Minireview

Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations

Leonardo Mancabelli,[1†] Christian Milani,[1†]
Gabriele Andrea Lugli,[1] Francesca Turroni,[1]
Chiara Ferrario,[1] Douwe van Sinderen,[2]
Marco Ventura[1]*

[1]*Laboratory of Probiogenomics, Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy.*
[2]*APC Microbiome Institute and School of Microbiology, Bioscience Institute, National University of Ireland, Cork, Ireland.*

Summary

Metagenomic studies of the human gut microbiome have only recently begun to explore the differences in taxonomic composition between subjects from diverse geographical origins. Here, we compared taxonomy, resistome and functional metabolic properties of publicly available shotgun datasets of human fecal samples collected from different geographical regions (Europe, North America, Asia and Oceania). Such datasets encompassed gut microbiota information corresponding to 13 developed/industrialized societies, as well as two traditional hunter-gatherer, pre-agricultural communities (Tanzanian and Peruvian individuals). Assessment of the retrieved taxonomic profiles allowed the most updated reconstruction of the global core-microbiome as based on currently available data, as well as the identification and targeted genome reconstruction of bacterial taxa that appear to have been lost and/or acquired during urbanization/ industrialization. Functional characterization of these metagenomic datasets indicates that the urbanization/ industrialization process which occurred in recent human history has shaped the gut microbiota through the acquisition and/or loss of specific gut microbes,

thereby potentially impacting on the overall functionality of the gut microbiome.

Introduction

The composite activities of the human gut microbiome impact on various functions of its host, including gut physiology, intestinal metabolism, and immune system modulation (Round and Mazmanian, 2009). During the life span of its host, the gut microbiota composition is influenced by factors such as diet, lifestyle and environment (Conlon and Bird, 2015). Recently, some major research efforts, such as the European Metagenomics of the Human Intestinal Tract (MetaHIT) (http://www.metahit.eu) project and the American Human Microbiome Project (HMP) (http://hmpdacc.org), have dissected the gut microbiome composition and functionality across different human populations. In order to better understand the role of the microbiota and its co-evolution with human host, it's pivotal to compare the microbiome of urbanized/industrialized populations with pre-agricultural/isolated populations. Currently, only a small number of studies have compared the microbiomes of pre-agricultural/isolated communities and urbanized/industrialized populations in order to detect possible differences in composition and their potential correlation with disease (risk) and/or metabolic disorders (De Filippo *et al.*, 2010; Yatsunenko *et al.*, 2012; Schnorr *et al.*, 2014; Clemente *et al.*, 2015; Martinez *et al.*, 2015; Obregon-Tito *et al.*, 2015; Rampelli *et al.*, 2015).

Moreover, many investigations of the gut microbiota were simply interested in microbial cataloguing by means of 16S rRNA gene-based amplicon sequencing (De Filippo *et al.*, 2010; Yatsunenko *et al.*, 2012; Schnorr *et al.*, 2014; Clemente *et al.*, 2015; Dehingia *et al.*, 2015; Martinez *et al.*, 2015). Nonetheless, an increasing number of investigative efforts are based on shotgun metagenomics sequencing, aimed at functionally characterizing gut microbiomes (Qin *et al.*, 2010; Qin *et al.*, 2012; Karlsson *et al.*, 2013; Li *et al.*, 2014; Lim *et al.*, 2014; Zeller *et al.*, 2014; Feng *et al.*, 2015; Obregon-Tito *et al.*, 2015; Rampelli *et al.*, 2015; Voigt

et al., 2015). The majority of gut microbiota analyses have focused on specific populations, yet do not provide any comparative analysis of gut microbiomes from different geographical regions or habitation conditions (Karlsson *et al.*, 2013; Lim *et al.*, 2014; Feng *et al.*, 2015; Voigt *et al.*, 2015; Raymond *et al.*, 2016).

So far, only a small number of studies have investigated how the urbanization/industrialization process may have affected gut microbiota composition (De Filippo *et al.*, 2010; Yatsunenko *et al.*, 2012; Schnorr *et al.*, 2014; Clemente *et al.*, 2015; Dehingia *et al.*, 2015; Martinez *et al.*, 2015; Obregon-Tito *et al.*, 2015; Rampelli *et al.*, 2015). However, most of these investigations were interested in the identification of specific microbial taxa that correlate with the adaptation of human to different lifestyles without supporting their findings with shotgun metagenomics analyses (De Filippo *et al.*, 2010; Yatsunenko *et al.*, 2012; Schnorr *et al.*, 2014; Clemente *et al.*, 2015; Dehingia *et al.*, 2015; Martinez *et al.*, 2015). These studies reported a range of differences in the gut microbiota composition between industrialized and pre-agricultural societies, reflecting the dietary and environmental factors typical of their lifestyle. Particularly, pre-agricultural communities displayed high abundance of members of the *Prevotella* genus, known to harbour genetic features for the breakdown of cellulose and xylan. While the taxa *Treponema* and *Brachyspira* were undetected in industrialized populations (Schnorr *et al.*, 2014; Obregon-Tito *et al.*, 2015; Rampelli *et al.*, 2015). Members of the genus *Treponema* were also found in non-human primates and all traditional populations studied to date, suggesting that these gut commensals have been lost from the gut microbiota of human beings associated with urban-industrialized societies (Obregon-Tito *et al.*, 2015).

So far, only two publications have reported on the use of shotgun sequencing to perform an in depth functional analysis of the gut microbiome of Hadza and Matses populations, which represent pre-agricultural societies (Obregon-Tito *et al.*, 2015; Rampelli *et al.*, 2015). These studies reported that Hazda and Matses microbiomes possess a lower abundance of antibiotic resistance genes and an extended metabolic potential toward utilization of carbohydrates when compared with human gut microbiomes from industrialized areas (Obregon-Tito *et al.*, 2015; Rampelli *et al.*, 2015). Nevertheless, both studies compared the gut microbiomes of pre-agricultural populations to very small cohorts of fecal samples collected from individuals living in urbanized societies that can not be considered representative of the general western population (Obregon-Tito *et al.*, 2015; Rampelli *et al.*, 2015).

In this minireview, we evaluated the notion that urbanization/industrialization processes have substantially influenced the composition and functionality of the human gut microbiome. This evaluation was performed from a taxonomic and functional perspective by means of a meta-analysis of all publicly available human gut shotgun metagenomic datasets corresponding to urbanized and pre-agricultural societies. An overall summary of the observations discussed in this study, accompanied by a comparison with previously published data, is reported in Supporting Information Table S1.

