J Physiol 587.17 (2009) pp 4153–4158

SYMPOSIUM REVIEW

The core gut microbiome, energy balance and obesity

Peter J. Turnbaugh and Jeffrey I. Gordon

Center for Genome Sciences, Washington University School of Medicine, St Louis, MO 63108, USA

Metagenomics is an emerging field focused on characterizing the structures, functions and dynamic operations of microbial communities sampled in their native habitats without the need for culture. Here, we present findings from a 16S rRNA gene sequence- and whole community DNA shotgun sequencing-based analysis of the adult human gut microbiomes of lean and obese mono- and dizygotic twins. Our findings indicate that a core microbiome can be found at the gene level, despite large variation in community membership, and that variations from the core are associated with obesity. These findings have implications for ongoing Human Microbiome Project(s), and highlight important challenges to the field of metagenomics.

(Received 18 April 2009; accepted after revision 28 May 2009; first published online 2 June 2009)
Corresponding author J. I. Gordon: Washington University School of Medicine, 4444 Forest Park, Campus Box 8510, St Louis, MO 63108, USA. Email: jgordon@wustl.edu

The human 'metagenome' is a composite of *Homo sapiens* genes and genes present in the genomes of the trillions of microbes that colonize our adult bodies (the 'microbiome'). Our largest collection of microbes resides in the gut, where an estimated 10–100 trillion organisms reside (the gut microbiota; see Table 1 for a glossary of terms). The gut microbiome encodes metabolic capacities that remain largely unexplored but include the degradation of otherwise indigestible components of our diet (Backhed *et al.* 2005; Sonnenburg *et al.* 2005; Flint *et al.* 2008).

Colonization of adult germ-free mice with a distal gut microbial community harvested from conventionally raised mice produces a dramatic increase in body fat within 10–14 days, despite an associated decrease in food consumption (Backhed *et al.* 2004). This change involves several linked mechanisms including microbial fermentation of dietary polysaccharides that cannot be digested by the host, subsequent intestinal absorption of monosaccharides and short-chain fatty acids, their conversion to more complex lipids in the liver; and microbial regulation of host genes that promote deposition of the lipids in adipocytes (Backhed *et al.* 2004). Additionally, germ-free mice are resistant to diet-induced obesity caused by consumption of a high-fat/high-sugar 'Western' diet (Backhed *et al.* 2007). Other studies of lean and genetically obese (*ob*/*ob*) mice (Ley *et al.* 2005; Turnbaugh *et al.* 2006), as well as lean and genetically obese (*fa*/*fa*) rats (Waldram *et al.* 2009) have revealed differences in their metabotypes that have been ascribed in part to differences in their gut microbial ecology (these differences involve the representation of members of the Bacteroidetes, Firmicutes and Actinobacteria phyla of bacteria).

Although the primary cause of obesity is excess caloric intake compared with expenditure, one intriguing hypothesis that follows from these studies links differences in gut microbial ecology between humans to energy homeostasis; i.e. individuals with a microbial community more efficient at energy extraction from the diet, or with an increased ability to promote adiposity through manipulation of host genes and metabolism, may be predisposed to obesity. This hypothesis predicts that obese and lean individuals will have distinct microbiotas, with measurable differences in their ability to extract energy from their diet and to deposit that energy in fat.

Additionally, there are a number of unresolved questions regarding the organismal and genetic diversity of the human gut microbiome. How diverse is the human gut microbiome between individuals and how does it vary within an individual over time? How is a microbiota and microbiome transmitted to an individual following birth: what is the relative role of early environmental exposures (to microbes and diets) *versus* our human genotype in defining postnatal microbial community assembly, and the structures of our adult microbiota? Is there a core microbiome shared between humans, and should this core be defined in terms of organisms, genes,

This review was presented at *The Journal of Physiology* Symposium on *The world within – impact of the intestinal microbiota on whole body physiology and pathophysiology*, which took place at Experimental Biology 2009 at New Orleans, LA, USA on 20 April 2009. It was commissioned by the Editorial Board and reflects the views of the authors.

DOI: 10.1113/jphysiol.2009.174136

Table 1. Glossary

Metagenomics. A rapidly evolving field that emerged from rapid advances in DNA sequencing methods, with a focus on the use of culture-independent methods to study the structures, functions and dynamic operations of microbial communities. Frequently used to refer to the technique of isolating and sequencing community DNA directly from an environment, but extends to the profiling of gene expression at the level of RNA (metatranscriptomics), and protein (metaproteomics), as well as community metabolism (metabolomics).

