Environmental Microbiology – Research Article

Gut microbial functional maturation and succession during human early life

Tomás Cerdó*[1,2], Alicia Ruiz*[2,3], Inmaculada Acuña[3], Ruy Jáuregui[4], Nico Jehmlich[5], Sven-Bastian Haange[5], Martin von Bergen [5], Antonio Suárez[3] and Cristina Campoy[1,2,6]

[1]Department of Paediatrics, School of Medicine, University of Granada, Granada, Spain.

[2]EURISTIKOS Excellence Centre for Paediatric Research, Biomedical Research Centre, University of Granada, Granada, Spain.

[3]Department of Biochemistry and Molecular Biology 2, Biomedical Research Centre, University of Granada, Spain.

[4]AgResearch Grasslands, Tennent Drive, Private Bag 11008, Palmerston North, New Zealand.

[5]Department of Molecular System Biology, Helmholtz Centre for Environmental Research- UFZ, Permoserstraße 15, Leipzig, Germany

[6]Spanish Network of Biomedical Research in Epidemiology and Public Health (CIBERESP), Carlos III Institute, Granada node, Spain.

*T.C and A.R contributed equally to this work.

Corresponding author:

Dr.Antonio Suárez

Department of Biochemistry and Molecular Biology 2, Biomedical Research Centre, University of Granada, Spain.

Email: asuarez@ugr.es Phone: +34-958-243833/+34-958-241000 ext.20318

Running title

Gut microbial functionality during early life

ORIGINALITY-SIGNIFICANCE STATEMENT

A major challenge in microbial ecology is to identify its functional members and understand how their functional and phylogenetic dynamics ultimately influence human physiology and health. Critically important are the initial stages of microbiota colonization and maturation in the gut because early dysbiosis has been shown to affect human health later in life. Though a high number of papers have described community membership in gut microbiota, few studies have attempted to provide insight into the dynamics of functional and metabolic processes associated to gut microbial evolution and maturation. In this study, we reasoned that a metaproteome analysis would provide biological information on the relative importance of gut microbial taxa in ecosystem functioning, their collective functional pattern and the network topology in relation to host physiology during human early life.

The new findings are the following:

First, we assessed the metaproteomes of 56 infants from 6- and 18-months of age, where the transition from weaning to solid food consumption occurs, to obtain an extensive catalogue of 9,173 bacterial proteins groups compressed into 1,117 COG functions that were assigned to 134 genera. *Bifidobacterium* was the genus to which the highest number of distinct protein groups were assigned.

Second, we observed a high number of significant differences between 16S rRNA gene sequences (community membership) and origin of peptides (biologically active members) at all taxa levels.

Third, age had a major impact on early gut microbiota assembly and function.

Fourth, the functional community was more similar among individuals than the total community.

Fifth, the early gut microbiota acquired a significant capacity to transport bicarbonate, ion metals, amino acids and inorganic oxides.

Sixth, our results underscore the role of host and dietary glycan degradation, central carbon metabolism and short chain fatty acid fermentation in the progression to a mature profile in the gut microbiota, providing insights into the metabolic strategies of gut microbiota taxa.

Seventh, our results show that the maturation of the gut microbiota is a non-random process where two mutually exclusive modules of functional families, built around *Bifidobacteriaceae* and *Lachnospiraceae* respectively, metabolically succeeded each other.

Our results contribute to this field because there is a need to understand the functional maturation of the gut microbiota which may constitute an important research tool for indicators of future healthy or diseases states and for the design of microbiota-targeted health-promoting strategies early in life.

SUMMARY

The evolutional trajectory of gut microbial colonization from birth has been shown to prime for health later in life. Here, we combined cultivation-independent 16S rRNA gene sequencing and metaproteomics to investigate the functional maturation of gut microbiota in faecal samples from full-term healthy infants collected at 6 and 18 months of age. Phylogenetic analysis of the metaproteomes showed that *Bifidobacterium* provided the highest number of distinct protein groups. Considerable divergences between taxa abundance and protein phylogeny were observed at all taxonomic ranks. Age had a profound effect on early microbiota where compositional and functional diversity of less dissimilar communities increased with time. Comparisons of the relative abundances of proteins revealed the transition of taxon-associated saccharolytic and fermentation strategies from milk and mucin-derived monosaccharide catabolism feeding acetate/propanoate synthesis to complex food-derived hexoses fuelling butanoate production. Furthermore, co-occurrence network analysis uncovered two anti-correlated modules of functional taxa. A low-connected *Bifidobacteriaceae*-centred guild of facultative anaerobes was succeeded by a rich club of obligate anaerobes densely interconnected around *Lachnospiraceae*, underpinning their pivotal roles in microbial ecosystem assemblies. Our findings establish a framework to visualize whole microbial community metabolism and ecosystem succession dynamics, proposing opportunities for microbiota-targeted health-promoting strategies early in life.

INTRODUCTION

The human gut is a bioreactor with a microbiota that consists of hundreds or thousands of bacterial species-level taxa, dominated by the phyla *Firmicutes* and *Bacteroidetes* with the less abundant phyla *Proteobacteria*, *Actinobacteria* and *Verrumicrobia* (Harmsen and de Goffau, 2016). The composition of the gut microbiota is influenced by genetic and environmental factors starting early in life (Charbonneau et al., 2016; Milani et al., 2017). Considerable efforts have focused on cataloguing the composition of the infants' gut microbiota (Palmer et al., 2007; Qin et al., 2010; Gosalbes et al., 2011; Koenig et al., 2011; Franzosa et al., 2014; Valles et al., 2014; Backhed et al., 2015; Asnicar et al., 2017; Cerdó et al., 2017) because the evolutional trajectory of gut microbiota from birth has been shown to prime for health later in life (Cryan and Dinan, 2012; Cox et al., 2014; Cerdó et al., 2016; Tamburini et al., 2016). These studies have shown that the initial microbiota evolves over time, increasing diversity and adapting to the anaerobic environment and nutrient availability.

However, current knowledge on infants' gut microbiota has been almost exclusively obtained from 16S rRNA gene sequencing, metagenomics and metatranscriptomics analyses. A major limitation of these DNA-based studies is that they infer potential functions, providing limited insight into the metabolic landscape and dynamic interplay of the gut microbiota. Because protein abundance is a reflection of specific microbial activities in a given ecosystem, metaproteomics exploits the power of high performance mass spectrometry (MS) to simultaneously address composition and function in microbial communities (Hettich et al., 2013). So far, three intestinal metaproteomics studies on preterm neonates and one on term infants have been carried out (Klaassens et al., 2007; Brooks et al., 2015; Young et al.,

2015; Zwittink et al., 2017), leaving the dynamics of microbial functional maturation during early life largely unexplored. Here, we combined cultivation-independent 16S rRNA gene sequencing and metaproteomics to investigate the functional differentiation of gut microbiota in faecal samples from full-term healthy infants collected at 6 and 18 months of age. We used the metaproteomics data in a systematic comparative strategy to provide direct evidence of active microbial taxa, functional signatures and topological architecture of gut microbial interactions characteristic to each chronological state.

RESULTS AND DISCUSSION

Considering the metaproteome analysis as indicators of current metabolic activity and physiological adaptation provides microbial ecologists with a robust framework, facilitating a closer understanding of the complex dynamics that drive ecosystem functional and compositional responses to environmental pressures (Hettich et al., 2013). Our hypothesis is that the metaproteome provides insight about the relative importance of its members in ecosystem functioning, their collective functional pattern and the network topology in relation to host physiology. Our goal was to address those questions during early human life.