The urbanized/industrialized populations include individuals residing in high-income geographical regions that are densely populated, i.e. metropolitan areas, towns, but also individuals inhabiting rural areas that have access to medical care and obtain high hygiene standards and follow a globalized westernized diet (Karlsson *et al.*, 2013; Li *et al.*, 2014; Lim *et al.*, 2014; Zeller *et al.*, 2014; Voigt *et al.*, 2015; Raymond *et al.*, 2016). In contrast, the pre-agricultural communities encompass individuals living in isolated areas, who have no or limited access to medical care, and whose diet is based on foods gathered and/or huntered from their immediate environment, which was further processed only by cooking (Schnorr *et al.*, 2014; Obregon-Tito *et al.*, 2015). The lifestyle (including diet) of these human communities resembles that of people from ancestral human populations (Clemente *et al.*, 2015). We would also have liked to evaluate the microbiome of ancestral human populations and microbiota-host co-evolution at various stages of human evolution, from the Neolithic to the modern urbanized/industrilized populations. Unfortunately, such meta-studies that cover human evolutionary history are currently not possible due to a lack of corresponding data sets.

These data allow the identification of specific compositional and functional differences between the gut microbiome of urbanized/-industrialized vs. pre-agricultural populations. Furthermore, this metagenomic information may allow genome reconstruction of bacterial taxa that seem to have been lost from or gained by individuals living in urban-industrialized countries.

The worldwide gut microbiome database

All datasets included in this meta-analysis were collected from published human gut microbiome studies. Literature searches allowed us to exclusively select shotgun metagenomic data generated by Illumina technology, i.e. the (currently) most preferred and reliable technology to perform shotgun metagenomics studies (Quail *et al.*, 2012). Datasets had to be obtained starting from DNA extracted from fecal samples of healthy and adult human individuals. In order to obtain comparable sequence information, it was necessary to exclude those metagenomic projects that did not have at least six datasets with an average quality value of > 25 and an average

read length of > 95 bp (following quality filtering). These criteria resulted in the selection of 18 publicly available shotgun metagenomic sequencing projects from 14 different geographical regions covering Africa, Asia, Europe, North/South America and Oceania (Supporting Information Table S2).

Metadata of the sequencing projects was employed to select only datasets of healthy individuals whose age ranged between 21 and 65, not undergoing any antibiotic or probiotic treatment, and not suffering from gut-related diseases/disorders. Unfortunately, the metadata of these collected datasets frequently do not provide information related to eating habits, diet and associated nutritional proprieties. This represents a crucial limitation which prevents us from assessing the role of diet in shaping the gut microbiota composition. In particular, human gut microbiomes were obtained for individuals from Austria, Denmark, France, Germany, Spain, Sweden, China, South Korea, Canada (Qin *et al.*, 2010; 2012; Karlsson *et al.*, 2013; Li *et al.*, 2014; Lim *et al.*, 2014; Zeller *et al.*, 2014; Feng *et al.*, 2015; Voigt *et al.*, 2015; Nishijima *et al.*, 2016; Raymond *et al.*, 2016), Italy (SRP079680), Australia (PRJEB6092) and United States (HMP DACC, http://www.hmpdacc.org). Furthermore, the gut microbiomes of just two pre-agricultural communities, i.e. Hadza (from Tanzania) (Schnorr *et al.*, 2014; Rampelli *et al.*, 2015) and Matses (from Peru) (Obregon-Tito *et al.*, 2015), were publicly avaiable and thus included in our analyses. While the avaibility of two datasets is not sufficient for a comprehensive and statistically significant representation of pre-agricultural populations in general, these data at least allow us to gain some initial insights into the effects of urbanization/industrialization on gut microbiome.

Samples were taxonomically profiled through Meta-PhlAn2 software (Truong *et al.*, 2015) and the 10 samples with profiles closest to the average of each population were chosen as representatives for in-depth functional analyses. These datasets were included in the Worldwide Gut Microbiome Database (WGMD), and encompassed a total of 142 samples, facilitating both a taxonomic and a functional overview of the covered countries (Supporting Information Table S2).

Notably, the use of shotgun metagenomic data minimizes biases, in particular when compared with other taxonomic profiling methods that are based on marker genes, such as the widely used 16S rRNA profiling. In fact, for this latter approach there are serious methodological concerns (e.g. efficiency and universality of the PCR primers and DNA extraction protocols) (Milani *et al.*, 2013) that might prevent a faithful comparison between data sets derived from different studies. Moreover, the use of shotgun datasets allows us to perform functional investigation of the human gut microbiome, which is not possible with the use of 16S rRNA profiling data sets. Thus, despite limitations related to different procedures concerning stool collection/processing and DNA extraction protocols followed for the analysed samples (Qin *et al.*, 2010; 2012; Karlsson *et al.*, 2013; Li *et al.*, 2014; Lim *et al.*, 2014; Schnorr *et al.*, 2014; Zeller *et al.*, 2014; Feng *et al.*, 2015; Obregon-Tito *et al.*, 2015; Rampelli *et al.*, 2015; Voigt *et al.*, 2015; Nishijima *et al.*, 2016; Raymond *et al.*, 2016), the geographical coverage of the WGMD allows a global gut microbiota assessment related to both composition and functionality (Supporting Information Table S2).

Profiling of the global human gut microbiota

A 3-Dimensional Principal Coordinate Analysis (PCoA) representing the beta-diversity (Caporaso *et al.*, 2010) based on Bray-curtis dissimilarity index of the different gut microbiomes encompassing the WGMD, highlighted an intriguing profiling based on geographical regions (Fig. 1). In this context, the majority of the samples group together based on their geographic origin, whereas Australian, German and South Korean individuals elicit an uneven distribution (Fig. 1). Notably, the Hadza and Matses pre-agricultural individuals grouped as two separate clusters with respect to all other datasets, likely reflecting their distinct geographical origin, as well as their unique life style and diet (Supporting Information Table S3). Nevertheless, an extension of this meta-analysis with additional pre-agricultural samples is needed in order to statistically validate these results. Overall, the obtained results are statistically supported by a PERMANOVA *p*-value of < 0.001. Such findings may indicate that lifestyle and diet are important factors influencing the gut microbiota composition. In contrast, the urbanized populations show partial overlap, which may point towards the presence of a shared core microbiome (see below).