Microbiota. A microbial community; commonly referred to according to the habitat that it occupies (e.g. the gut microbiota).

Microbiome. The aggregate genomes and genes found in the members of a microbiota.

Phylotype (or operational taxonomic unit, OTU). A group of microbes, commonly defined by the level of sequence similarity (percentage identity) between small subunit (16S) rRNA genes (e.g. $\geq$ 97% for a 'species'-level phylotype).

'Core' human gut microbiome. A set of features shared across all or the vast majority of gut microbiomes (e.g. genes and/or metabolic capabilities).

or functional characteristics? And finally, are there genes and/or metabolic pathways in the human microbiome that can be identified as being associated with obesity?

To begin to address these questions, we conducted a large-scale analysis of the faecal microbiotas and microbiomes of 154 individuals: this group consisted of 31 monozygotic and 23 dizygotic young adult female twin-pairs and the majority of their mothers ($n = 46$). Twins were either concordant for obesity or leanness. Faecal samples were obtained from all study participants at $t = 0$ and from 127 individuals ~2 months later. Although variations in microbial community composition along the length and width of the gut remain poorly defined in humans, we chose to analyse the faecal microbiota. We did so for several reasons: sample collection is non-invasive; samples contain a large biomass of microbes; and the faecal microbiota is representative of interpersonal differences in distal gut microbial ecology (Eckburg *et al.* 2005). Combined, our analysis yielded 9920 near full-length and 1,937,461 partial bacterial 16S rRNA sequences (the latter representing data from the gene's V2 and V6 regions), plus 2.14 gigabases from their microbiomes (Turnbaugh *et al.* 2009).

The results revealed that the human gut microbiome is shared among family members, but that each person's gut microbial community varies in the specific bacterial lineages present, with a comparable degree of co-variation between adult monozygotic (MZ) and dizygotic (DZ) twin pairs. Remarkably, there was not a single abundant (defined as $> 0.5\%$ of the community) bacterial species shared by all 154 individuals. However, there was a wide array of shared microbial genes among the sampled population: together, they comprise an extensive, identifiable 'core microbiome' at the gene, rather than at the organismal lineage level. Obesity was associated with phylum-level changes in the microbiota, reduced bacterial diversity, and altered representation of bacterial genes and metabolic pathways, including those involved in nutrient harvest. These results emphasize the importance of early environmental exposures in shaping the human gut microbiota (Table 2), demonstrate that a diversity of organismal assemblages can nonetheless yield a core microbiome at a functional level, and show that deviations from this core are associated with different physiological states (obese compared with lean) (Turnbaugh *et al.* 2009).

Implications for future studies of the gut microbiome

Our results have implications for a number of recently initiated Human Microbiome Projects which seek to explore the microbiomes associated with various body habitats including nose, mouth, skin, gut and vagina (e.g. the NIH Roadmap HMP and the EU MetaHIT). One of the initial goals of the HMP was to generate a set of reference genomes and metagenomes. However, given the high levels of inter-individual variation in body habitat-associated microbial communities, and the limited availability of cultured representatives of lineages present in these communities, it is likely that only a minority of abundant 'species'-level phylotypes found in individuals will be characterized by sequencing in the next few years. Likewise, defining a single reference human microbiome is not a reasonable goal for most body-associated habitats; however, it is likely that gene families can be identified that encode common and abundantly represented functional features expressed by microbial communities that inhabit, shape, and are shaped by these ecosystems. Studies in model organisms and/or less diverse communities will probably be essential to understand basic principles that govern community assembly, diversity, robustness and function.

No other studies of lean and obese humans have been published to date that use large-scale 16S rRNA gene or metagenomic sequencing. However, a number of published reports have used qPCR and/or FISH to quantify specific taxa in lean and obese individuals. For example, a FISH-based study of overweight and obese adolescents revealed a decrease in Firmicutes and an increase in Bacteroidetes during weight loss (Nadal *et al.* 2008). A prospective Finnish study of children used qPCR to show that there was an increased representation of members of the Firmicutes in individuals who became overweight by

