape. The gut ecological landscape or 'enteroscape' (universe of sampled human microbiota constellations) was visualized as the first plane of the PCoA on normalized genus abundances of MetaHIT samples (Bray–Curtis distance, R package Vegan). Facets of the enteroscape grid were coloured according to selected variables (metabolic potentials) with smoothing restricted to immediate neighbour facets (average).

Decomposition of taxa contribution to metabolic shifts. Deconvolution of the functional shifts in saccharolytic, proteolytic and lipolytic potential between the three enterotypes was performed using FishTaco[32], a permutation-based approach to decompose functional shifts observed in a comparative analysis into individual taxon-level contributions. FishTaco (version 1.0.5) was used with default parameters, with as input the functional matrix corresponding to total saccharolytic, proteolytic and lypolitic potential in the MetaHIT samples and the genus abundance matrix for the same samples. The –inf option was used to infer the genomic content of taxa from the data.

Statistical analyses. Non-parametric testing was used because it is distribution-free and robust to outliers. The non-parametric Mann–Whitney *U* test (MWU) was used for categorical variables (group comparisons; Figs 1–5) and Spearman correlations for continuous variables (Fig. 2). The one-tailed Fisher's exact test was used for the analysis of contingency tables (over-representation of specific modules in certain taxa compared to all others). Correction for false discovery rate (Benjamini–Hochberg (FDR)) was applied whenever multiple comparisons were executed. Statistical analyses were performed in R.

Accession codes. Gut reference genomes can be downloaded from the JGI Genome Portal with the taxon_oid provided in Supplementary Table 3. The MetaHIT project samples can be downloaded from EBI with accession numbers ERA000116 and ERP003612.

Computer code. The computer code used to compute GMM abundances from a KO abundance table is shared on GitHub (https://github.com/raeslab/GMMs).