Inspection of predicted taxonomic profiles of the WGMD at phylum level show a preponderant presence of members of the *Bacteroidetes* phylum (Supporting Information Fig. S1), in particular of the *Bacteroides* genus in the urban-industrialized populations (average of 28.76%) (*p*-value < 0.05) (Supporting Information Fig. S2), whereas the *Prevotella* genus is highly abundant in the Hadza community (34.42%) (*p*-value < 0.05) (Supporting Information Fig. S2). Although some members of the *Bacteroides* genus are known to metabolize complex polysaccharides (Xu and Gordon, 2003), recent studies have shown that the presence of members of the *Bacteroides* genus positively correlates with a diet enriched in protein and animal fat. The presence of members of the *Prevotella* genus on the other hand is associated with regular consumption of a carbohydrate-based diet,
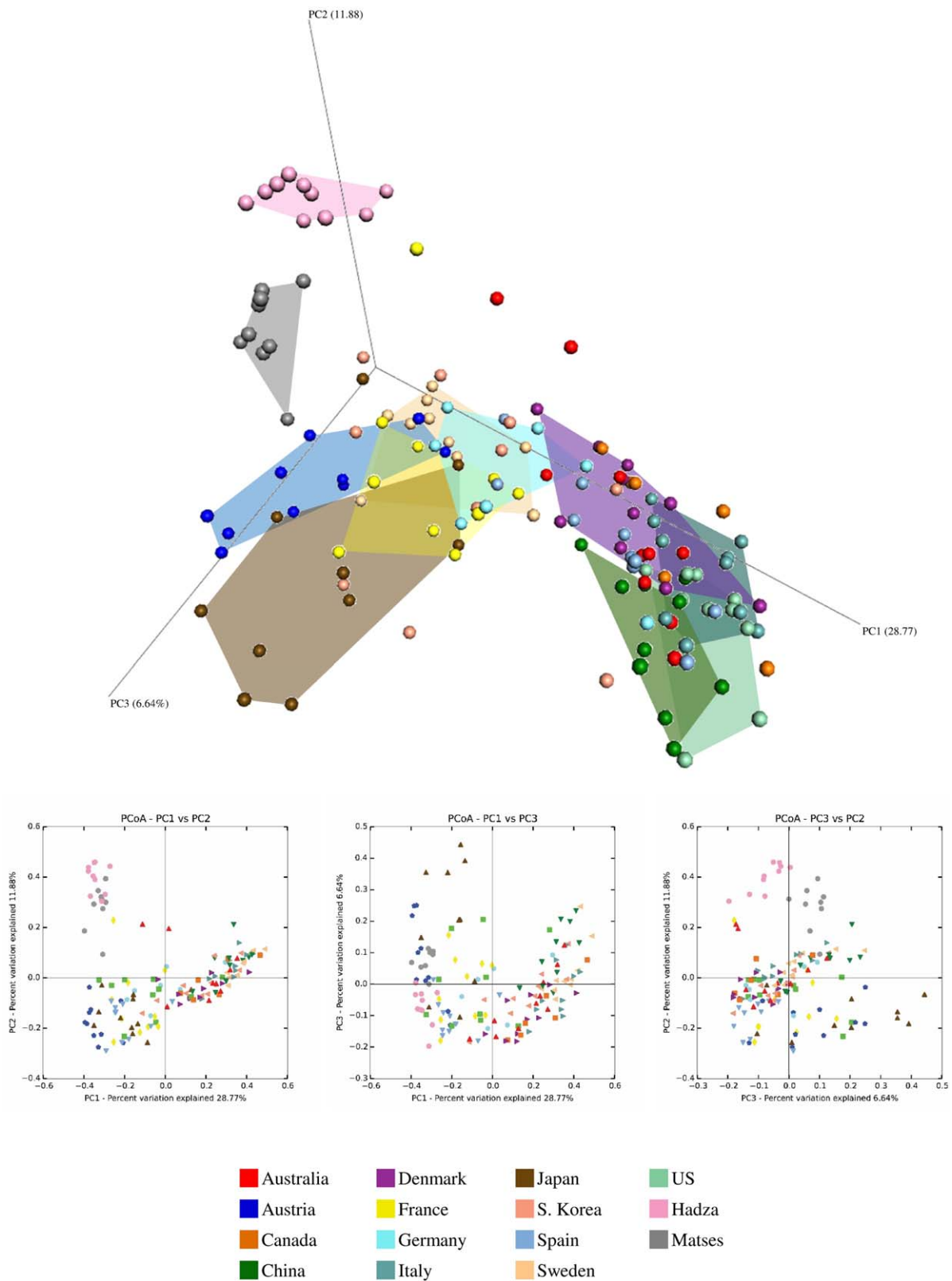
Fig. 1. Evaluation of the beta-diversity in the 142 analysed samples. The predicted PCoA is reported through two three-dimensional images as well as two-dimensional sections. The Panel depicts the beta-diversity of the samples subdivided according to their geographical origin. Colored areas highlight the main identified clusters. [Colour figure can be viewed at wileyonlinelibrary.com]

as commonly used in rural communities (Wu *et al.*, 2011).

Matses samples are characterized by low abundance of *Bacteroidetes* (3.91%) and high abundance of the phylum *Euryarchaeota* (15.03%), represented mainly by the *Methanobrevibacter* genus (14.90%) (Supporting Information Fig. S2). Members of this genus, such as *Methanobrevibacter smithii*, have been shown to increase the efficiency of energy extraction from dietary polysaccharides with consequent impact on host energy harvest (Samuel *et al.*, 2007). Therefore, *M. smithii* has been proposed as a target to reduce energy harvest in obese individuals (Samuel *et al.*, 2007).