Table 2. Key observations and challenges

Observations	Challenges
The gut microbiota is highly variable between individuals.	Time series studies to determine the extent of variation of community membership (i) within habitats (with careful attention paid to the biogeography of the habitat), (ii) between body habitats within a given host, (iii) between hosts, and (iv) at various ages. A key question is which host variables to monitor during the collection period (e.g. diet). Knowledge of variance provides a foundation for determining the significance of associations between host physiology or pathophysiology and community membership. If intrapersonal variation was significantly less than interpersonal variation, then each individual would serve as his/her best control for correlating community ecology with host biology. Need new methods to scale up sample processing and better PCR primers to more fully capture diversity in all three domains of life when performing SSU rRNA gene-based surveys.
Family members share more similar gut microbiotas than unrelated individuals	Need more extensive family surveys, both intragenerational (mother, father, siblings) as well as intergenerational, with particular attention paid to first days/months/year of life. Environmental surveys need to be performed as well as knowledge of cultural traditions (who and how many people handle newborns).
The gut microbiota of monozyotic twin pairs have a similar degree of variance as do the gut microbiota of dizygotic twin pairs, indicating that early environmental exposures are key determinants of adult gut community structures.	Examine twins at earlier stages of life (are the gut communities of monozygotic twins more similar than dizygotic twins, indicating that host genetic factors help shape ecology). Need to understand the factors that regulate assembly of a gut microbiota (e.g. patterns of exposure to microbes, selective forces that operate in the habitat such as diet, the co-evolving immune system). This will require studies of model organisms where potentially confounding variables can be constrained, as in wild-type and genetically engineered inbred strains of gnotobiotic mice. Use microbiota transplants (from one body habitat to another in humans, or from one environmental habitat to a body habitat in gnotobiotic mice) to explore the relative contributions of legacy (history of exposure to microbes) *versus* habitat in selecting a microbial community
Despite marked interpersonal variation in species assemblages, there is an identifiable core gut microbiome composed of genes encoding various signalling and metabolic pathways. This convergence of disparate species assemblages onto common functional states is a feature of macro-ecosystems and implies considerable functional redundancy between different taxa comprising gut communities.	Extend efforts to sequence genomes of cultured representatives of different microbial lineages represented in the gut (including multiple strains of a given species-level phyotype in order to define its pan-genome). Make *in silico* predictions of niches (professions) from these annotated genomes and then conduct experimental tests of hypotheses *in vitro* and *in vivo* (e.g. with gnotobiotic mice) using defined consortia of sequenced microbes, plus metagenomic methods (DNA-, mRNA-, protein- and metabolite-level studies). The experimental results and modelling exercises that follow acquisition of these datasets should yield insights about the genetic, genomic and metabolic foundations, and ecological principles, that regulate microbial–microbial interactions (e.g. syntrophy, competition) and generate testable hypotheses about the significance of functional redundancy in body habitat-associated microbial communities. Such studies are also important to developing effective strategies for intentionally altering community structure and operations in ways that benefit the host and that are sustainable.

the age of 7 (Kalliomaki *et al.* 2008). A study of three adults following gastric bypass surgery revealed lower levels of Firmicutes in faecal samples with weight loss (Zhang *et al.* 2009). Furthermore, a qPCR-based characterization of gut microbial ecology in obese and lean pigs revealed a decreased level of Bacteroidetes in obese pigs (Guo *et al.* 2008). However, a FISH-based study of obese adults on 2 month-long diets and a FISH/qPCR-based study of pregnant overweight and lean women failed to show consistent phylum-level differences in community structure (Collado *et al.* 2008; Duncan *et al.* 2008). Given the high-level of inter-individual variation in community structure, differences in age, diet, lifestyle and geographic location, combined with potential experimental biases (e.g. which primers are used for PCR of various regions of 16S rDNA genes), it remains difficult to conclude if these discrepant findings reflect actual differences in gut microbial community structures across different cohorts.

There are a number of future experiments that could help move towards a mechanistic understanding of the influence of the gut microbiome on host energy balance. We are currently conducting studies of the microbiome associated with twins discordant for leanness or obesity. In addition, we are characterizing the faecal microbiomes of twins at younger ages, plus the microbiomes of their mothers, fathers and siblings: these latter studies will allow us to determine whether at an earlier age MZ twin pairs have lesser or similar levels of variance in their microbiotas compared to DZ twin pairs, and the degree to which microbial communities are shared among family members. Recent advances will also allow transcriptional profiling of obese and lean microbiomes, to determine if there are identifiable obesity-associated differences in gene expression. In addition to characterization of their microbiomes, detailed studies of energy balance will need to be performed in MZ twins discordant for obesity. If done in an ethically acceptable manner, similar combinations of comparative metagenomics and energy balance studies should be performed in individuals prior to, during, and after intentional weight gain on a defined diet, to test the hypothesis that differences in community structure are a cause rather than an effect of obesity.

