Analysis and Interpretation of the Human Microbiome

James J. Ashton, BMBS,[*,†] R. Mark Beattie, FRCPCH,*[*] Sarah Ennis, PhD,[†] and David W. Cleary, PhD*[‡,§]

Abstract: Microbiome research has experienced an unprecedented level of growth over the last decade. This is largely due to revolutionary developments in, and accessibility to, DNA sequencing technologies that have enabled laboratories with even modest budgets to undertake projects. Study of the human microbiome in particular has seen a surge in interest, and although a lot of time and money has been focused on health and disease, the clinical interpretation of these data and the ability of clinicians to understand these studies in the context of disease are less straightforward. Conditions such as inflammatory bowel disease, asthma, and cancer have seen a huge increase in research focused on the role of microbiome in disease pathogenesis, but the ability of clinicians to appraise and use these data is largely lacking. The purpose of this article is to provide an introduction for clinicians and nonclinicians wishing to learn about and engage in microbiome research. It details the background of microbiome research and discusses the process of generating 16S rRNA sequencing data, the most commonly used method for microbiome analysis. We discuss the interpretation of results in a clinical context, commonly used metrics for analysis and discuss future impact and direction for microbiome research. The meteoric rise of genomic medicine to the brink of routine clinical use should be seen as a blueprint for the microbiome; the ability for physicians to understand and interpret these data is vital to this growth and aiding clinicians (and researchers) to participate in further microbiome research.

(Inflamm Bowel Dis 2016;0:1–10)

Key Words: microbiome, next-generation sequencing, inflammatory bowel disease

The advent of high-throughput next-generation sequencing has revolutionized genetic research; since the completion of the human genome project in 2003, a whole human genome can now be sequenced for a fraction of the cost and in a fraction of the time. Over the last 5 to 10 years, this technology has begun to be applied to bacteria, with coining of the term microbiome to describe "the collective genomes and gene products of all microbes residing within an organism."[1]

Study of the human microbiome in the context of disease has seen a surge in interest. Although a lot of time and money has been focused on several disease areas, the clinical interpretation of these data and the ability of clinicians to interpret these studies in the context of understanding and management of disease has lagged behind. To address this gap, this review is targeted at those wanting to understand and interpret microbiome results and data; we discuss basic science aspects, techniques for investigation, analysis, and interpretation of data.

The virome and mycobiome are, respectively, the viral and fungal equivalent of microbiome,[2,3] details of which are beyond the scope of this review.

WHAT IS THE MICROBIOME AND WHY IS IT IMPORTANT?

The microbiome is "the collective genomes and gene products of the microbiota residing within an organism."[1] In normal healthy individuals, microbiota exists on all parts of the body that come into contact with "the outside world." Bacteria at certain body sites have long been believed to be involved in immune modulation/programming, in the development of disease,[4,5] and in maintaining health.[6,7] Currently, 16S rRNA sequencing for microbiome analysis has been used solely as a research tool. The potential to use rapid sequencing to characterize and understand the impact of bacteria on diseases (not classical infectious disease) is huge, and a parallel to the rapid emergence of human genetics and genomics in a clinical setting cannot be overlooked.

The Human Microbiome Project is an attempt by the National Institute of Health to characterize microbial communities at various sites on the human body.[1] This project has been ongoing since 2007 and has widely published on health and disease. There is an aim to produce reference data alongside detailing the role that the microbiome has in disease.

Received for publication February 12, 2016; Accepted March 3, 2016.

From the *Department of Paediatric Gastroenterology, Southampton Children's Hospital, Southampton, United Kingdom; †Department of Human Genetics and Genomics Medicine, University of Southampton, Southampton, United Kingdom; ‡Faculty of Medicine, Academic Unit of Clinical and Experimental Sciences, University of Southampton, Southampton, United Kingdom; and §Faculty of Natural and Environmental Sciences, Institute for Life Sciences, University of Southampton, Southampton, United Kingdom.