Taxonomic profiling of gut microbiota and their proteins

We collected faecal samples from healthy infants of 6- and 18-months of age (Supporting information Table S1) to characterize the gut microbial composition by 16S rRNA gene sequencing and the expressed proteins by shotgun metaproteomics. The phylogenetic composition and categorical breakdown of identified OTUs and proteins in our samples are presented in the supporting material (Supporting information Table S2). After quality filtering, 7,890,853 read sequences rendered a gut microbial profile consisting of 679 species-level bacterial operational taxonomic units (OTUs) that narrowed to 89 distinct genera belonging to 40 families after high confidence phylogenetic annotation (Fig 1A). In total, 11,901 peptides were identified, of which 9,173 bacterial protein groups were assigned and unambiguously quantified. These protein groups were assigned to 134 genera belonging to 61 families (Fig 1A). This is the highest number of distinct proteins groups identified in human gut metaproteomics studies published so far, but indicate the high variability thorough the investigated cohorts. Moreover, taxonomic analysis of metaproteomics data showed that protein coverage and abundances within each phylum were highly diverse

(Supporting information Fig. S1). In *Actinobacteria*, *Bifidobacterium* accounted for most protein group abundances. Notably, the greatest number of distinct protein groups annotated to a genus belonged to *Bifidobacterium*, which emphasizes its functional significance in the gut during early life (Charbonneau et al., 2016). *Akkermansia* within *Verrucomicrobia* accounted for 2.7% of total protein groups. *Parabacteroides, Prevotella* and *Alistipes* had the highest number of protein groups in *Bacteroidetes*. We were able to identify protein groups across sixty-six different genera in *Firmicutes* and forty-five genera in *Proteobacteria*. Bins of groups that represented more than 50% of proteins in *Firmicutes* were assigned to *Faecalibacterium*, *Ruminococcus*, *Veillonella*, *Roseburia* and *Eubacterium*. In *Proteobacteria*, *Pseudomonas*, *Parasuterella*, *Haemophilus*, *Bilophila*, *Escherichia*, and *Desulfovibrio* were the highest contributors to protein group abundances. Our results are in agreement and further extend previous reports on the phylogenetic diversity of microbial proteins within the human gut (Klaassens et al., 2007; Verberkmoes et al., 2009; Rooijers et al., 2011; Kolmeder et al., 2012; Ferrer et al., 2013; Perez-Cobas et al., 2013; Juste et al., 2014; Brooks et al., 2015; Kolmeder et al., 2016; Tanca et al., 2017; Zwittink et al., 2017).

Comparison of metaproteomics and compositional data derived from 16S rRNA analysis

Kolmeder In human gut microbial ecology, a major challenge is to identify its active members whose response to environmental factors or disease-induced dysbiosis ultimately influence host homeostasis (Mao and Franke, 2015). Conventional DNA-based approaches inform about gene content and metabolic potential but do not inform about biological activity since all microbial DNA will be sequenced (Cangelosi and Meschke, 2014). Metaproteomics offers large-scale functional and

phylogenetic profiling of expressed proteins and, thus, a feasible approximation to characterize biological activity within microbial ecosystems (Hettich et al., 2013). To estimate the biological activity of the members of the gut microbial community, we compared the log ratio of abundances between organism-origin of protein groups and 16S rRNA gene abundances. Though previous studies reported good correlations between 16S rRNA gene abundances and microbial source of protein groups (Rooijers et al., 2011; Kolmeder et al., 2016), our analysis yielded a high number of deviations even at upper taxonomic ranks (Fig. 1B), suggesting discordance between microbial membership and biological activity. The deviations were in a many cases of an order-of-magnitude (highly significant absolute fold change >10) either lower or higher than expected from 16S rRNA gene abundances of their corresponding taxa (Supporting Information Table S3). The most significant deviation was the high relative proportion of protein groups identified for *Verrucomicrobia* versus its null detection by 16S rRNA gene sequencing. *Actinobacteria* showed a very high log ratio with *Bifidobacterium* accounting for high protein abundance, suggesting a high activity of this taxon in early life microbiota. Despite *Bacteroides* was the most abundant genus, most protein groups were assigned to *Parabacteroides*, *Prevotella* and *Alistipes*. Significant differences in the ratios for *Lachnospiraceae* and *Ruminococcaceae* in *Firmicutes* were also observed. The strongest negative ratios were observed in *Streptococcus*, *Veillonella*, *Enterococcus* and *Blautia* whereas the positive ratios observed for *Faecalibacterium, Flavonifractor* and *Oscillibater* suggested high biological activity. Thus, our metaproteomics data revealed that the gut microbiota harbours a distinctive subset of biologically active microorganisms as consistently shown in other reports (McNulty et al., 2011; Ferrer et al., 2013; Maurice et al., 2013; Maurice and Turnbaugh, 2013; Perez-Cobas et al., 2013). While we cannot exclude

analytical biases in DNA and protein extraction methods, these discordances suggest that using bacterial taxa as input information to build predictive theoretical models of microbial activity and contribution to community functioning in human gut microbial ecosystems may be highly misleading.

Temporal patterns of qualitative diversity in the gut microbiota

We initially calculated the proportion of variance (coefficient of determination, R^2) in microbiota composition and function that was explained by study variables and individuals (Supporting information Table S4A-B). In microbiota composition, age and individuality explained 7% and 11% of the total variation in agreement with previous reports (Zoetendal et al., 2011; Goodrich et al., 2014; Salonen et al., 2014; Backhed et al., 2015; Falony et al., 2016). Mode of delivery or pre-pregnancy mother's body mass index did not influence microbial composition in our dataset. Age was the single significant variable explaining a relevant proportion (13.5%) of total variation in microbiota function. The unprecedented impact of age upon microbiota function is remarkable, given the high variation assigned to individuality and/or mode of delivery consistently reported in compositional studies of gut microbiota. β-diversity metrics of total (phylogeny of OTUs) and functional (phylogeny of proteins) gut microbial communities confirmed that samples clustered by age (Fig. 2A-C). We observed increased α-diversity but reduced β-diversity as a function of time, suggesting that both total and functional communities accumulated diversity into less heterogeneous structures (Fig. 2D). Despite taxonomic divergence, Bray-Curtis dissimilarity metrics showed that the functional community was increasingly more conserved among infants than the total community (Fig. 2E). This result is in line with the hypothesis suggesting that gut microbiota is assembled around a between-subject more conserved consortium of biologically active

microorganisms (Turnbaugh et al., 2009; Burke et al., 2011; Consortium, 2012; Franzosa et al., 2014; Moya and Ferrer, 2016; Ruiz et al., 2017). We used LEfSe, a tool for metagenomic biomarker discovery (Segata et al., 2011) to further explore age-associated shifts in total and functional gut microbiota. Many taxa at multiple phylogenetic depths were found at significantly different relative abundances between time points (Fig. 2F-G and Supporting information Table S5). *Firmicutes* dominated the total community and its functional subset that enriched in *Bacteroidetes* and *Firmicutes* and depleted in *Proteobacteria* and *Actinobacteria* with time. In total gut microbiota, signature highly abundant genera (>1% mean relative abundance) at 6-months were *Enterococcus*, *Lactobacillus*, *Erysipelotrichaceae_Incertae_Sedis*, *unclassified_Veillonellaceae* and *unclassified_Enterobacteriaceae* while the 18-month's total gut microbiota was significantly enriched in obligate anaerobes from the genera *Bacteroides*, *Anaerostipes*, *Blautia*, *Fusicatenibacter*, *Lachnospiraceae_incertae_sedis*, *Roseburia*, *Ruminococcus2*, *Faecalibacterium* and *unclassified_Clostridiales*. The functional gut microbiota was characterized by few signature genera due to the high inter-individual variability. In 6-month's functional gut microbiota, signature highly abundant genera were *Bifidobacterium* and *Veillonella* while *Eubacterium* and *Faecalibacterium* were enriched at 18 months. In agreement with previous studies (Koenig et al., 2011; Valles et al., 2014; Backhed et al., 2015), our findings reflected the shift of gut microbiota towards an adult-like structure and composition as infants grew, possibly associated to physiological fitness to persist in increasingly lower oxygen levels.

Enrichment analysis identifies age-specific functional signatures in the gut microbiota

To determine how the functional capacity of the gut microbiota developed during early life, we analysed the metaproteome using Clusters of Orthologous Groups (COG) categories, hierarchically organized in 3 tiers where each tier is increasingly a more specific functional assignment (main, secondary and function categories). Due to the functional redundancy of orthologous proteins in our metaproteomics dataset, 9,173 protein groups narrowed to 1,117 COG functions (Supporting information Table S6). The mean number of protein groups per sample was 895±49. Overall distribution exhibited a rather even pattern across the samples where the most abundant secondary COGs belonged to 'Metabolism' category: 'Carbohydrate Transport and Metabolism', 'Amino Acid Transport and Metabolism', 'Energy Production and Conversion' and 'Inorganic Ion Transport and Metabolism' (Fig. 3A). This result is consistent with previous reports on the functional profile of protein groups expressed by gut microbiota (Verberkmoes et al., 2009; Rooijers et al., 2011; Ferrer et al., 2013; Perez-Cobas et al., 2013; Kolmeder et al., 2016). A group of 30 COG functions were identified in the gut microbiota of 90% of the subjects and may represent a functional core of biological processes (Supporting Information Table S6). Interestingly, seven of these core functions were enzymes also identified in adults (Verberkmoes et al., 2009; Kolmeder et al., 2012) suggesting a high stability of these proteins across human metaproteomes. These enzymes were glutamine synthetase and glutamate dehydrogenase in 'Amino acid transport and metabolism', and enolase, glyceraldehyde-3-phosphate dehydrogenase and fucose, glucuronate and xylose isomerases in 'Carbohydrate Transport and Metabolism'.