General challenges and opportunities for metagenomics

The major challenges in the emerging field of metagenomics are experimental, computational and conceptual. A number of studies across diverse environments have begun to associate organismal and gene lineages with chemical or host meta-data (Hugenholtz & Tyson, 2008). More recently, 'metatranscriptomic' and 'metaproteomic' studies have begun to characterize gene expression, as well as the population structure of microbial communities (Ram *et al.* 2005; Lo *et al.* 2007; Frias-Lopez *et al.* 2008; Wilmes *et al.* 2008; Poretsky *et al.* 2009; Verberkmoes *et al.* 2009). However, new tools for multivariate analyses will be required to integrate these disparate data and to develop testable hypotheses about the functions of members of a complex community.

Stable isotope probing has recently been used to identify microbes involved in oxidizing single carbon (C_1) compounds in Lake Washington sediments, revealing niche partitioning in the environment (Kalyuzhnaya *et al.* 2008). This method has the potential to provide more direct links between community membership, gene content, and metabolism. Additionally, methods designed to enrich microbial clades or lineages through selective culture and/or cell sorting, as well as methods that isolate single cells (Marcy *et al.* 2007), are needed to explore the potential niches (professions) of uncultured groups identified in the gut microbiota as a result of 16S rRNA gene sequencing efforts: these groups include members of the Cyanobacteria and TM7 phyla, and even lineages within the more frequently studied Bacteroidetes and Firmicutes. These techniques, coupled with the creation of defined communities of sequenced gut microbes in genetically defined gnotobiotic animal models, reared under conditions where potentially confounding environmental factors can be constrained and/or systematically varied, could provide opportunities to better understand the metabolic activities and adaptations of complex communities.

Finally, an artificial divide has been placed between 'environmental' and 'medical' metagenomics (Ley *et al.* 2007). Lessons learned from mathematical modelling of the biogeography of ocean microbial communities (Follows *et al.* 2007), metagenomic sequencing of microbial symbionts associated with insect hosts (Wu *et al.* 2006), and population genomic characterizations of 'simple' acid mine drainage microbial consortia (Lo *et al.* 2007) need to be applied to human communities. These 'meta'-metagenomic analyses have already begun to reveal insights about the ecological and evolutionary factors that shape community membership and functional capabilities (Tringe *et al.* 2005; Lozupone & Knight, 2007; Turnbaugh *et al.* 2007; Dinsdale *et al.* 2008; Ley *et al.* 2008), and should help address key issues related to the assembly and evolution of human gut communities.

References

Backhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF & Gordon JI (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A* **101**, 15718–15723.

Backhed F, Ley RE, Sonnenburg JL, Peterson DA & Gordon JI (2005). Host-bacterial mutualism in the human intestine. *Science* **307**, 1915–1920.

Backhed F, Manchester JK, Semenkovich CF & Gordon JI (2007). Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proc Natl Acad Sci U S A* **104**, 979–984.

Collado MC, Isolauri E, Laitinen K & Salminen S (2008). Distinct composition of gut microbiota during pregnancy in overweight and normal-weight women. *Am J Clin Nutr* **88**, 894–899.

Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues C, Haynes M, Li L, McDaniel L, Moran MA, Nelson KE, Nilsson C, Olson R, Paul J, Brito BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White BA & Rohwer F (2008). Functional metagenomic profiling of nine biomes. *Nature* **452**, 629–632.

Duncan SH, Lobley GE, Holtrop G, Ince J, Johnstone AM, Louis P & Flint HJ (2008). Human colonic microbiota associated with diet, obesity and weight loss. *Int J Obes (Lond)* **32**, 1720–1724.

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE & Relman DA (2005). Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638.

Flint HJ, Bayer EA, Rincon MT, Lamed R & White BA (2008). Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol* **6**, 121–131.

Follows MJ, Dutkiewicz S, Grant S & Chisholm SW (2007). Emergent biogeography of microbial communities in a model ocean. *Science* **315**, 1843–1846.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW & Delong EF (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* **105**, 3805–3810.