Received 17 August 2015; accepted 9 May 2016; published 13 June 2016

References

1. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124,** 837–848 (2006).
2. Rajilić-Stojanović, M. & de Vos, W. M. The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol. Rev.* **38,** 996–1047 (2014).
3. Tremaroli, V. & Bäckhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* **489,** 242–249 (2012).
4. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).
5. Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nature Rev. Microbiol.* **10,** 538–550 (2012).
6. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500,** 541–546 (2013).
7. Levy, R. & Borenstein, E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc. Natl Acad. Sci. USA* **110,** 12804–12809 (2013).
8. Langille, M. G. I. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnol.* **31,** 814–821 (2013).
9. Cummings, J. H. & Macfarlane, G. T. The control and consequences of bacterial fermentation in the human colon. *J. Appl. Bacteriol.* **70,** 443–459 (1991).
10. Hughes, R., Magee, E. A. & Bingham, S. Protein degradation in the large intestine: relevance to colorectal cancer. *Curr. Issues Intest. Microbiol.* **1,** 51–58 (2000).
11. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **42,** D459–D471 (2014).
12. Kanehisa, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* **42,** D199–D205 (2014).
13. Darzi, Y., Falony, G., Vieira-Silva, S. & Raes, J. Towards biome-specific analysis of meta-omics data. *ISME J.* **10,** 1025–1028 (2016).
14. Koch, A. L. Oligotrophs versus copiotrophs. *Bioessays* **23,** 657–661 (2001).
15. Pokusaeva, K., Fitzgerald, G. F. & van Sinderen, D. Carbohydrate metabolism in Bifidobacteria. *Genes Nutr.* **6,** 285–306 (2011).
16. Thomas, F., Hehemann, J.-H., Rebuffet, E., Czjzek, M. & Michel, G. Environmental and gut bacteroidetes: the food connection. *Front. Microbiol.* **2,** 93 (2011).
17. Tyrrell, K. L., Warren, Y. A., Citron, D. M. & Goldstein, E. J. C. Re-assessment of phenotypic identifications of Bacteroides putredinis to Alistipes species using molecular methods. *Anaerobe* **17,** 130–134 (2011).
18. Carbonero, F., Benefiel, A. C. & Gaskins, H. R. Contributions of the microbial hydrogen economy to colonic homeostasis. *Nature Rev. Gastroenterol. Hepatol.* **9,** 504–518 (2012).
19. Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P. & Forano, E. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **3,** 289–306 (2014).
20. Vieira-Silva, S. & Rocha, E. P. C. The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genet.* **6,** e1000808 (2010).
21. Macfarlane, G. T., Gibson, G. R. & Cummings, J. H. Comparison of fermentation reactions in different regions of the human colon. *J. Appl. Bacteriol.* **72,** 57–64 (1992).
22. Macfarlane, S. & Macfarlane, G. T. Regulation of short-chain fatty acid production. *Proc. Nutr. Soc.* **62,** 67–72 (2003).
23. Macfarlane, S., Quigley, M., Hopkins, M., Newton, D. F. & Macfarlane, G. Polysaccharide degradation by human intestinal bacteria during growth under multi-substrate limiting conditions in a three-stage continuous culture system. *FEMS Microbiol. Ecol.* **26,** 231–243 (1998).
24. De Preter, V. *et al.* The prebiotic, oligofructose-enriched inulin modulates the faecal metabolite profile: an *in vitro* analysis. *Mol. Nutr. Food Res.* **54,** 1791–1801 (2010).
25. Zaneveld, J. R., Lozupone, C., Gordon, J. I. & Knight, R. Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives. *Nucleic Acids Res.* **38,** 3869–3879 (2010).
26. Greenblum, S., Turnbaugh, P. J. & Borenstein, E. 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 (2012).
27. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473,** 174–180 (2011).
28. Holmes, I., Harris, K. & Quince, C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* **7,** e30126 (2012).
29. Jeffery, I. B., Claesson, M. J., O'Toole, P. W. & Shanahan, F. Categorization of the gut microbiota: enterotypes or gradients? *Nature Rev. Microbiol.* **10,** 591–592 (2012).
30. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334,** 105–108 (2011).
31. Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host–bacterial mutualism in the human intestine. *Science* **307,** 1915–1920 (2005).
32. Manor, O. & Borenstein, E. Systematic characterization and analysis of taxonomic drivers of functional shifts in the human microbiome (2016); http//elbo.gs.washington.edu/software_fishtaco.html
33. Vandeputte, D. *et al.* Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut* **65,** 57–62 (2016).
34. Gorkiewicz, G. *et al.* Alterations in the colonic microbiota in response to osmotic diarrhea. *PLoS One* **8,** e55817 (2013).
35. Flores, G. E. *et al.* Temporal variability is a personalized feature of the human microbiome. *Genome Biol.* **15,** 531 (2014).
36. McCann, K. S. The diversity–stability debate. *Nature* **405,** 228–233 (2000).
37. Korem, T. *et al.* Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349,** 1101–1106 (2015).
38. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489,** 220–230 (2012).
39. Relman, D. A. The human microbiome: ecosystem resilience and health. *Nutr. Rev.* **70**(Suppl 1), S2–S9 (2012).
40. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
41. Markowitz, V. M. *et al.* The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. *Nucleic Acids Res.* **36,** D528–D533 (2008).
42. Federhen, S. The NCBI Taxonomy database. *Nucleic Acids Res.* **40,** D136–D143 (2012).
43. Forslund, K. *et al.* Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528,** 262–266 (2015).
44. Sunagawa, S. *et al.* Metagenomic species profiling using universal phylogenetic marker genes. *Nature Methods* **10,** 1196–1199 (2013).
45. Campbell, J. W., Morgan-Kiss, R. M. & Cronan, J. E. A new *Escherichia coli* metabolic competency: growth on fatty acids by a novel anaerobic beta-oxidation pathway. *Mol. Microbiol.* **47,** 793–805 (2003).

46. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41,** D590–D596 (2013).
47. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–1797 (2004).
48. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56,** 564–577 (2007).

Acknowledgements

This research was funded by FP7 METACARDIS HEALTH-F4-2012- 305312 and IWT-SBO 100016. S.V.S. is funded by Marie Curie Actions FP7 People COFUND Proposal 267139 (acronym OMICS@VIB) and the Fund for Scientific Research-Flanders (FWO-V). S.C., M.V.C. and G.L.M. are funded by FWO-V. D.V. is funded by the Agency for Innovation by Science and Technology (IWT). The authors thank Dr Greenblum and collaborators and Dr Korem and collaborators for providing data.

Author contributions

S.V.S. and G.F. contributed equally to this work. S.V.S. and G.F. designed the study, collected and processed data, performed the analyses and wrote the paper. G.F., S.V.S., D.V., M.V.C. and Y.D. curated the gut-specific metabolic modules. J.R. designed the study and co-wrote the paper. Y.D., R.G.Y., F.H. and S.O. participated in data collection and processing. G.L.M., Y.D., R.G.Y., S.O., D.V. and S.C. participated in statistical analyses. Y.D. participated in graphical representation design.

Additional information

Supplementary information is available online. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.R.

Competing interests

The authors declare no competing financial interests.