To look for significantly over- and under-represented COG functions in the gut metaproteomes of 6- and 18-months old infants, a comparative analysis was performed. The mean number of COG functions per sample was significantly higher in the 18-months metaproteome (1118±94) than in the 6 months one (767±59), indicating enrichment in microbial functionalities with age (p<0.01). Principal component analysis based on COG functions plot revealed a clear segregation between the two time points, with 40.6% of variance explained by the first component (Fig. 3B). These results indicated that functional complexity increased with time to create more similar inter-individual functional communities.

Metaproteomics analyses revealed significant differences between the sampled time points (Fig. 3C and Supporting Information Table S7). In the 6 months' gut microbiota, we observed over-representation of COGs classified into the main COG category 'Cellular processes and signalling', distributed within 'Cell wall membrane envelope biogenesis", "Cell motility", "Intracellular Trafficking Secretion and Vesicular Transport" and "Signal transduction mechanisms". The 18-month metaproteome was enriched in COGs classified into the main COG category 'Metabolism', distributed within 'Lipid transport and metabolism" and "Nucleotide transport and metabolism". The most significant COG functions within 'Cell wall membrane envelope biogenesis" were a protein translocase and a S-ribosylhomocysteine lyase mostly assigned to *Bifidobacterium*, involved in the control of gut colonization and protection against pathogens during early life (Christiaen et al., 2014). Significant abundances of an outer membrane adhesion protein involved in β-lactam resistance, an attachment invasion locus protein and a lipopolysaccharide transport and assembly protein binned to *Enterobacteriaceae*, and an autotransporter adhesin assigned to *Veillonellaceae* were observed. In addition, we

identified significant abundances of a Dps/Dpr ferritin-like protein involved in iron incorporation and six ABC-type transporters for bicarbonate, ion metals (nickel), amino acids (arginine, lysine, histidine and glutamine), dipeptides (cationic peptide) and inorganic oxides (phosphate, molybdate and tungstate) (Supporting Information Fig S2). According to its central role, these transporters were assigned to multiple taxa within *Actinobacteria* (*Bifidobacteriaceae*), *Firmicutes* (mostly *Ruminococcaceae*), and *Proteobacteria* (mostly *Enterobacteriaceae*). In contrast, only two COG functions involved in transport were enriched in the 18-months' microbiota, an oligopeptide ABC transporter binding lipoprotein binned to *Bifidobacteriaceae* and *Ruminococcaceae*, and a C_4-dicarboxylate-transport protein assigned to *Ruminococcaceae*. The identification of proteins involved in cellular transport is consistent with the observations of previous metaproteomes (Kolmeder et al., 2012). Different taxa within *Clostridia* were responsible for the abundance of lysozyme and carbon starvation protein A involved in cell defence, motility and agglutination. Taken together, this pattern of COG functions suggested that surface and signalling proteins of the gut microbiota were highly abundant at early stages since they regulate gut colonization and interaction with host cells.

Metabolic signatures differentiate chronological states of infants' gut microbiota

An interesting finding was that 32 COG functions within "Metabolism" with relevant roles in polysaccharide catabolism, central carbon metabolism and fermentation were differentially abundant between 6- and 18-months gut microbiota. To facilitate the understanding of these differences, we manually parsed and illustrated these functions into a hierarchically clustered heatmap and a comprehensive model of carbon metabolism (Fig. 4). The phylogenetic assignment of the functions at phylum level is provided (Supporting information Table S7).

(I) *Polysaccharide catabolism*: In an initial step, primary fermenters provide the cocktail of glycoside hydrolases (GH) or glycosidases to breakdown host glycans and dietary polysaccharides. Several metagenomic studies have determined the diversity of GH encoding genes in the infant and adult gut microbiota that ranged from fourteen to twenty-five GH families according to the carbohydrate-active enzymes database (CAZy) (Cantarel et al., 2009; Tasse et al., 2010; Cecchini et al., 2013; El Kaoutari et al., 2013; Backhed et al., 2015). We detected abundances of 24 GH COGs that belong to 20 GH families. To the best of our knowledge, metaproteomics results on the GH repertoire of gut microbiota have not been reported in such detail. The most abundant GH COG was β-galactosidase/β-glucuronidase, consistent with its high activity in gut microbiota (Hernandez et al., 2013), to which the largest number of peptides could be mapped in GH family. The mean number, protein abundance and catalytic activities of GH increased with age (Fig. 4A). These results reflected the high GH potential of the gut microorganisms metabolically prepared to degrade human mucin, milk oligosaccharides, plant and animal polysaccharides, even in exclusive breast-fed infants and before the introduction of solid foods (Koenig et al., 2011; Flint et al., 2012; Tailford et al., 2015). This is not surprising since mucin-adapted resident mutualists can alternatively forage on dietary plant polysaccharides to ensure gut microbial stability as the infant diet transitions to solid food (Marcobal et al., 2013). Consistent with the contribution of human and formula milk to infant diet, the microbiota of 6-months infants was enriched in β-galactosidase and arabinogalactan endo-1,4-β-galactanase, mostly expressed by *Actinobacteria*. Additionally, *Actinobacteria* and *Firmicutes* expressed an α-glucoside phosphotransferase IIC subunit, involved in the phosphorylative transport of glucose, glucosamine and n-acetylneuraminic acid while *Proteobacteria* expressed

maltoporin involved in maltose and maltodextrin transport. In contrast, the higher diversity and complexity of dietary carbohydrates in 18-months' infants resulted in a significant enrichment in α-amylase, α-glucosidase and β-glucosidase that were expressed by multiple taxa within *Bacteroidetes* and *Firmicutes*. Endo-β-N-acetylglucosaminidase D involved in the hydrolysis of branched oligosaccharides was expressed only by *Bacteroidetes* while cellobiose phosphorylase involved in the phosphate-dependent hydrolysis of cellulose was assigned to *Firmicutes*. The determination of gut microbial β-galactosidase, α-glucosidase and β-glucosidase activities in 6- and 18-monts infants confirmed their enrichment in the metaproteomes (Supporting Information Figure S3). Taken together, these results indicated that the gut microbiota used the proper upper glycolytic pathways depending on the availability of the carbohydrate source in a diet shifting from breast milk or formula to solid foods.

(II) *Central carbon metabolism*: Once a monosaccharide enters a cell, it flows through the Embden-Meyerhoff-Parnas (EMP), the pentose phosphate (PP) and the Entner-Doudoroff (ED) pathways that convert monosaccharides into phosphoenolpyruvate (PEP) (Fig. 4B). As expected, all COG functions in the EMP pathway were detected in the metaproteomes due to its central metabolic role. We observed very little abundance of pyruvate kinase, indicating that synthesis of PEP was the main outcome of EMP pathway in gut microbiota. Instead, PEP carboxylase was expressed in both metaproteomes because it allows bacteria to extract the second equivalent of ATP and generate oxaloacetate in an anaerobic environment (Macy and Probst, 1979). In EMP, two COG functions were significantly enriched in the 6 months' gut microbiota, glyceraldehyde-3-phosphate dehydrogenase assigned to taxa within *Bacteroidetes*, *Firmicutes* and *Proteobacteria*, and enolase binned to taxa in

Actinobacteria, *Firmicutes* and *Proteobacteria*. Glyceraldehyde-3-phosphate dehydrogenase and glutamate dehydrogenase were the COG functions to which the largest number of distinct peptides could be mapped in central carbon metabolism. While glyceraldehyde-3-phosphate dehydrogenase bridges PP and ED pathways to the lower EMP pathway, glutamate dehydrogenase has been shown to link the nitrogen and the carbon-cycle and to act as an electron sink in strict anaerobes (Kengen and Stams, 1994). The 6-month's metaproteome was also enriched in transketolase, an enzyme of PP pathway necessary for the *Bifidobacterium* shunt of glycolysis that yields acetate, glyceraldehyde-3-phosphate and ATP (de Vries et al., 1967). Finally, the null-detection of citrate synthase and the enrichment in citrate lyase in TCA cycle, assigned to *Firmicutes* and *Proteobacteria*, and isocitrate dehydrogenase, expressed by *Actinobacteria*, suggested that, in an environment that does not support aerobic respiration, bacteria may use the glut of PEP and the high availability of amino acids and CO_2 in a reverse TCA cycle to synthesise oxaloacetate (Macy et al., 1978).