J. J. Ashton is funded by a University of Southampton National Institute of Health Research Academic Clinical Fellowship.

The authors have no conflict of interest to disclose.

J. J. Ashton wrote the article in conjunction with D. W. Cleary. S. Ennis and R. M. Beattie commented on the article, along with D. W. Cleary. All authors approved the article before submission.

Reprints: James J. Ashton, BMBS, Department of Paediatric Gastroenterology, Paediatric Medical Unit, University Hospital Southampton, Tremona Road, SO16 6YD Southampton, United Kingdom (e-mail: j.ashton@soton.ac.uk).

DOI 10.1097/MIB.0000000000000809

Published online.

Definitions

1. The Human Microbiome Project a National Institutes of Health (NIH) project centered on sequencing and identification of the microbiome from all body sites.[8]
2. Microbiome: the collective genomes and gene products of the microbiota residing within an organism.[1]
3. Microbiota: a collective term for a group of microscopic organisms of any (specific) region.
4. Metagenomics: analysis of microorganisms by direct extraction of DNA from all genomes within a sample.[9]
5. 16S ribosomal RNA: bacterial small ribosomal subunit; the DNA coding for this contains highly conserved areas situated between 9 highly variable regions (V1–9).[10] A number of sequence repositories are available, such as Greengenes, which is a comprehensive collection of regularly updated freesource 16S rRNA gene databases.[11]
6. Dysbiosis: an imbalance or shift in microbiota at a given body site to an altered state.[12]
7. Taxonomy: classification of organisms into groups based on their characteristics (for bacteria, this is often their genetic similarity).
8. Phylogeny: a description of evolutionary relatedness among a group of organisms.
9. Operational taxonomic unit (OTU): microbial 16S sequences are grouped together at a level of similarity (variable) to indicate a single "taxonomic unit." For example, all sequences with 94% similarity are grouped together into a single genus.[13,14]
10. Next-generation sequencing: a catch-all term used to describe several modalities of high-throughput sequencing technologies.[15]
11. Primer sequence: in the context of microbiome sequencing, these are the DNA primer sequences used to amplify a specific variable region(s) within the 16S rDNA gene.
12. FASTQ file: text-based format containing multiple nucleotide sequences each with a per-base quality scores.
13. Barcode sequence: a sample-specific DNA sequence (code or tag) that is incorporated into the DNA to be sequenced at the point of amplification. It allows linking of output sequences back to their samples of origin.
14. LINUX/UNIX: a modular operating system that uses a single "kernel" for all processes. Additional devices/programs can be added as modules. It used a command-line interface (or graphical user interface). LINUX is an example of an open-source software operating system.

WHAT SAMPLES CAN YIELD MICROBIOME DATA?

Almost any biological sample can be used to derive bacterial DNA.[8] The most commonly used samples are biopsies (GI, skin, and lung tissue), washings/scrapings (such as bronchoalveolar lavage), and fluids/bodily products (such as faeces and sputum). The source of sampling does not impact on the technology used for subsequent 16S analysis, but all samples undergo specific processing (dependant on their type) to extract bacterial DNA. Certain samples with low bacterial biomass are likely to have low yields of DNA; it is vital to be able to recognize this and interpret with adequate controls in these scenarios.

Storage of samples before processing is critical; some bacteria will lyse and their DNA will degrade, whereas some will multiply after sampling, leading to misleading or incorrect results. Generally, samples are frozen as soon as possible and some (such as tissue samples) are stored in solutions to preserve DNA/RNA (such as RNAlater [various manufacturers]).