Guo X, Xia X, Tang R, Zhou J, Zhao H & Wang K (2008). Development of a real-time PCR method for Firmicutes and Bacteroidetes in faeces and its application to quantify intestinal population of obese and lean pigs. *Lett Appl Microbiol* **47**, 367–373.

Hugenholtz P & Tyson GW (2008). Microbiology: metagenomics. *Nature* **455**, 481–483.

Kalliomaki M, Collado MC, Salminen S & Isolauri E (2008). Early differences in fecal microbiota composition in children may predict overweight. *Am J Clin Nutr* **87**, 534–538.

Kalyuzhnaya MG, Lapidus A, Ivanova N, Copeland AC, McHardy AC, Szeto E, Salamov A, Grigoriev IV, Suciu D, Levine SR, Markowitz VM, Rigoutsos I, Tringe SG, Bruce DC, Richardson PM, Lidstrom ME & Chistoserdova L (2008). High-resolution metagenomics targets specific functional types in complex microbial communities. *Nat Biotechnol* **26**, 1029–1034.

Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD & Gordon JI (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* **102**, 11070–11075.

Ley RE, Knight R & Gordon JI (2007). The human microbiome: eliminating the biomedical/environmental dichotomy in microbial ecology. *Environ Microbiol* **9**, 3–4.

Ley RE, Lozupone CA, Hamady M, Knight R & Gordon JI (2008). Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol* **6**, 776–788.

Lo I, Denef VJ, Verberkmoes NC, Shah MB, Goltsman D, DiBartolo G, Tyson GW, Allen EE, Ram RJ, Detter JC, Richardson P, Thelen MP, Hettich RL & Banfield JF (2007). Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature* **446**, 537–541.

Lozupone CA & Knight R (2007). Global patterns in bacterial diversity. *Proc Natl Acad Sci U S A* **104**, 11436–11440.

Marcy Y, Ouverney C, Bik EM, Losekann T, Ivanova N, Martin HG, Szeto E, Platt D, Hugenholtz P, Relman DA & Quake SR (2007). Dissecting biological 'dark matter' with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc Natl Acad Sci U S A* **104**, 11889–11894.

Nadal I, Santacruz A, Marcos A, Warnberg J, Garagorri M, Moreno LA, Martin-Matillas M, Campoy C, Marti A, Moleres A, Delgado M, Veiga OL, Garcia-Fuentes M, Redondo CG & Sanz Y (2008). Shifts in clostridia, bacteroides and immunoglobulin-coating fecal bacteria associated with weight loss in obese adolescents. *Int J Obes (Lond)*; DOI: 10.1038/ijo.2008.260.

Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP & Moran MA (2009). Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ Microbiol* **11**, 1358–1375.

Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC 2nd, Shah M, Hettich RL & Banfield JF (2005). Community proteomics of a natural microbial biofilm. *Science* **308**, 1915–1920.

Sonnenburg JL, Xu J, Leip DD, Chen CH, Westover BP, Weatherford J, Buhler JD & Gordon JI (2005). Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307**, 1955–1959.

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P & Rubin EM (2005). Comparative metagenomics of microbial communities. *Science* **308**, 554–557.

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R & Gordon JI (2009). A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R & Gordon JI (2007). The human microbiome project. *Nature* **449**, 804–810.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER & Gordon JI (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031.

Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, Lefsrud MG, Apajalahti J, Tysk C, Hettich RL & Jansson JK (2009). Shotgun metaproteomics of the human distal gut microbiota. *Isme J* **3**, 179–189.

Waldram A, Holmes E, Wang Y, Rantalainen M, Wilson ID, Tuohy KM, McCartney AL, Gibson GR & Nicholson JK (2009). Top-down systems biology modeling of host metabotype-microbiome associations in obese rodents. *J Proteome Res* **8**, 2361–2375.

Wilmes P, Andersson AF, Lefsrud MG, Wexler M, Shah M, Zhang B, Hettich RL, Bond PL, VerBerkmoes NC & Banfield JF (2008). Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *Isme J* **2**, 853–864.

Wu D, Daugherty SC, Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE, Tran PL, Moran NA & Eisen JA (2006). Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol* **4**, e188.

Zhang H, DiBaise JK, Zuccolo A, Kudrna D, Braidotti M, Yu Y, Parameswaran P, Crowell MD, Wing R, Rittmann BE & Krajmalnik-Brown R (2009). Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci U S A* **106**, 2365–2370.

Acknowledgements

This work was supported in part by the NIH (DK078669, DK70977), and the Crohn's and Colitis Foundation of America.