With age, the 18-months metaproteome was also enriched in phosphoglyceromutase in EMP, mainly assigned to *Firmicutes*. In TCA cycle, we observed the increased abundances of two structurally and functionally-related membrane-bound enzymes, succinate dehydrogenase (SDH) and fumarate reductase (FRD). SDH, the enzyme that catalyses succinate oxidation, was mainly assigned to *Firmicutes* while *Bacteroidetes* used FRD. Fumarate reduction by FRD is the most used electron transport chain that generates ATP and succinate as a metabolic end product (Lu and Imlay, 2017). The fact that the subsequent step, succinyl-CoA synthesis, was very low abundant in *Bacteroidetes* is in line with the hypothesis of metabolic cross-feeding between *Bacteroidetes* and *Firmicutes* (Fischbach and Sonnenburg, 2011). An unexpected finding was the lack of detection of succinyl-CoA

synthetase (SCS). Rather, our metaproteomics data suggested that taxa within *Firmicutes* and to a minor extent in *Bacteroidetes*, *Proteobacteria* and *Verrumicrobia* employed a variation of the classical TCA cycle based on an unorthodox enzyme, acetate:succinate CoA-transferase (ASCT) to synthesize succinyl-CoA. Kwong *et al*. recently showed that ASCT genes were widespread in prokaryotic genomes and functionally replaced SCS in TCA cycle of human microbial commensals (Kwong et al., 2017). In a carbon-rich anaerobic gut ecosystem, this strategy may be a result of niche specialization where gut microbiota may use acetyl-CoA, the keystone molecule of central metabolism produced from monosaccharides, amino acids, fatty acids and other secondary metabolites, as driver of a reverse TCA cycle to maintain redox balance and obtain energy for growth. These results indicated that the gut microbial community displayed distinct strategies for central carbon transformations for PEP synthesis and biosynthetic reactions.

(III) *Fermentation*: Depending on the carbohydrate source and oxygen concentration, gut microbiota use distinct pathways of monosaccharide catabolism that end in the production of the main non-gaseous products of microbial fermentation: lactate and the short-chain fatty acids (SCFA) acetate, propanoate and butyrate (Fig. 4B). The profile of COG functions revealed distinct pathways of monosaccharide catabolism in the infant gut at the sampled time points. In particular, the enrichment in galactokinase, galactose mutarotase, gluconate/galactonate dehydratase, N-acetyl-glucosamine 6-phosphate 2-epimerase, 2-dehydro-3-deoxy-rhamnonate aldolase (KDRA) and fucose dehydrogenase suggested active catabolism of milk and mucin-derived monosaccharides by early gut microbiota. The phylogenetic assignments of these enzymes showed that *Proteobacteria* expressed KDRA suggesting that *Proteobacteria* may contribute to propanoate fermentation by

the propanediol pathway in early life microbiota. *Actinobacteria* was the major contributor to the catabolism of galactose via EMP, gluconate and galactonate via ED, and the unique taxa to catabolize mucin-derived n-acetylneuraminic acid to acetate and fucose to lactate. Since we did not detect lactoyl-CoA dehydratase in the acrylate pathway, the fact that lactate dehydrogenase was highly abundant and phylogenetically assigned to all phyla suggested that fucose-derived lactate may be a central substrate for metabolic cross-feeding in early life microbiota (Pham et al., 2016). Moreover, we observed enrichment in acetate kinase that produces ATP and acetate as end product, a strategy mainly used by *Actinobacteria*. This result highlighted the importance of *Bifidobacteria* metabolic contribution in a gut that is starting to be colonized where they may benefit host physiology by fermenting host-derive glycans to provide acetate that reduces faecal pH and protects host epithelial cells from enterotoxins (Fukuda et al., 2011).

With age, the action of GH on solid foods with a high diversity in glycan compositions generates a richer repertoire of released monosaccharides available for microbial metabolism. Accordingly, the 18-months metaproteome was enriched in glucuronate isomerase, 5-dehydro-4-deoxy-glucuronate ketol-isomerase and 2-dehydro-3-deoxygluconokinase that channel these acid hexoses to pyruvate and glyceraldehyde-3-phosphate synthesis by 2-dehydro-3-deoxy-phosphogluconate aldolase (KDPGA) in a semi-phosphorylative ED pathway. KDPGA was the enzyme with the highest protein abundance in our metaproteomes suggesting that this catabolic pathway is metabolically important for *Firmicutes* and *Bacteroidetes*. In SCFA metabolism, the enrichment in succinyl-CoA reductase, acetyl-CoA acyltransferase, enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenase, suggested that gut microbial metabolism shifted towards butyrate fermentation. The

fact that these protein groups were mainly assigned to *Clostridia* support their important role in the metabolic welfare of colonocytes by releasing *butyrate* as a fermentation end-product (Lopetuso et al., 2013). The enrichments in ASCT and succinyl-CoA reductase suggest a link between succinate fermentation to butyrate and acetate production, as has been observed in *Clostridium* (Sohling and Gottschalk, 1996). Butyrate kinase protein group was not detected, confirming that butyril-CoA:acetate-CoA transferase pathway is preferred by gut microbiota (Flint et al., 2012). Notably, our metaproteome revealed an alternate route for acetate synthesis in 18-months gut microbiota. We observed an enrichment in ethanolamine ammonia-lyase that catalyzes the adenosylcobalamin (AdoCbl)-dependent conversion of ethanolamine to acetaldehyde and ammonia (O'Brien et al., 1985). Ethanolamine is abundant in the human gut because the constant renewal of the intestinal epithelium daily releases 25% of enterocytes whose membranes are rich in phosphatidylethanolamine (Snoeck et al., 2005). Thus, our data suggested that ethanolamine may be used as a source of carbon and energy under aerobic and anaerobic conditions by gut microbiota. Taken together, our metaproteomics revealed the gut microbial age-associated maturation of fermentative strategies to harvest energy from diverse carbon resources in a shifting glycobiome environment.

Bifidobacteriaceae and _Lachnospiraceae_ are the hubs of succeeding anti-correlated functional co-occurrence modules in infant's gut microbiota

Analysis of microbial contributions to overall community activity indicated that the ecological network was remodelled as the functional gut microbiota of infants evolved with time. The topology of the co-occurrence networks of active taxa collapsed at family level determined by Pearson's correlation coefficient showed two mutually exclusive modules clustered by age (Fig. 5A). A low connected module built

by six families (20% of total nodes, 6.8% of total edges with a mean of 2 edges/node) evolved to a more complex and enriched one (80% of total nodes, 93% of total edges with a mean of 6.8 edges/node). It is plausible that these modules are built around ecologically relevant taxa whose pattern of functional interactions has a greater-than-average influence on network architecture. We calculated node parameters to identify families with the highest centrality in the modules (Supporting information Table S8). *Bifidobacteriaceae* was the node with the largest fraction of shortest edge paths and highest betweenness centrality in 6-months microbiota. These node properties defined *Bifidobacteriaceae* as a gatekeeper, cooperating simultaneously with different nodes of the module (Freeman, 1980). Its removal resulted in the fragmentation of this co-occurrence module, indicating that *Bifidobacteriaceae* was crucial for ecological module structure and persistence (Pocock et al., 2012). Eight *Bifidobacterium* species have been consistently identified in the human gastrointestinal tract (*Bifidobacterium adolescentis*, *B.breve*, *B.longum*, *B.pseudolongum*, *B.bifidum*, *B.pseudocatenulanum*, *B.dentium* and *B.animalis*) of which protein affiliation in our metaproteomes ruled out *B.dentium* but included other twenty *Bifidobacterium* species, suggesting that functional diversity in the *Bifidobacteriaceae* family may be richer than taxonomic one (Turroni et al., 2009). This low connected *Bifidobacteriaceae*-centred module anti-correlated with the highly connected (324 edges) and richer one (48 nodes) observed in 18-months' microbiota, suggesting competition or diversifying selection among modules. *Lachnospiraceae* was the hub (highest degree) in this cooperative module where *Desulfovibrionaceae*, *Ruminococcaceae*, *Rikenellaceae*, *Eubacteriaceae* and *Porphyromonadaceae* had a great importance in cooperative interactions in this module. *Lachnospiraceae, Eubacteriaceae* and *Clostridiaceae* may play a role in the transfer of biological information because these nodes shared the

highest betweenness centrality values in this module. At the genus level, *Veillonella*, *Escherichia*, *Bifidobacterium* and *Pseudomonas* were mutually exclusive with *Eubacterium*, *Ruminococcus*, *Faecalibacterium*, *Alistipes*, and *Bilophila*. Facultative anaerobes were not only more abundant but also were more positively correlated with obligate anaerobes in 6-months microbiota while showing many negative correlations in 18-months microbiota.