WHICH DISEASES TYPES MAY BE OF INTEREST TO MICROBIOME RESEARCHERS

Microbiome research has not focused on classical infectious illnesses; most clinicians will be aware of targeted sequencing or PCR testing for common infectious bacteria such as *Neisseria meningitidis*. 16S sequencing is much broader and gives little or no information about subspecies and antibiotic sensitivities that are important to the clinicians. Most publications relate to disease in which host–microbe interaction is likely to play a role in disease pathogenesis and ongoing disease flare; these are mostly chronic conditions such as inflammatory bowel disease (IBD),[4,16–18] asthma,[19–21] skin conditions,[22] and cancers.[23–25]

The immunomodulatory effect of the microbiome seems likely to hold a key to understanding several diseases, and correlation of host gene expression (transcriptome) with the microbiome and microbial gene expression may yield useful insights.[16,18,26] Trials have begun to look at the effect of specific modulatory diets on the gut microbiome, largely in the context of IBD[17,27] and IBS.[28] These have revealed initial dysbiosis with changes in diversity through treatment and have started to understand the functional impact, such as on short-chain fatty acid synthesis. This uses additional software (such as PICRUSt) after initial microbiome analysis.[28,29]

Increasingly, there is interest in the role the microbiome may play in the development of obesity and type 2 diabetes,[30,31] and alongside this, there has been a small amount of work into acute conditions such as necrotizing enterocolitis in preterm infants looking for evidence of a role of bacteria in such conditions.[32–35]

NEXT-GENERATION SEQUENCING (NGS)

NGS is a catch-all term used to describe several modalities of high-throughput sequencing technologies that were developed in mid 2000s.[15] Classical DNA sequencing (Sanger sequencing[36]) was developed in the 1970s and uses chain termination by dideoxynucleotides and detection of fluorescence of these molecules. It was this technology that led to our first insights into the human genome,[37] although at great cost in both time and expense. The lack of throughput made this technology less than ideal for studies of complex microbial communities, and it was technical advances

in sequencing chemistry that led to the widespread adoption of NGS as a tool for microbiome research. Two recent reviews by Land and Loman in 2015 provide excellent insights into the development of bacterial genetic sequencing over the last 20 years.[38,39]

Illumina

Illumina sequencing by synthesis is the most widely used sequencing technology for microbiome research.[40] The method works by a reversible dye terminator technology; DNA is ligated into short sections with the addition of identification and primer sequences at either end and attached to a lane (flow cell); it is amplified using polymerase-forming DNA clusters. Fluorescently labeled reversible terminator bases (1 color for each base) are added, which fluoresce when added to the sequence. A camera captures each base as it is added and can relate this to the location of the DNA on the lane. For each sequence, several 100 reads are generated.[40] The Illumina MiSeq is typically used and has a single read length of up to 300 bp and can produce up to 50 million paired-end reads per run.

For microbiome research, primers are used in presequencing PCR to amplify a specific region of the 16S rRNA sequence. These primers include the additional primer and identification sequences mentioned above, and thus, when loaded into the instrument, this PCR product becomes the target for the sequencing reaction generating many 16S rRNA gene-specific reads that then can be related to individual bacteria.

Roche 454

Originally an incredibly popular sequencing technology for microbiome studies, the iterative increases in read length by Illumina (longer reads being the original benefit of using 454) have largely rendered it redundant.[41]

Others—IonTorrent, PacBio, and MinIon

The overwhelming majority of 16S sequencing has been done using Illumina and 454 technologies. Several other companies have their own sequencing technology; some, such as PacBio and MinIon, are real-time single-molecule platforms that may, in time, become highly useful for 16S rRNA sequencing for microbiome analysis.[42]

HOW SEQUENCING 16S rRNA ALLOWS BACTERIA TO BE IDENTIFIED?

The ubiquitous and phylogenetically stable bacterial 16S rRNA offers a very useful target for the identification bacteria down to species level in some instances. The 16S ribosomal subunit in encoded for by DNA that has highly conserved regions (very similar sequences) between all bacterial species. Between these areas are highly variable regions (V1-9) that can be used to identify a specific genus or species of bacteria through the sequence.[43–45] Universal primers are designed to amplify a specific variable region of the 16S rRNA sequence. Based on a combination of best primer sites and information gathered from variable regions, the most commonly targeted regions are V3, V4, and V6.[46] Different primer sets exist for amplification of variable regions,[47] and the choice should be based on the sample type and importance of taxonomy or phylogeny.[48,49] See Figure 1 for the structure of 16S rDNA.