Moreover, individuals from industrialized countries were shown to elicit a higher abundance of *Alistipes* genus (average of 5.12%) as compared with the two hunter-gatherer populations (average of 0.02%) (p-value < 0.01). This is in accordance with previous reports, suggesting a correlation between the *Bacteroides* enterotype and the presence of *Alistipes* (Wu *et al.*, 2011). Furthermore, the two hunter-gatherer communities show a higher abundance of the *Phascolarctobacterium* genus (average of 7.03%) compared with the abundance of this genus in individuals from industrialized nations (average of 0.35%) (p-value < 0.01), being notably absent from US, Swedish and Asian microbiome datasets. Members of the genus *Phascolarctobacterium* are known to produce high amounts of the short chain fatty acids (SCFA) acetate and propionate (Watanabe *et al.*, 2012). The majority of SCFA in the gut is derived from bacterial fermentation of complex carbohydrates present in this body compartment, such as dietary soluble fibres or resistant starch, which represent two of the main glycan components of the hunter-gatherer community diet (Watanabe *et al.*, 2012).
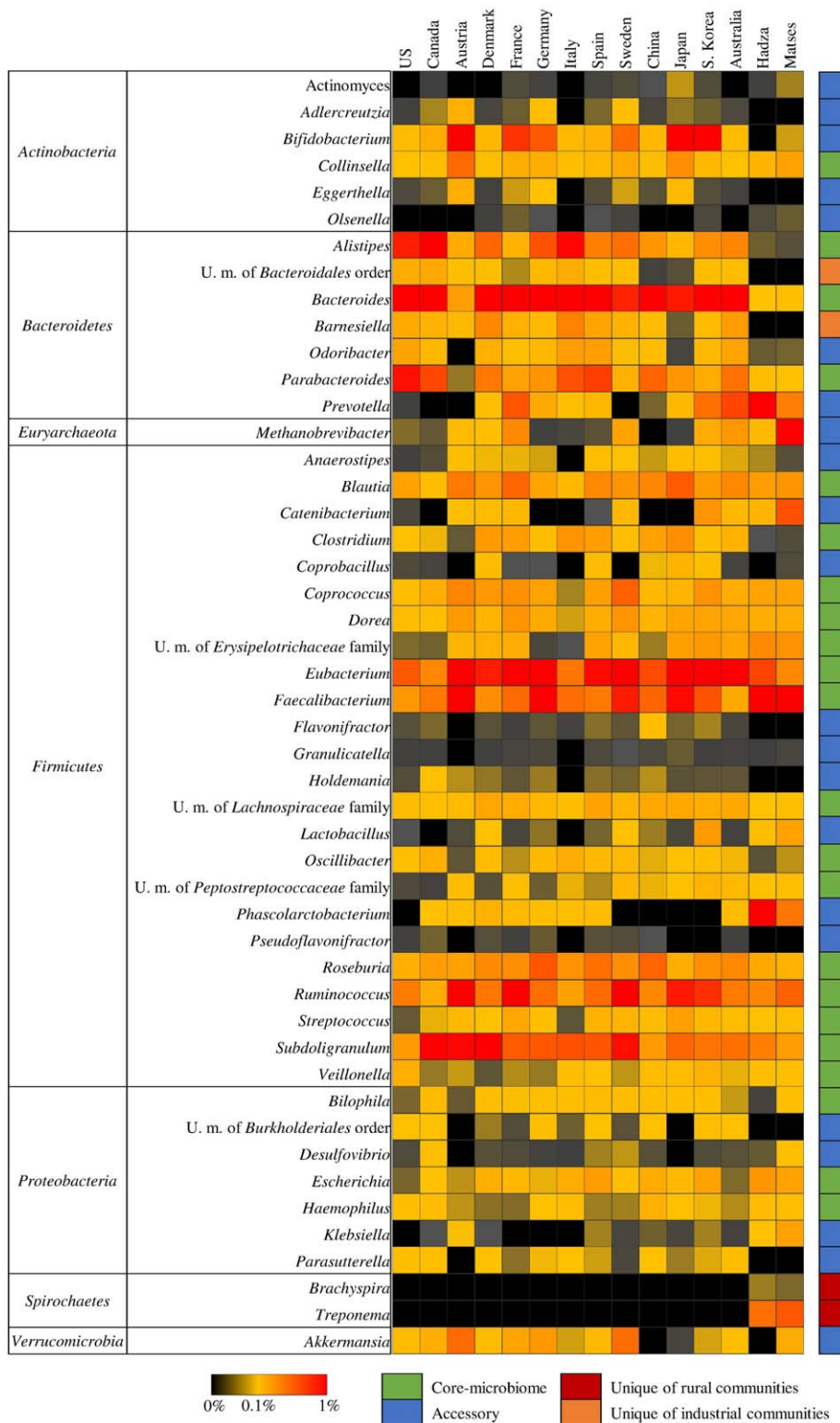
The 'pan-microbiome' of the human gut

Comparisons between different microbiomes are known to allow identification of bacterial taxa that are unique and thus characteristic of a specific metagenomic sample as well as those microbial groups that are commonly shared by all the microbiomes (Li *et al.*, 2014). Such analyses are thus important in the reconstruction of the so-called 'pan-microbiome' (Leung *et al.*, 2016). Interestingly, the pan-microbiome of WGMD reconstructed by means of MetaPhlAn2 software consists of 147 bacterial genera of which 48 were found to be present in > 90% of the samples in at least one population (Fig. 2).

The profiling data sets allow the identification of the human intestinal core microbiota, which represents bacteria that are present among all analysed human populations (Salonen *et al.*, 2012). While the number of core taxa will be reduced with the addition of future shotgun datasets covering new populations across the world, we believe that our iteration represents the most current and accurate reconstruction of the human core gut microbiota. These microorganisms are presumed to include evolutionary selected symbionts that strictly co-evolved with human beings and exert key roles in the biology of their host, such as added metabolic abilities, pathogen resistance and enhanced immune functionalities (Salonen *et al.*, 2012). Comparative analyses led to the identification of 22 genera present in all fecal samples of the 15 populations included in this study, thus representing the global human core microbiota (Fig. 2). Interestingly, in the urban-industrialized and pre-agricultural communities, this core microbiota encompass on average 82.13% ± 8.35% and 46.95% ± 6.34% (p-value < 0.01), respectively, of the total gut microbiota (Fig. 2). Furthermore, the analysis of the 'pan-microbiome' shows 15 accessory genera, which are not present in all populations. These taxa represent 11.63% ± 9.40% and 39.65% ± 9.20% of the microbiota in individuals living in urbanized-industrialized and pre-agricultural environments respectively (Fig. 2). Notably, these findings highlight that the gut microbiota of Hadza and Matses populations possess higher biodiversity, thus supporting the notion that urbanization/industrialization somehow caused a simplification of the microbial gut community (Schnorr *et al.*, 2014).

Furthermore, the pre-agricultural communities Hadza and Matses were found to be characterized by the presence of certain genera, that are absent in all other analysed populations, examples of which are *Treponema* and *Brachyspira* with average relative abundance of 4.95% and 0.04% respectively (Fig. 2). Notably, while *Treponema* is generally linked to infectious diseases (Giacani and Lukehart, 2014), a recent study reported that members of this genus may play a functional role in nutrient extraction from fibrous foods typically abundant in pre-agricultural diets (Schnorr *et al.*, 2014). Intriguingly, despite the limited availability of datasets from pre-agricultural populations, the present meta-analysis highlights that these taxa may represent microorganisms that were lost during urbanization/industrialization. Future integration of these datasets with additional shotgun metagenomics data from pre-agricultural populations will be necessary in order to confirm these results. To explore the functional roles exerted by these possibly lost gut microbiota members, targeted genome reconstruction and functional characterization of the main identified representative of this genus, i.e. *Treponema succinifaciens*, starting from shotgun metagenomics data collected for Hadza and Matses populations, allow us to obtain insights into the predicted metabolic properties of these bacteria (see below).