Further analysis revealed the succession of metabolic functions between taxa in these mutually exclusive consortia, suggesting a high level of functional redundancy between taxa to (Moya and Ferrer, 2016; Ruiz et al., 2017). Taxa within *Bifidobacteriaceae*, *Enterobacteriaceae* and *Veillonellaceae* constituted a functional consortium responsible for an important proportion of amino acid, carbohydrate, coenzyme, energy, inorganic, nucleotide and secondary metabolisms in 6-months microbiota (Fig. 5B). Age-related maturation restructured the contributions of taxa to metabolic performance where key co-occurrent families assembled an evolved functional consortium to fulfil overall gut microbial ecosystem requirements. These data supported that the maturation of the human microbiota during early life may be proposed as an example of ecological succession, in which communities undergo consecutive compositional and functional transitions in dominant taxa to establish physiological syntrophy among microbiota for niche adaptation (Koenig et al., 2011; Lozupone et al., 2012). Due to the variety of available nutrients, energy substrates and oxygen levels in the gut, it is reasonable to hypothesize that gut microbial taxa with diverse functional traits cooperate syntrophically to maximize energy yield and growth, as has been shown in other ecosystems (Morris et al., 2013). Comprehensive mathematical analysis of the characteristics of network edges between all genera pairs and their expressed functions will shed light on community-wide interactions via

primary degradation, resource competition and interspecies cross-feeding between gut microbes.

Concluding remarks

Although comparisons between individual children showed great differences in the dynamics of colonization, functional changes occurred more similarly across individuals, highlighting the non-stochastic nature of the bacterial functional community succession. Our results showed that considerable discordance existed between microbial composition and phylogenetic origin of proteins at all taxonomic levels. Age was the major driver of the rewiring of networks around succeeding key functional taxa and of the restructuring of community metabolic performances. Taken together, the detailed reconstruction of the gut microbial carbon metabolism presented here, including the assignment of enzymes to microbial taxa, revealed alternate temporary microbial and metabolic configurations where community-wide metabolic relationships to harvest energy by fermentation of prevailing dietary and host-derived carbon substrates, mainly glycans, differentiated chronological states. Our data provide a proteomic catalogue of the functional maturation of early gut microbiota, which may constitute an important research tool for indicators of future healthy or diseases states and for the design of microbiota-targeted health-promoting strategies early in life.

EXPERIMENTAL PROCEDURES

Subjects, experimental design and ethical guidelines

In the present study, full-term healthy infants aged 6- and 18-months, who did not present any intestinal disorders and had not taken antibiotics, were chosen from the panel of infants that belonged to PREOBE study cohort (Berglund et al., 2016). In this period of life, the transition from weaning to solid food consumption occurs. Characteristics of the study population are shown in Supporting information Table S1. In this project, pregnant women were recruited between 2007 and 2012 at San Cecilio and Mother-Infant University Hospitals in Granada, Spain. The study exclusion criteria for mothers were: simultaneous participation in any other research study, any kind of drug treatment, diagnosed diseases (e.g. pre-gestational diabetes, hypertension or preeclampsia, intrauterine growth retardation, maternal infection, hypo/hyperthyroidism, hepatic or renal disease), and vegan diet. Fresh stools were collected at 6- and 18-months after delivery and were immediately stored at -80ºC, until processing. The study included anthropometric measurements, health questionnaires and medical assessments of the child. This project followed the ethical standards recognized by the Declaration of Helsinki (reviewed in Hong-Kong 1989 and in Edinburgh 2000) and the EEC Good Clinical Practice recommendations (document 111/3976/88 1990), and current Spanish legislation regulating clinical research in humans (Royal Decree 561/1993). The study was explained to the participants before starting, and the parents signed an informed consent.

DNA extraction from stool samples

Genomic DNA was extracted from faecal bacteria of 6-month (n =68) and 18-month (n =72) old infants as previously described (Ferrer et al., 2013). Briefly, faecal samples were resuspended in 1mL of TN150 buffer (10mM Tris-HCl pH 8.0 and

150mM NaCl). Zirconium glass beads (0.3g) and 150µL of buffered phenol were added and bacteria were disrupted with a mini bead beater set to 5000rpm at 4°C for 15s (Biospec Products, USA). After centrifugation, genomic DNA was purified from the supernatant using phenol-chloroform extraction. Quality was checked by agarose gel electrophoresis and quantified with Quant-iT PicoGreen dsDNA assay kit (Invitrogen, Darmstadt, Germany).

16S rRNA gene sequencing and data processing

Genomic DNA from faecal bacteria was used as templates for 16S rRNA gene amplification using 27F and 338R universal primers and two consecutive PCR reactions to integrate Illumina multiplexing sequences as previously described (Camarinha-Silva et al., 2014). The library was prepared by pooling equimolar ratios of amplicons and was sequenced using an Illumina MiSeq platform (Genetic Service, University of Granada). Reads were demultiplexed and sorted, and paired ends were matched to give 240nt reads. Data set was filtered and OTUs were defined at 99% similarity with MOTHUR programs *unique.seqs* and *pre.cluster* (Schloss et al., 2009). Taxonomic classifications of OTUs were assigned using the naïve Bayesian algorithm CLASSIFIER of Ribosomal Database Project (Wang et al., 2007). OTUs were considered unassigned when confidence value score was lower than 0.8, and were annotated using upper taxonomic ranks.

Protein extraction, separation, identification and data processing

Protein extraction was performed from faecal bacteria of 6-months (n =29) and 18-months (n =27) old infants as previously described (Ferrer et al., 2013). Faecal samples (0.5g) were thawed and diluted in 1mL of 0.05% L-cysteine phosphate saline buffer solution (PBS) under anaerobic conditions. After differential centrifugation, faecal bacteria were disrupted by mechanical lysis in BugBuster Protein Extraction

Reagent (Novagen) for 30 min at room temperature, followed by sonication for 2.5min on ice. Protein extracts were centrifuged for 10min at 12.000rpm to separate cell debris. Protein concentrations were determined with the Bradford assay (Bradford, 1976). For 1-DE analysis, two 75-µgprotein samples (technical replicates denoted by a or b) were precipitated with five-fold volumes of ice-cold acetone and separated on a 12% acrylamide separating gel with the Laemmli buffer system (Laemmli, 1970).

After electrophoresis, protein bands were stained with Coomassie Brilliant Blue G-250. Entire protein lanes were individually cut into one band prior to performing in-gel tryptic digestion. Peptide lysates were desalted using C18 ZipTip prior to MS analysis. Peptides were analysed by nano-HPLC system Advion NanoMate and Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific). The peptides were eluted over 115 min with a gradient of 2 to 55% solvent (acetonitrile, 0.1% formic acid). MS scans were measured at a resolution of 120,000 in the scan range of 400-1600 *m/z*, MS2 in the Iontrap (rapid mode). Raw data files were searched with Proteome Discoverer (v1.4, Thermo Fisher Scientific) using the SequestHT algorithm against a database containing protein-coding entries of bacterial taxa selected via 16S rRNA gene sequencing. Only rank 1 peptides were considered to be identified with a threshold of FDR <1%. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaino et al., 2016) partner repository with the dataset identifier PXD009056. Higher protein abundance is represented by a higher number of MS/MS spectra acquired from peptides of the respective protein. Thus, protein abundances were calculated based on normalized spectral abundances that allow relative comparison of protein abundances over different samples (Bantscheff et al., 2012; Ferrer et al., 2013; Guazzaroni et al.,

2013). "PROteomics results Pruning & Homology group ANotation Engine" (PROPHANE) (Schneider et al., 2011) was used to assign proteins to their taxonomic and functional groups using the functional annotation of COGs. For KEGG pathway reconstructions, a BLASTP v2.2.27 search of the original protein sequences against NCBInr to retrieve KEGG Orthology identifiers was performed (Kanehisa et al., 2014). The use of a metaproteome-specific database containing fully sequenced genomes from closely related genera to the sample's strains and other documented gut genera in the database of proteomes together with the specificity of the identification procedure resulted in a high proportion of taxonomic and functional annotation (Denef et al., 2007).