After sequencing, each read of a 16S rRNA gene is clustered together with others by computer analysis into operational taxonomic units (OTUs) based on how similar the sequences are. It is prudent to remember that grouping together bacteria based on their genetic similarity is somewhat arbitrary. Historically, all species were defined on phenotypic characteristics, and then their genetic similarity was discovered. In microbiome sequencing, bacteria are clustered depending on their genetic similarity and then defined as a genus, species, etc. For example, all sequences with 94% similarity may be grouped together,[8,9] and referencing a representative sequence for the cluster (OTU) to a reference database allows for identification of the bacteria present at genus level. Grouping at 97% similarity may allow identification at a species level. The relative frequency of the sequences allows for relative abundances of bacteria to be identified within a sample.

HOW ARE SEQUENCING DATA ANALYZED?—COMPUTING PIPELINES

Output data from nearly all sequencing technology is in the form of ".fastq" files; these contain the specific sequence, primer sequences, barcode sequences, and quality information on the sequencing read, i.e., the per-base sequencing reliability score.

Analysis of these data is through computing pipelines using command-line inputs; typically, these are based on UNIX/LINUX operating system commands. Commands (specific actions for a computer to run) are entered into the command-line, which then runs scripts (programs) and the output are files or directories (equivalent to folders).

Several software packages exist for processing sequencing data. Many of these are freely available and have been produced by academic collaboration. In this study, we focus on the 2 most commonly used ones.

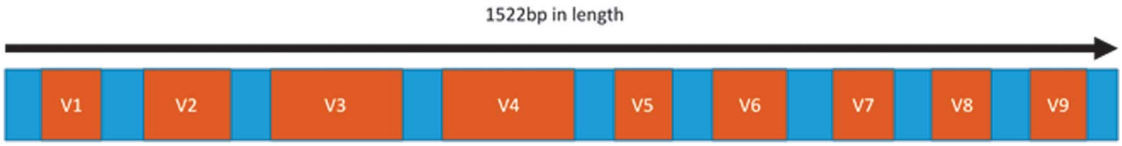

FIGURE 1. Schematic structure of the 16S rDNA displaying conserved (blue) and variable regions V1 to V9 (orange). Variable regions are represented to scale.

Quantitative Insights into Microbial Ecology (QIIME)

QIIME was developed by collaboration between the University of Colorado and Northern Arizona University. It is freely available as a download and can be installed on a home PC (through a VirtualBox) or Mac.[13] QIIME is actually a collection of many programs allowing for a huge amount of customization of output data; however, as each stage is analyzed separately, there is some added complexity to processing data.

QIIME uses command-line inputs. A basic pathway for analysis of data is shown in Figure 2. Output from initial analysis in the form of a simplified taxonomy summary table is provided in Table 1. Further analysis of this output data can be performed within QIIME, and this is discussed below.

Mothur[50]

Mothur is a widely used software package, developed at the University of Michigan that allows for analysis of 16S rRNA sequencing data using a very similar pipeline to QIIME. It has the advantage of being a single program to run but lacks some of the easy customization of QIIME. Output from QIIME and Mothur has been compared and is most often extremely similar.[51]

INTERPRETING RESULTS—WHAT DOES MY DATA MEAN?

The output data from basic 16S rRNA analysis requires interpretation. Although basic output such as a taxonomic summary shows the bacteria present (to all taxonomic levels) and their relative abundances within a sample, further analysis is needed to understand the quality of the data, the diversity within/between the samples, the species richness of the samples, and ultimately the statistical comparisons needed to determine whether a microbiome has experienced flux or dysbiosis. It must be remembered that many of the underlying tests included in these analyses have been directly drawn from ecological methods of species observations and are not specific for microbes.