Fig. 2. The global human gut 'Pan-microbiome' at genus level. The heat map shows the relative abundance of bacterial taxa observed in 90% of the samples in at least one population. The right column indicates a classification of the taxa based on cell color: green cells represent taxa of the core-microbiome, blue cells indicate the accessory genera, and red and orange cells represent unique taxa of rural-primitive and urbanized-industrialized communities respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

In contrast, the taxa Unclassified member of *Bacteroidales* order and *Barnesiella* (represented mainly by the species *Barnesiella intestinihominis*) are identified only in populations living in developed countries, suggesting that their presence was promoted by the urbanization/industrialization process.

Global microbiome functional characterization

The WGMD analysis also provided an updated snapshot of the global gut microbiome functionalities. *In silico* analyses of the predicted glycobiomes of shotgun metagenomic reads based on the Carbohydrate-Active enZYmes (CAZy) database, which includes the overall enzyme collection involved in glycan metabolism (Lombard *et al.*, 2014), allowed the detection of significant differences in relative abundance of Glycoside Hydrolase families (GH) between the communities included in the WGMD. Interestingly, Matses and Hadza communities show reduced abundance of GH-encoding reads (average of 1.18%) as compared with urban-industrialized populations (average of 2.92%) (Fig. 3a). Moreover, the microbiomes of hunter-gatherer communities possess a lower number of GH families (an average of 26) compared with those of the industrialized populations (an average of 46), based on GH families with an abundance of $> 0.01\%$ of the total GH pool. A detailed analysis of these predicted glycobiomes, i.e., the enzymatic arsenal involved in the metabolism of carboydrates (Lombard *et al.*, 2014), revealed that industrialized communities encompass a higher abundance (an average of 1.35%) of GH families dedicated to the breakdown of multiple carbohydrate substrates (e.g. GH2, GH3 and GH43) compared with pre-agriculture microbiomes (with average of 0.50%) (*p*-value < 0.05), which may be the consequence of the former group following a diet that is more diverse in glycan content. The described diet of pre-agricultural communities (Supporting Information Table S3) predominantly consists of a limited variety of tubers, vegetables and fruits gathered from the surrounding environment, while consumption of meat and eggs is infrequent (Schnorr *et al.*, 2014). Notably, the advent of agriculture and industrialization brought novel foods and advanced food processing technologies developed for food making as well as preservation and nutrition enhancement, which in turn led to new food and nutrient combinations that were absent in pre-agricultural human diets (Cordain *et al.*, 2005). Moreover, industrialization led to the development of global markets that further extended the variety of tubers, vegetables and fruits of the Western diet (Popkin *et al.*, 2012). In this context, our findings indicate that, even if the diet of pre-agricultural communities primarily consists of vegetables and fruits, it is limited in its glycan variety as compared with a typical Western diet (Cordain *et al.*, 2005), which may be responsible for a low GH diversity in the former case (Fig. 3a). This is in contrast with previous observations derived from limited comparisons between small cohorts of samples (Obregon-Tito *et al.*, 2015; Rampelli *et al.*, 2015) and highlights the critical role of a comprehensive and world-wide meta-analysis.

Pathway prediction of the WGMD gut microbiomes, as based on the MetaCyc database (Caspi *et al.*, 2012), show that, on average, $1.4\% \pm 0.47\%$ (Fig. 3b) of the microbiome data retrieved from pre-agricultural communities is involved in carbohydrate degradation (Supporting Information Table S4), constituting 30 different pathways with an abundance of $> 0.01\%$. In contrast, gut microbiome metadata obtained from communities living in developed countries contained an average of $2.15\% \pm 0.65\%$ (Fig. 3b) reads encoding an average of 38 carbohydrate utilization pathways with an abundance of $> 0.01\%$. Furthermore, Hadza and Matses populations contain on average a lower number of pathways involved in alginate (-98.68%), carrageenan (-100.00%) and gellan (-77.62%) degradation, as compared with urban populations. Alginate, carrageenan and gellan gum are three polysaccharides, which are derived from algae or bacteria, and which are widely used in the food industry as stabilizers, thickeners or emulsifying agents (Brownlee *et al.*, 2005; Burges Watson, 2008; Prajapati *et al.*, 2013).

Moreover, profiling of pathways involved in fatty acid and lipid metabolism indicates on average a lower level of triacylglycerol degradation in hunter-gatherer communities (-96.47%) as compared with industrialized populations, possibly reflecting the lower (animal-derived) fat intake of pre-agricultural populations (Supporting Information Table S3) (Cordain *et al.*, 2005).

These differences reinforce the notion that increased diet variety, as associated with an urbanized life style, has shaped the human microbiome towards an expansion of its carbohydrate degradative abilities.