Measurement of glycosidase activities

Glycosidase activities were quantified in protein extracts from purified faecal bacteria by measuring the release of p-nitrophenol-α-D-glucoside, p-nitrophenol-β-D-glucoside and p-nitrophenol-β-D-galactoside (Sigma Chemical Co., St. Louis, MO, USA) at 410 nm. One unit (U) of enzyme activity was defined as the amount of protein producing 1 μmol of reducing sugars in 1 min under the assay conditions.

Statistical and data analysis

Statistical analyses were carried out using SPSS version v19.0 (IBM, IL) and R statistical package (Team, 2014). Sankey flow chart was created with SankeyMATIC web tool (http://sankeymatic.com/). KEGG Mapper was used to visualize metabolic pathways. To quantify the amount of variability explained by each variable and subject in our different data sets, we calculated the coefficient of determination (R^2). For the response of composition and function of the microbiota, multivariate analysis of variance using distance matrices was performed, based on Bray-Curtis distance metrics. The matrices were partitioned in sources of variation

with subject and characteristics of the study population as explanatory variables. Significance of the pseudo-F ratios was assessed by permutation test (999 permutations, using the *adonis* function from the R package vegan) (Oksanen, 2011). β-diversity for compositional data was calculated as Unifrac distance with GUnifrac package. Permanova analysis of the distance between different time points was calculated with *adonis* function from *vegan* package. Bray-Curtis dissimilarity measures were calculated with *vegan* package and *anosim* test was used to establish significant differences between time points. Statistical Analysis of Metagenomic Profiles v2.0 was used to compare the abundances of taxa, COG categories and subcategories between time points (Parks et al., 2014). α-diversity indices were calculated with PAST software (Hammer et al., 2001). Significant differences were identified with the White's non-parametric *t* test. Benjamini & Hochberg FDR method was used to correct for multiple comparisons, and results with a *q*-value (corrected *p*-value <0.05) were retained. Pearson's correlation network analysis and visualization were carried out using Calypso v8.20 (Zakrzewski et al., 2017). Network node parameters were calculated using Cytoscape v3.1.1 (Shannon et al., 2003).

ACKNOWLEDGMENTS

CC and AS designed this study. TC performed the metaproteomic analyses under the supervision of NJ, SBH and MvB. TC, AR and AS analysed the data. AS wrote the manuscript. This work was supported by the European Union's 7th Framework Programme grant agreement no. 613979 (MyNewGut Project 2013/KB/613979) by the Spanish Ministry of Economy and Competitiveness Project BFU2012-40254-C03-01 and by Carlos III Institute of Health Projects 051579 and 021513. We thank Dr.Luz García-Valdés, Dr.Francisco Torres-Espínola and Hatim

Azaryah for their contribution. Tomás Cerdó participated in the PhD Program in Biomedicine of the University of Granada and is a fellow of the FPI (BES-2013-065133) Program at the Spanish Ministry of Economy and Competitiveness. This article will be part of Tomás Cerdó PhD. The authors are especially grateful for the use of the analytical facilities of the Centre for Chemical Microscopy (ProVIS) at the Helmholtz Centre for Environmental Research, which is supported by European Regional Development Funds (EFRE - Europe funds Saxony) and the Helmholtz Association.

CONFLICT OF INTEREST

The authors declare no conflict of interest

FIGURE LEGENDS

Figure 1. Taxonomic distribution and comparison of 16S rRNA sequences and protein groups characterized in faecal samples from the studied cohort of infants. (A) Phylogeny at phylum (circular graph) and family (bar graph) levels of total (16S rRNA gene sequences, top plots) and functional (organism-origin of protein groups, bottom plots) gut microbiota. Results for the same sample are lined up in bar plots. (B) Comparisons between the phylogeny of total (right) and functional (left) gut microbiota are tracked and confronted using a Sankey plot. To reduce the size of the Sankey plot, only highly abundant genera (mean relative abundance >1%) are shown. The heights of the rectangles indicate mean relative abundances in the datasets. Full description of the differences is shown in Supporting Information Table S3.

Figure 2. Total (16S rRNA gene sequences) and functional (organism-origin of

protein groups) gut microbial communities clustered according to age. (A, B) Scatterplot from principal coordinate analysis using weighted and unweighted Unifrac metrics in 6-months (red) and 18-months (green) old infants. *p* value for PERMANOVA test with 999 permutations. (C) Principal component analysis based on the phylogeny at phylum level of protein groups according to Bray-Curtis dissimilarity metrics in 6-months (red) and 18-months (green) old infants. (D) α-diversity and β-diversity of total and functional gut microbiota determined by Rao's diversity at phylum level. (E) β-diversity (Bray-Curtis dissimilarity metrics) of total and functional gut microbiota at phylum level. (F, G) Differentially abundant microbial taxa in total and functional gut microbiota of 6-months (red) and 18-months (green) old infants. Significantly discriminant taxon nodes are coloured and branch areas are shaded according to the highest ranked community for that taxon. If the taxon is not significantly different between sample communities, the corresponding node is coloured in yellow. For simplicity, only taxa meeting a linear discriminant analysis significant threshold >2 are shown. Full description of discriminant genera is shown in Supporting Information Table S5.

Figure 3. Comparison of metaproteomics profiles between 6-months (red) and 18-months (green) old infants. (A) COG distribution of the protein groups detected in the metaproteomes of 6-months and 18-months old infants (B) Principal component analysis based on COG function patterns according to Bray-Curtis dissimilarity metrics in 6-months and 18-months old infants (C) Differences in functional comparisons of metaproteomes from 6-months and 18-months infants at main (top plot) and secondary (bottom plot) COG hierarchy levels. Left: histogram: relative mean proportions and deviations; right plot: differences between proportions and significances.

Figure 4. Metabolic signatures that differentiate between the metaproteomes of 6-months and 18-months old infants. (A) Hierarchical clustering and heatmap of the abundances (log_{10} values) of glycoside hydrolases (GH) characterized in the metaproteomes of 6-months and 18-months old infants. CAZy families (right) and summation of GH abundances (bottom) in samples are shown. Values are indicated by colours ranging from orange to blue in log_{10} (bottom, -14 to 4) and absolute scales (top, 0 to 80). GH names with significant differences between infant groups are highlighted in bold. Each vertical line corresponds to one sample, identified at the bottom of the bar plot by a code that specifies the corresponding time-point (red, F24 for 6 months; green, F72 for 18-months) (B) Enzymes with significantly different abundances between the metaproteomes of 6-months (red arrows) and 18-months (green arrows) old infants are highlighted over a schematic carbon metabolism summary based on KEGG pathway maps. Semi-transparent boxes delimitate major central metabolic pathways (Embden-Meyerhoff-Parnas, EMP; pentose phosphate, PP; Entner-Doudoroff, ED) and short-chain fatty acids. Black arrows indicate enzymes detected in both metaproteomes with no differential abundance. Orange and purple boxes mark starting carbon substrates. Enzymes are indicated in the map as follows: 1 glyceraldehyde-3-phosphate dehydrogenase. 2 phosphoglyceromutase. 3 enolase. 4 transketolase. 5 citrate lyase. 6 isocitrate dehydrogenase. 7 succinate dehydrogenase. 8 fumarate reductase. 9 acetate:succinate CoA-transferase. 10 galactokinase. 11 galactose mutarotase. 12 gluconate/galactonate dehydratase. 13 N-acetyl-glucosamine 6-phosphate 2-epimerase. 14 2-dehydro-3-deoxy-rhamnonate aldolase. 15 fucose dehydrogenase. 16 acetate kinase. 17 glucuronate isomerase. 18 2-dehydro-3-deoxy-phosphogluconate aldolase. 19 succinyl-CoA reductase. 20 acetyl-CoA acyltransferase. 21 enoyl-CoA hydratase. 22 3-hydroxyacyl-CoA

dehydrogenase. 23 ethanolamine ammonia-lyase.