Diversity Analysis—How Many Different Taxa are in My Sample?

Alpha diversity is the measure of diversity within a single sample (community).[49] Sample-specific values for alpha diversity can be compared. The most basic form of alpha diversity is simply the number of species seen in a sample; however, this is often unhelpful as there is no measure of frequency of species, estimates of unseen species, etc.

Chao1 is a commonly used measure of alpha diversity[52]; this is mathematically derived and estimates the number of OTUs that are undetected by sequencing based on the number of bacterial taxa represented by a single read, "singletons." Samples with higher numbers of singletons are estimated by Chao1 as having increased species richness, and therefore higher alpha diversity. Values range from 0 to greater than 10,000.

Shannon diversity is another measure of alpha diversity[53]; originally proposed as a measure of entropy (disorder) within text. This measure can be interpreted as follows; if all bacteria were placed in a list, the Shannon diversity measures the likelihood that one would guess the next bacteria (species) in the list. Hence, if there were 100 bacteria, 90 of which were *firmicutes*, the likelihood of predicting the next bacteria as

1. Removal of barcodes and linking each sequence to the sample using a mapping file (containing barcode and sample information)
2. Picking operational taxonomic units (OTUs) for each sequence. Similar sequences are clustered together based on the similarity of the DNA sequence
3. A single representative DNA sequence is chosen to represent each OUT
4. Output data (representative sequences) are compared to reference sequence data (such as greengenes database) and taxonomy is assigned to the representative sequences.
5. Data is collated into an 'OTU table' detailing the number of times an OTU is seen and which taxa this represents.
6. From the 'OTU table' different taxonomic summary tables are produced detailing the bacteria present and relative proportions at a phylum, class, order, family, genus and species level.

FIGURE 2. Basic data-processing pathway (pipeline) in QIIME.

TABLE 1. Example of Taxonomic Summary Table (Class/Level 2)

Class of Bacteria	Relative Abundance
k__Bacteria;p__Actinobacteria	0.0165982121715
k__Bacteria;p__Bacteroidetes	0.725658596336
k__Bacteria;p__Cyanobacteria	$3.20463176111 \times 10^{-5}$
k__Bacteria;p__Firmicutes	0.240162225581
k__Bacteria;p__Fusobacteria	$2.13642117407 \times 10^{-5}$
k__Bacteria;p__Lentisphaerae	$1.78035097839 \times 10^{-6}$
k__Bacteria;p__Proteobacteria	0.0169364788574
k__Bacteria;p__Tenericutes	0.000194058256645
k__Bacteria;p__Verrucomicrobia	$2.49249136975 \times 10^{-5}$

a *firmicutes* is very high; and hence the Shannon diversity would be low (close to 0). If there are 100 bacteria, 10 each of 10 different species, then the ability to predict the next bacteria is low and the Shannon diversity would be high. Values are typically 1.5 to 3.5 but may be higher in some microbiome analysis.

Rarefaction is another measure of alpha diversity and assesses species richness through construction of rarefaction curves. These are produced by plotting microbial diversity against depth of sequencing, an example is shown in Figure 3. These can be used to assess whether the depth of sequencing is sufficient to truly sample all species present within a community.[49]

Beta diversity is the measure of diversity between samples.

Bray–Curtis dissimilarity is possibly the most commonly used measure[54]; it describes the dissimilarity between the bacterial compositions of samples and is based on the species numbers at each site. It is expressed on a scale of 0 to 1, where 0 indicates identical samples and 1 represents no species shared between samples.

Unweighted UniFrac[55] is another measure of beta diversity, specific to microbiome research; it incorporates relative relatedness of bacteria with a sample (community) into the calculation and subsequent score. The score is derived by observing the taxa that are seen in both samples (and those that are individual to a sample) and then calculating total branch distances (an evolutionary measure of how closely related bacteria are) between shared and unshared bacteria on a phylogenetic tree. *Weighted UniFrac* also takes into account the relative abundances of bacteria within the samples.