In order to map and characterize the resistome of the microbial consortia residing in the gut of humans living in different geographic regions, the shotgun metagenomic datasets encompassing the WGMD were also screened for known bacterial Antibiotic Resistance Enzymes (AREs) (McArthur *et al.*, 2013) (Fig. 3c). *In silico* analyses show a high abundance of AREs in Japanese ($0.49\% \pm 0.10\%$), Korean ($0.30\% \pm 0.02\%$), Austrian ($0.32\% \pm 0.03\%$), Swedish ($0.31\% \pm 0.02\%$) and Italian individuals ($0.43\% \pm 0.05\%$), while the remaining eight industrialized countries display a somewhat lower average ARE abundance of $0.25\% \pm 0.03\%$ (*p*-value < 0.05) (Fig. 3c). Interestingly, the data obtained from Tanzanian and Peruvian individuals indicate a significantly lower abundance of AREs (average of 0.19%)
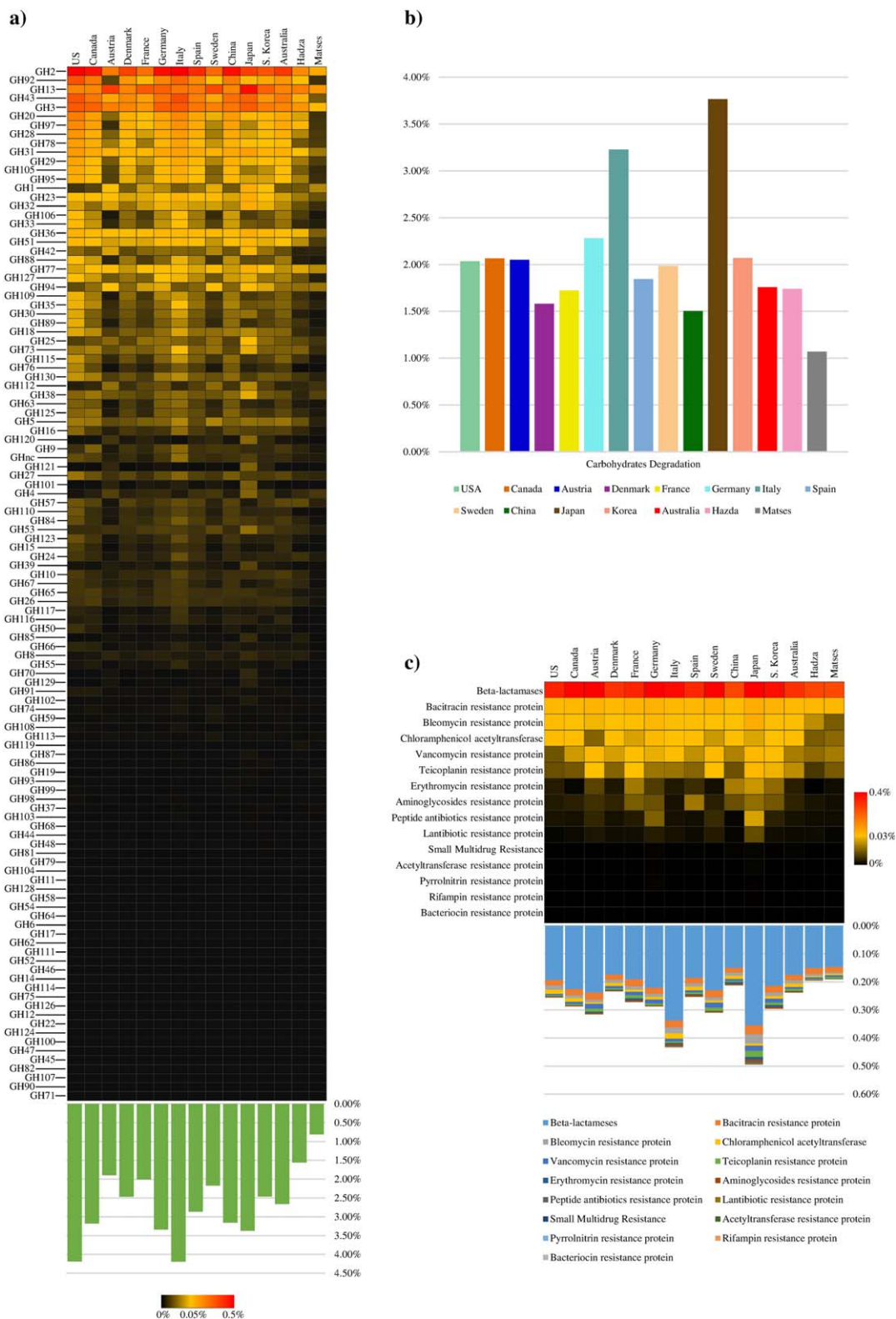
Fig. 3. Functional profile of the global human gut microbiome. Panel a shows a heat map reporting the abundance of glycosyl hydrolases in the various datasets, with a bar plot of their total abundance at the bottom. Panel b reports the % of the microbiome encoding pathways involved in carbohydrate degradation for each nation. Panel c exhibits AREs in the analysed populations, with a bar plot of the total abundance of these enzymes reported at the bottom. [Colour figure can be viewed at wileyonlinelibrary.com]
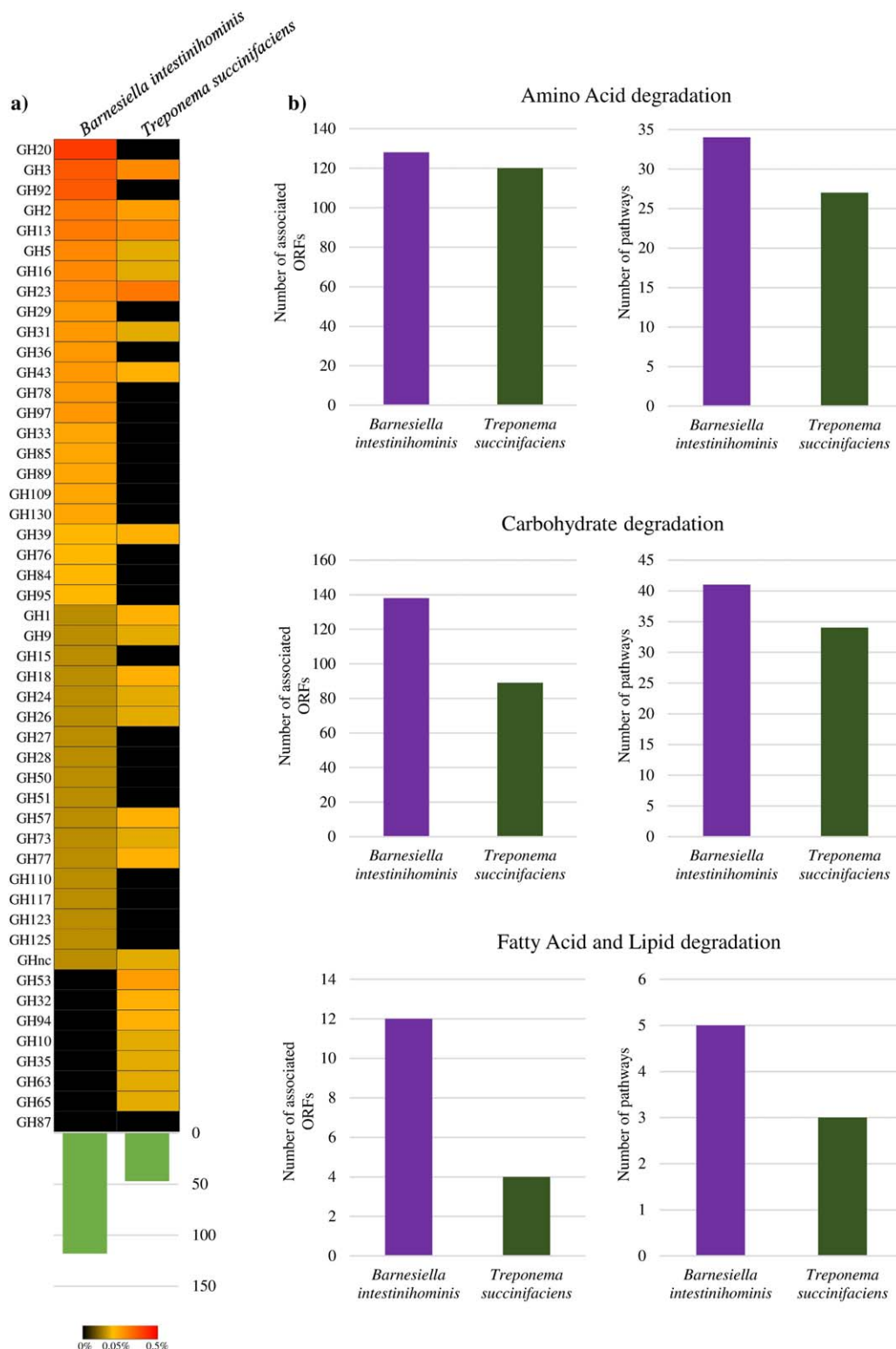
Fig. 4. Functional analysis of the reconstructed genome of *Barnesiella intestinihominis* and *Treponema succinifaciens*. Panel a indicates the abundance of glycoside hydrolases in *B. intestinihominis* and *T. succinifaciens*. The total number of GH-encoding genes is reported at the bottom. Panel b displays the number of genes and pathways involved in Amino Acid degradation, Carbohydrate degradation and Fatty Acid and Lipid degradation predicted in the analysed *B. intestinihominis* and *T. succinifaciens* genomes. [Colour figure can be viewed at wileyonlinelibrary.com]