Figure 5. Age-driven remodelling and functional succession of co-occurrent taxa in the metaproteomes. (A) Co-occurrence network of taxa at family level in the metaproteomes of 6-months and 18-months old infants. The nodes represent families connected by significantly positive (thick brown lines) and negative (thin blue lines) and the node colour denotes association to each chronological state, both defined by the Pearson's correlation coefficient, set to a minimum of 0.6. The numbers in nodes refer to families' names shown in Supporting Information Table S5. Phylogenetic assignment at phylum level of the nodes is coloured by rings in blue for *Actinobacteria*, orange for *Bacteroidetes*, black for *Fusobacteria*, purple for *Firmicutes*, yellow for *Proteobacteria*, and light blue for *Verrumicrobia*. Bif, *Bifidobacteriaceae*; Clo, *Clostridiaceae*; Des, *Desulfovibrionaceae*; Enb, *Enterobacteriaceae*; Eno, *Enterococcaceae*; Eub, *Eubacteriaceae*; Lac, *Lachnospiraceae*; Ox, *Oxalobacteraceae*; Por, *Porphyromonadaceae*; Pse, *Pseudomonadaceae*; Rik, *Rikenellaceae*; Rum, *Ruminococcaceae*; Vei, *Veillonellaceae*. (B) Mean relative abundances of COGs involved in metabolism expressed by key co-occurrent families in the modules. Data are expressed as $\log_{10}$ values. [C] Energy production and conversion. [E] Amino acid transport and metabolism. [F] Nucleotide transport and metabolism. [G] Carbohydrate transport and metabolism. [H] Coenzyme transport and metabolism. [I] Lipid transport and metabolism. [P] Inorganic ion transport and metabolism. [Q] Secondary metabolites biosynthesis, transport, and catabolism.

Supporting Information Figure S1. Phylogeny of the functional gut microbiota (organism-origin of protein groups) at genus level.

Supporting Information Figure S2. Transport functions with differential abundance between the metaproteomes of 6- and 18-months old infants.

Supporting Information Figure S3. Gut microbial glycosidase activities in 6- and 18-months old infants. Mean enzymatic specific activities (units per gram of total protein) ± SEM from faecal microbiota are represented.

REFERENCES

Asnicar, F., Manara, S., Zolfo, M., Truong, D.T., Scholz, M., Armanini, F. et al. (2017) Studying Vertical Microbiome Transmission from Mothers to Infants by Strain-Level Metagenomic Profiling. *mSystems* **2**: pii: e00164-00116.

Backhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P. et al. (2015) Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe* **17**: 690-703.

Bantscheff, M., Lemeer, S., Savitski, M.M., and Kuster, B. (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **404**: 939-965.

Berglund, S.K., Garcia-Valdes, L., Torres-Espinola, F.J., Segura, M.T., Martinez-Zaldivar, C., Aguilar, M.J. et al. (2016) Maternal, fetal and perinatal alterations associated with obesity, overweight and gestational diabetes: an observational cohort study (PREOBE). *BMC Public Health* **16**: 207.

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248-254.

Brooks, B., Mueller, R.S., Young, J.C., Morowitz, M.J., Hettich, R.L., and Banfield, J.F. (2015) Strain-resolved microbial community proteomics reveals simultaneous aerobic and anaerobic function during gastrointestinal tract colonization of a preterm infant. *Front Microbiol* **6**: 654.

Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S., and Thomas, T. (2011) Bacterial community assembly based on functional genes rather than species. *Proceedings of the National Academy of Sciences* **108**: 14288-14293.

Camarinha-Silva, A., Jauregui, R., Chaves-Moreno, D., Oxley, A.P., Schaumburg, F., Becker, K. et al. (2014) Comparing the anterior nare bacterial community of two discrete human populations using Illumina amplicon sequencing. *Environ Microbiol* **16**: 2939-2952.

Cangelosi, G.A., and Meschke, J.S. (2014) Dead or alive: molecular assessment of microbial viability. *Appl Environ Microbiol* **80**: 5884-5891.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**: D233-238.

Cecchini, D.A., Laville, E., Laguerre, S., Robe, P., Leclerc, M., Dore, J. et al. (2013) Functional metagenomics reveals novel pathways of prebiotic breakdown by human gut bacteria. *PLoS One* **8**: e72766.

Cerdó, T., García-Valdés, L., Altmae, S., Ruiz, A., Suárez, A., and Campoy, C. (2016) Role of microbiota function during early life on child's neurodevelopment. *Trends in Food Science & Technology* **57**: 15.

Cerdó, T., Ruiz, A., Jáuregui, R., Azaryah, H., Torres-Espínola, F.J., Garcia-Valdés, L. et al. (2017) Maternal obesity is associated with gut microbial metabolic potential in offspring during infancy. *J Physiol Biochem* **73**: 10.

Charbonneau, M.R., Blanton, L.V., DiGiulio, D.B., Relman, D.A., Lebrilla, C.B., Mills, D.A., and Gordon, J.I. (2016) A microbial perspective of human developmental biology. *Nature* **535**: 48-55.

Christiaen, S.E., O'Connell Motherway, M., Bottacini, F., Lanigan, N., Casey, P.G., Huys, G. et al. (2014) Autoinducer-2 plays a crucial role in gut colonization and probiotic functionality of Bifidobacterium breve UCC2003. *PLoS One* **9**: e98111.

Consortium, H.M.P. (2012) Structure, function and diversity of the healthy human microbiome. *Nature* **486**: 207-214.
Cox, L.M., Yamanishi, S., Sohn, J., Alekseyenko, A.V., Leung, J.M., Cho, I. et al. (2014) Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell* **158**: 705-721.
Cryan, J.F., and Dinan, T.G. (2012) Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat Rev Neurosci* **13**: 701-712.
de Vries, W., Gerbrandy, S.J., and Stouthamer, A.H. (1967) Carbohydrate metabolism in Bifidobacterium bifidum. *Biochim Biophys Acta* **136**: 415-425.
Denef, V.J., Shah, M.B., VerBerkmoes, N.C., Hettich, R.L., and Banfield, J.F. (2007) Implications of strain-and species-level sequence divergence for community and isolate shotgun proteomic analysis. *Journal of proteome research* **6**: 3152-3161.
El Kaoutari, A., Armougom, F., Gordon, J.I., Raoult, D., and Henrissat, B. (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol* **11**: 497-504.
Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K. et al. (2016) Population-level analysis of gut microbiome variation. *Science* **352**: 560-564.
Ferrer, M., Ruiz, A., Lanza, F., Haange, S.B., Oberbach, A., Till, H. et al. (2013) Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure. *Environ Microbiol* **15**: 211-226.
Fischbach, M.A., and Sonnenburg, J.L. (2011) Eating for two: how metabolism establishes interspecies interactions in the gut. *Cell Host Microbe* **10**: 336-347.
Flint, H.J., Scott, K.P., Duncan, S.H., Louis, P., and Forano, E. (2012) Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **3**: 289-306.
Franzosa, E.A., Morgan, X.C., Segata, N., Waldron, L., Reyes, J., Earl, A.M. et al. (2014) Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci U S A* **111**: E2329-2338.
Freeman, L.C. (1980) The gatekeeper, pair-dependency and structural centrality. *Quality and Quantity* **14**: 7.
Fukuda, S., Toh, H., Hase, K., Oshima, K., Nakanishi, Y., Yoshimura, K. et al. (2011) Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**: 543-547.
Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R. et al. (2014) Human genetics shape the gut microbiome. *Cell* **159**: 789-799.
Gosalbes, M.J., Durban, A., Pignatelli, M., Abellan, J.J., Jimenez-Hernandez, N., Perez-Cobas, A.E. et al. (2011) Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS One* **6**: e17447.
Guazzaroni, M.-E., Herbst, F.-A., Lores, I., Tamames, J., Peláez, A.I., Lopez-Cortés, N. et al. (2013) Metaproteogenomic insights beyond bacterial response to naphthalene exposure and bio-stimulation. *The ISME journal* **7**: 122.
Hammer, Ø., Harper, D.A.T., and Ryan, P.D. (2001) PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4**: 9.
Harmsen, H.J., and de Goffau, M.C. (2016) The Human Gut Microbiota. *Adv Exp Med Biol* **902**: 95-108.
Hernandez, E., Bargiela, R., Diez, M.S., Friedrichs, A., Perez-Cobas, A.E., Gosalbes, M.J. et al. (2013) Functional consequences of microbial shifts in the human gastrointestinal tract linked to antibiotic treatment and obesity. *Gut Microbes* **4**: 306-315.

Hettich, R.L., Pan, C., Chourey, K., and Giannone, R.J. (2013) Metaproteomics: harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. *Anal Chem* **85**: 4203-4214.

Juste, C., Kreil, D.P., Beauvallet, C., Guillot, A., Vaca, S., Carapito, C. et al. (2014) Bacterial protein signals are associated with Crohn's disease. *Gut* **63**: 1566-1577.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* **42**: D199-205.

Kengen, S.W.M., and Stams, A.J.M. (1994) Formation of l-alanine as a reduced end product in carbohydrate fermentation by the hyperthermophilic archaeon Pyrococcus furiosus. *Archives of Microbiology* **161**: 7.

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

Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R. et al. (2011) Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A* **108 Suppl 1**: 4578-4585.