What Further Statistical Analysis can be Done?

Principle coordinates analysis (PCoA) is increasingly common.[56] This presents each sample (bacterial community) as a data point on a 2- or 3-dimensional graph; similar samples cluster together. The axes of the graph represent the "principle coordinates" of the bacterial community. A simple example is provided in Figure 4.

Interesting Ways of Representing Data—Heat Maps, Networks

Heat maps have become a popular way of presenting 16S data; there are several ways that heat maps can represent various microbiome data. The most basic heat maps compare OTUs within samples, in which different colors represent different abundances in the samples. An example comparing bacterial genera is shown in Figure 5.

Correlation networks are a way of visualizing how bacterial species and other factors—fatty acids, host factors, diversity measures—interact. This is often used to compare diseased and healthy groups, demonstrating which bacteria/factors coexist and which do not.[57]

FUNCTIONAL IMPACTS OF THE MICROBIOME

The ability to infer functional impact of a bacterial community is a further stage of microbiome research. Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt)[29] is software that references known bacterial genomes to infer the genes present in bacteria that have been identified from 16S sequencing. This process allows for functional predictions about a bacterial community based on the metagenome, which has been inferred from 16S sequence.

An example of this can be found in articles by Gevers et al[4] and Morgan et al[58] on the microbiome in new onset pediatric Crohn's disease. Interestingly, both studies indicate a loss of normal bacterial biosynthetic function in diseased individuals.

CLINICAL VARIABLES IMPACTING THE MICROBIOME

Impact of Antibiotics

The number of variables that impact on, and alter, the human microbiome are too numerous to describe in this review. In a clinical context, antibiotics are perhaps the most important and common variable that can have a huge bearing on the microbiome of many sites; it is well described that antibiotic use in IBD amplifies the intestinal dysbiosis[4] and in non-IBD conditions, the gut flora is significantly altered by antibiotics, and this change can last for many months.[21,59–61]

Impact of Diet

Another important variable is the impact of diet, specifically on the gut microbiome. It is known that certain gut flora are associated with specific types of diet, and this can make interpretation of results more difficult.[62]

The Newborn

There has been an increase in interest centered around early bacterial colonization of the newborn infant; although vaginal birth is likely to result in a "normal" gut and skin flora in a baby in the first few days of life, the impact of a cesarean section on