compared with urban populations (p-value < 0.05). These results are in accordance with the very rare use of antibiotics by these pre-agricultural communities, consistent with a previous study (Clemente *et al.*, 2015).

Moreover, β-lactamase-encoding genes are the most abundant in all analysed samples, ranging from $0.15\% \pm 0.01\%$ to $0.36\% \pm 0.07\%$ in Matses and Japanese populations, respectively, possibly reflecting the fact that β-lactam-based antibiotics are the most widespread in both nature and common medical practice (Zeng and Lin, 2013).

Genome reconstruction of bacterial strains from shotgun metagenomic data

A valuable approach to assess functional roles exerted by bacterial taxa that were lost and/or acquired during urbanization/industrialization, is the targeted genome reconstruction [through the MEGAnnotator pipeline (Lugli *et al.*, 2016)] of unique gut commensals identified in the various gut microbiomes of WGMD. Such analyses targeted the reconstruction of *Treponema succinifaciens*, which has been uniquely identified in the gut microbiomes of pre-agricultural populations, as well as that of *Barnesiella intestinihominis*, which appears to be specifically present in the gut microbiomes of humans from developed countries.

Reconstruction of the *T. succinifaciens* genome was carried out using the Matses datasets due to its higher relative abundance in this population compared with that of Hadza datasets, and resulted in 97 contigs with a total length of 3.12 Mbp. Danish datasets were employed to reconstruct the *B. intestinihominis* genome, which generated 326 contigs encompassing 3.01 Mbp. *In silico* prediction of the glycobiome and identification of the pathways for degradation of carbohydrates and amino acids by means of the CAZy and MetaCyc databases provide detailed insights into the carbohydrate degradation capabilities of these two species (Fig. 4a).

Notably, glycobiome prediction revealed that *T. succinifaciens* and *B. intestinihominis* harbor 47 and 118 GH-encoding genes respectively (Fig. 4a). Furthermore, profiling of pathways for amino acid and carbohydrate degradation predicts that *B. intestinihominis* possesses a larger number of genes and pathways associated with the degradation of carbohydrates and amino acids, as well as with the metabolism of fatty acids and lipids as compared with *T. succinifaciens* (Fig. 4b).

These findings appear to reflect dietary differences between pre-agricultural and urbanized-industrialized communities (Cordain *et al.*, 2005). In fact, *B. intestinihominis* has only been identified in urbanized populations with a diet characterized by an increased diversity in polysaccharides and an augmented richness in proteins and lipids as compared with the Matses and Hadza pre-agricultural communities, which have been reported to follow a very simple (i.e. undiversified) diet (Cordain *et al.*, 2005). Functional data sets collected from human populations living in different countries across the globe highlight for the first time that changes in lifestyle and diet as caused by urbanization/industrialization appears to have created a selective pressure towards an expansion of the microbiome by acquisition of genes responsible for degradation and metabolism of a wider range of polysaccharides. In addition, an increase in protein and lipid intake also seems to have caused a marked expansion of the genetic repertoire involved in the metabolism of these energy sources.

Conclusions

In this mini-review, we explored the possibility that urbanization/industrialization processes have shaped the gut microbiomes as determined by a meta-analysis of various metagenomic datasets obtained from fecal samples of healthy human adults living in different countries across the world, including two pre-agricultural communities. There has been a considerable paucity of information regarding this topic. In fact, despite the increasing amount of scientific literature concerning the characterization of human gut microbiomes, very few are based on shotgun metagenomic sequencing focused on specific communities, while they typically do not involve analysis of datasets obtained from different geographical regions (Qin *et al.*, 2012; Karlsson *et al.*, 2013; Lim *et al.*, 2014; Feng *et al.*, 2015; Voigt *et al.*, 2015; Raymond *et al.*, 2016). The reconstructed WGMD allows the prediction of the global human gut core-microbiome, which was shown to encompass 22 genera present in the 18 datasets analysed, and representative of 15 nations. The comparison between pre-agricultural vs. urbanized/industrialized gut microbiomes of WGMD allowed the identification of particular taxa that seem to have been acquired (*Barnesiella intestinihominis*) or lost (*Treponema succinifaciens*) during the urbanization/industrialization process, perhaps as a result of dietary changes. In this context, the increase in dietary diversity and in protein and lipid intake appears to have caused an expansion of the metabolic capabilities of the human gut microbiome towards degradation of a wider range of polysaccharides and utilization of a higher number of available amino acids and lipids. However, the lack of metadata regarding dietary habits and lifestyle of the assessed populations prevents us to make any connection between diet and microbiome composition. Furthermore, antibiotic resistance profiling in the analysed datasets underlined a progressive increase in AREs proportional to the level of

urbanization/industrialization and corresponding intensity of antibiotic treatment.

As demonstrated in this mini-review, development and integration of worldwide metagenomic databases, e.g. the WGMD, will be pivotal for comprehensive studies combining novel metagenomic datasets and previous data from literature. In this context, meta-analyses of metagenomic datasets will aim to overcome the extensive fragmentation that still characterizes metagenomic studies, thus resulting in a more comprehensive overview of this rapidly expanding research area.

Acknowledgements

References

Brownlee, I.A., Allen, A., Pearson, J.P., Dettmar, P.W., Havler, M.E., Atherton, M.R., and Onsoyen, E. (2005) Alginate as a source of dietary fiber. *Crit Rev Food Sci Nutr* **45:** 497–510.

Burges Watson, D. (2008) Public health and carrageenan regulation: a review and analysis. *J Appl Phycol* **20:** 505–513.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7:** 335–336.