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

Kolmeder, C.A., Salojarvi, J., Ritari, J., de Been, M., Raes, J., Falony, G. et al. (2016) Faecal Metaproteomic Analysis Reveals a Personalized and Stable Functional Microbiome and Limited Effects of a Probiotic Intervention in Adults. *PLoS One* **11**: e0153294.

Kwong, W.K., Zheng, H., and Moran, N.A. (2017) Convergent evolution of a modified, acetate-driven TCA cycle in bacteria. *Nat Microbiol* **2**: 17067.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Lopetuso, L.R., Scaldaferri, F., Petito, V., and Gasbarrini, A. (2013) Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog* **5**: 23.

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012) Diversity, stability and resilience of the human gut microbiota. *Nature* **489**: 220-230.

Lu, Z., and Imlay, J.A. (2017) The Fumarate Reductase of Bacteroides thetaiotaomicron, unlike That of Escherichia coli, Is Configured so that It Does Not Generate Reactive Oxygen Species. *MBio* **8**.

Macy, J.M., and Probst, I. (1979) The biology of gastrointestinal bacteroides. *Annu Rev Microbiol* **33**: 561-594.

Macy, J.M., Ljungdahl, L.G., and Gottschalk, G. (1978) Pathway of succinate and propionate formation in Bacteroides fragilis. *J Bacteriol* **134**: 84-91.

Mao, L., and Franke, J. (2015) Symbiosis, dysbiosis, and rebiosis-the value of metaproteomics in human microbiome monitoring. *Proteomics* **15**: 1142-1151.

Marcobal, A., Southwick, A.M., Earle, K.A., and Sonnenburg, J.L. (2013) A refined palate: bacterial consumption of host glycans in the gut. *Glycobiology* **23**: 1038-1046.

Maurice, C.F., and Turnbaugh, P.J. (2013) Quantifying the metabolic activities of human-associated microbial communities across multiple ecological scales. *FEMS Microbiol Rev* **37**: 830-848.

Maurice, C.F., Haiser, H.J., and Turnbaugh, P.J. (2013) Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell* **152**: 39-50.

McNulty, N.P., Yatsunenko, T., Hsiao, A., Faith, J.J., Muegge, B.D., Goodman, A.L. et al. (2011) The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci Transl Med* **3**: 106ra106.
Milani, C., Duranti, S., Bottacini, F., Casey, E., Turroni, F., Mahony, J. et al. (2017) The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. *Microbiol Mol Biol Rev* **81**.
Morris, B.E., Henneberger, R., Huber, H., and Moissl-Eichinger, C. (2013) Microbial syntrophy: interaction for the common good. *FEMS Microbiol Rev* **37**: 384-406.
Moya, A., and Ferrer, M. (2016) Functional Redundancy-Induced Stability of Gut Microbiota Subjected to Disturbance. *Trends Microbiol* **24**: 402-413.
O'Brien, R.J., Fox, J.A., Kopczynski, M.G., and Babior, B.M. (1985) The mechanism of action of ethanolamine ammonia-lyase, an adenosylcobalamin-dependent enzyme. Evidence that the hydrogen transfer mechanism involves a second intermediate hydrogen carrier in addition to the cofactor. *J Biol Chem* **260**: 16131-16136.
Oksanen, J.F.G.B., R. Kindt, P. Legendre, P.R. Minchin, R.B. O'Hara, G.L. Simpson, P. Solymos, M.H.H. Stevens, H.Wagner. (2011) Package 'vegan' version 2.0-2. .
Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., and Brown, P.O. (2007) Development of the human infant intestinal microbiota. *PLoS Biol* **5**: e177.
Parks, D.H., Tyson, G.W., Hugenholtz, P., and Beiko, R.G. (2014) STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**: 3123-3124.
Perez-Cobas, A.E., Gosalbes, M.J., Friedrichs, A., Knecht, H., Artacho, A., Eismann, K. et al. (2013) Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut* **62**: 1591-1601.
Pham, V.T., Lacroix, C., Braegger, C.P., and Chassard, C. (2016) Early colonization of functional groups of microbes in the infant gut. *Environ Microbiol* **18**: 2246-2258.
Pocock, M.J., Evans, D.M., and Memmott, J. (2012) The robustness and restoration of a network of ecological networks. *Science* **335**: 973-977.
Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C. et al. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**: 59-65.
Rooijers, K., Kolmeder, C., Juste, C., Doré, J., De Been, M., Boeren, S. et al. (2011) An iterative workflow for mining the human intestinal metaproteome. *BMC genomics* **12**: 6.
Ruiz, A., Cerdó, T., Jáuregui, R., Pieper, D.H., Marcos, A., Clemente, A. et al. (2017) One-year calorie restriction impacts gut microbial composition but not its metabolic performance in obese adolescents. *Environ Microbiol* **19**: 1536-1551.
Salonen, A., Lahti, L., Salojarvi, J., Holtrop, G., Korpela, K., Duncan, S.H. et al. (2014) Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. *Isme j* **8**: 2218-2230.
Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.
Schneider, T., Schmid, E., de Castro, J.V., Jr., Cardinale, M., Eberl, L., Grube, M. et al. (2011) Structure and function of the symbiosis partners of the lung lichen (Lobaria pulmonaria L. Hoffm.) analyzed by metaproteomics. *Proteomics* **11**: 2752-2756.
Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011) Metagenomic biomarker discovery and explanation. *Genome Biol* **12**: R60.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498-2504.
Snoeck, V., Goddeeris, B., and Cox, E. (2005) The role of enterocytes in the intestinal barrier function and antigen uptake. *Microbes Infect* **7**: 997-1004.
Sohling, B., and Gottschalk, G. (1996) Molecular analysis of the anaerobic succinate degradation pathway in Clostridium kluyveri. *J Bacteriol* **178**: 871-880.
Tailford, L.E., Crost, E.H., Kavanaugh, D., and Juge, N. (2015) Mucin glycan foraging in the human gut microbiome. *Front Genet* **6**: 81.
Tamburini, S., Shen, N., Wu, H.C., and Clemente, J.C. (2016) The microbiome in early life: implications for health outcomes. *Nat Med* **22**: 713-722.
Tanca, A., Abbondio, M., Palomba, A., Fraumene, C., Manghina, V., Cucca, F. et al. (2017) Potential and active functions in the gut microbiota of a healthy human cohort. *Microbiome* **5**: 79.
Tasse, L., Bercovici, J., Pizzut-Serin, S., Robe, P., Tap, J., Klopp, C. et al. (2010) Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res* **20**: 1605-1612.
Team, R.C. (2014) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. In.
Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E. et al. (2009) A core gut microbiome in obese and lean twins. *Nature* **457**: 480-484.
Turroni, F., Marchesi, J.R., Foroni, E., Gueimonde, M., Shanahan, F., Margolles, A. et al. (2009) Microbiomic analysis of the bifidobacterial population in the human distal gut. *Isme j* **3**: 745-751.
Valles, Y., Artacho, A., Pascual-Garcia, A., Ferrus, M.L., Gosalbes, M.J., Abellan, J.J., and Francino, M.P. (2014) Microbial succession in the gut: directional trends of taxonomic and functional change in a birth cohort of Spanish infants. *PLoS Genet* **10**: e1004406.
Verberkmoes, N.C., Russell, A.L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J. et al. (2009) Shotgun metaproteomics of the human distal gut microbiota. *Isme j* **3**: 179-189.
Vizcaino, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I. et al. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* **44**: D447-456.
Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261-5267.
Young, J.C., Pan, C., Adams, R.M., Brooks, B., Banfield, J.F., Morowitz, M.J., and Hettich, R.L. (2015) Metaproteomics reveals functional shifts in microbial and human proteins during a preterm infant gut colonization case. *Proteomics* **15**: 3463-3473.
Zakrzewski, M., Proietti, C., Ellis, J.J., Hasan, S., Brion, M.J., Berger, B., and Krause, L. (2017) Calypso: a user-friendly web-server for mining and visualizing microbiome-environment interactions. *Bioinformatics* **33**: 782-783.
Zoetendal, E.G., Akkermans, A.D.L., Akkermans-van Vliet, W.M., de Visser, J.A.G.M., and de Vos, W.M. (2011) The Host Genotype Affects the Bacterial Community in the Human Gastrointestinal Tract. *Microbial Ecology in Health and Disease* **13**.
Zwittink, R.D., van Zoeren-Grobben, D., Martin, R., van Lingen, R.A., Groot Jebbink, L.J., Boeren, S. et al. (2017) Metaproteomics reveals functional differences

in intestinal microbiota development of preterm infants. *Mol Cell Proteomics* **16**: 1610-1620.