Caspi, R., Altman, T., Dreher, K., Fulcher, C.A., Subhraveti, P., Keseler, I.M., *et al.* (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **40:** D742–D753.

Clemente, J.C., Pehrsson, E.C., Blaser, M.J., Sandhu, K., Gao, Z., Wang, B., *et al.* (2015) The microbiome of uncontacted Amerindians. *Sci Adv* **1:** pii e1500183.

Conlon, M.A., and Bird, A.R. (2015) The impact of diet and lifestyle on gut microbiota and human health. *Nutrients* **7:** 17–44.

Cordain, L., Eaton, S.B., Sebastian, A., Mann, N., Lindeberg, S., Watkins, B.A., *et al.* (2005) Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr* **81:** 341–354.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., *et al.* (2010) Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A* **107:** 14691–14696.

Dehingia, M., Devi, K.T., Talukdar, N.C., Talukdar, R., Reddy, N., Mande, S.S., *et al.* (2015) Gut bacterial diversity of the tribes of India and comparison with the worldwide data. *Sci Rep* **5:** 18563.

Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., *et al.* (2015) Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun* **6:** 6528.

Giacani, L., and Lukehart, S.A. (2014) The endemic treponematoses. *Clin Microbiol Rev* **27:** 89–115.

Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergstrom, G., Behre, C.J., Fagerberg, B., *et al.* (2013) Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498:** 99–103.

Leung, M.H., Wilkins, D., and Lee, P.K. (2016) Erratum: Insights into the pan-microbiome: skin microbial communities of Chinese individuals differ from other racial groups. *Sci Rep* **6:** 21355.

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., *et al.* (2014) An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* **32:** 834–841.

Lim, M.Y., Rho, M., Song, Y.M., Lee, K., Sung, J., and Ko, G. (2014) Stability of gut enterotypes in Korean monozygotic twins and their association with biomarkers and diet. *Sci Rep* **4:** 7348.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42:** D490–D495.

Lugli, G.A., Milani, C., Mancabelli, L., van Sinderen, D., and Ventura, M. (2016) MEGAnnotator: a user-friendly pipeline for microbial genomes assembly and annotation. *FEMS Microbiol Lett* **363:** pii fnw049.

Martinez, I., Stegen, J.C., Maldonado-Gomez, M.X., Eren, A.M., Siba, P.M., Greenhill, A.R., and Walter, J. (2015) The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. *Cell Rep* **11:** 527–538.

McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., *et al.* (2013) The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother* **57:** 3348–3357.

Milani, C., Hevia, A., Foroni, E., Duranti, S., Turroni, F., Lugli, G.A., *et al.* (2013) Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. *PLoS One* **8:** e68739.

Nishijima, S., Suda, W., Oshima, K., Kim, S.W., Hirose, Y., Morita, H., and Hattori, M. (2016) The gut microbiome of healthy Japanese and its microbial and functional uniqueness. *DNA Res* **23:** 125–133.

Obregon-Tito, A.J., Tito, R.Y., Metcalf, J., Sankaranarayanan, K., Clemente, J.C., Ursell, L.K., *et al.* (2015) Subsistence strategies in traditional societies distinguish gut microbiomes. *Nat Commun* **6:** 6505.

Popkin, B.M., Adair, L.S., and Ng, S.W. (2012) Global nutrition transition and the pandemic of obesity in developing countries. *Nutr Rev* **70:** 3–21.

Prajapati, V.D., Jani, G.K., Zala, B.S., and Khutliwala, T.A. (2013) An insight into the emerging exopolysaccharide gellan gum as a novel polymer. *Carbohydr Polym* **93:** 670–678.

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.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., *et al.* (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490:** 55–60.

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13:** 341.

Rampelli, S., Schnorr, S.L., Consolandi, C., Turroni, S., Severgnini, M., Peano, C., *et al.* (2015) Metagenome sequencing of the Hadza Hunter-gatherer gut microbiota. *Curr Biol* **25:** 1682–1693.

Raymond, F., Ouameur, A.A., Deraspe, M., Iqbal, N., Gingras, H., Dridi, B., *et al.* (2016) The initial state of the human gut microbiome determines its reshaping by antibiotics. *ISME J* **10:** 707–720.

Round, J.L., and Mazmanian, S.K. (2009) The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9:** 313–323.

Salonen, A., Salojarvi, J., Lahti, L., and de Vos, W.M. (2012) The adult intestinal core microbiota is determined by analysis depth and health status. *Clin Microbiol Infect* **18(Suppl 4):** 16–20.

Samuel, B.S., Hansen, E.E., Manchester, J.K., Coutinho, P.M., Henrissat, B., Fulton, R., *et al.* (2007) Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. *Proc Natl Acad Sci U S A* **104:** 10643–10648.

Schnorr, S.L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., *et al.* (2014) Gut microbiome of the Hadza hunter-gatherers. *Nat Commun* **5:** 3654.

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., *et al.* (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* **12:** 902–903.

Voigt, A.Y., Costea, P.I., Kultima, J.R., Li, S.S., Zeller, G., Sunagawa, S., and Bork, P. (2015) Temporal and technical variability of human gut metagenomes. *Genome Biol* **16:** 73.

Watanabe, Y., Nagai, F., and Morotomi, M. (2012) Characterization of Phascolarctobacterium succinatutens sp. nov., an asaccharolytic, succinate-utilizing bacterium isolated from human feces. *Appl Environ Microbiol* **78:** 511–518.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., *et al.* (2011) Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334:** 105–108.

Xu, J., and Gordon, J.I. (2003) Honor thy symbionts. *Proc Natl Acad Sci U S A* **100:** 10452–10459.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., *et al.* (2012) Human gut microbiome viewed across age and geography. *Nature* **486:** 222–227.

Zeller, G., Tap, J., Voigt, A.Y., Sunagawa, S., Kultima, J.R., Costea, P.I., *et al.* (2014) Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* **10:** 766.

Zeng, X., and Lin, J. (2013) Beta-lactamase induction and cell wall metabolism in Gram-negative bacteria. *Front Microbiol* **4:** 128.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Microbial profiling of the 142 samples at phylum level. Sample names and origins are reported in Supporting Information Tables S1.

Fig. S2. Taxonomic profiling of the 142 analyzed samples at genus level. Sample names and origins are reported in Supporting Information Tables S1.

Table S1. Summary of the observations discussed in this study, accompanied by a comparison with previously published data

Table S2. List of publicly available data used in this study.

Table S3. Diet of pre-agricultural communities.

Table S4. List of pathway involved in carbohydrate degradation detected in analysis.