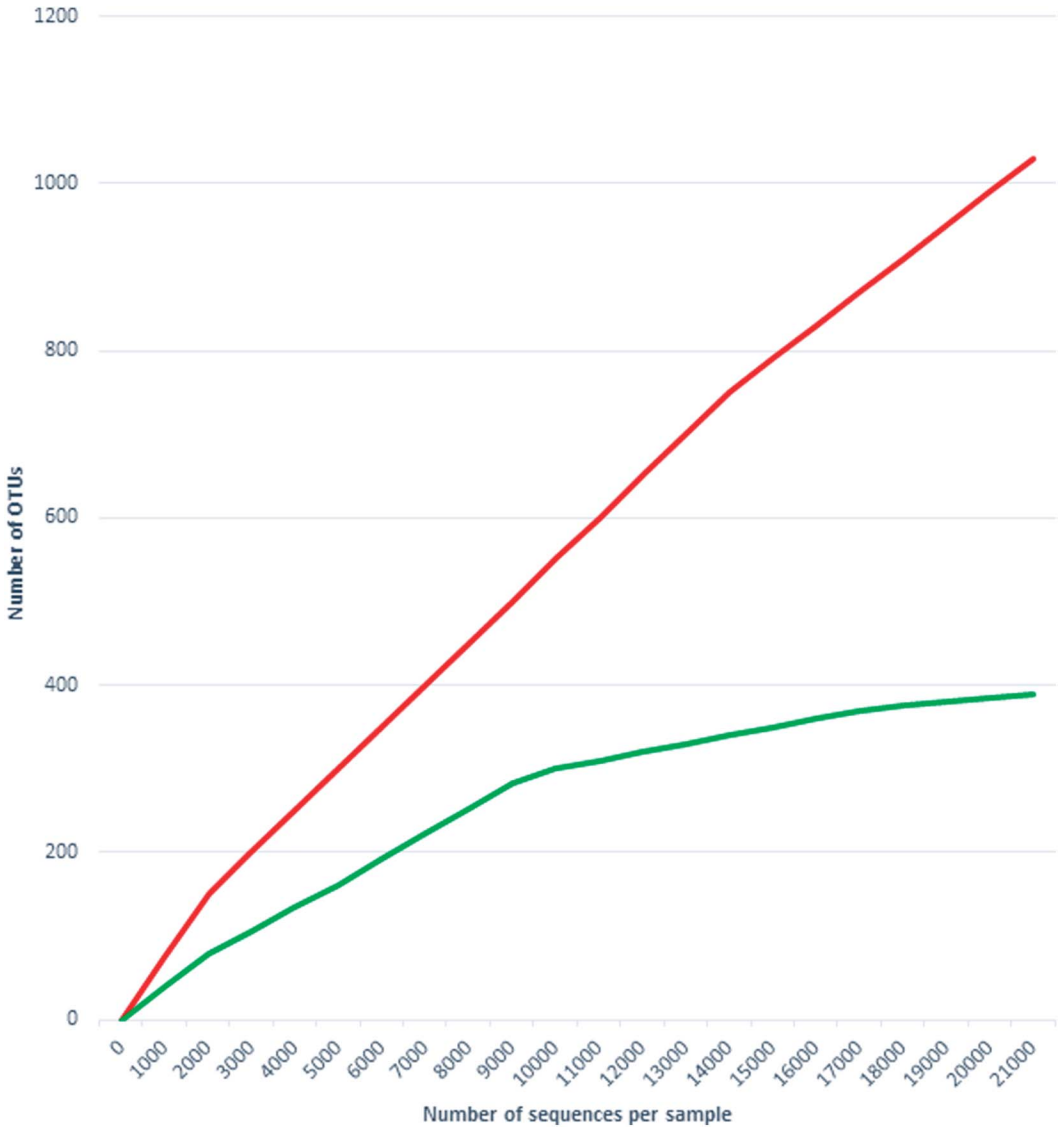

FIGURE 3. Rarefaction curves. Sample 1 (red line) may represent inadequate sequencing depth; the curve indicates that more OTUs would be found with the additional sequence number as the curve continues to rise beyond the maximum sequence number (not flattening out). Sample 2 (green) represents adequate sequencing depth; the curve flattens indicating that few additional OTUs will be identified even if the sample was sequenced in more depth.

developing "abnormal" flora and the relationship with disease requires further investigation.[63,64]

CURRENT APPLICATIONS AND CAUTION IN MICROBIOME INTERPRETATION

Although our ability to characterize microbiomes in both health and disease states has led to the swathe of publications, clinical interpretation of such findings has lagged behind. The ability to robustly interpret microbiome data has emerged over the last 5 to 10 years. The science is reliable, and there are an increasing number of large studies in the literature. However, despite many publications using current technology, there are still reports of potential pitfalls, such as over- or undersequencing of bacteria and misrepresentation of true numbers, in the interpretation of data.[65] Furthermore, the interpretation of results from a clinical perspective remains extremely uncertain with reliability and relevance of data needing to be examined before routine clinical application.

Without consensus on what is normal in health, it is difficult to understand the impact of the microbiome on individuals; inferring causality from correlation of microbiome results and disease is dangerous and further experimental work is needed.[66] It is known that there is huge interindividual microbiome variation at many body sites, especially the gut[67,68]; the work of the Human Microbiome Project to vastly increase the number of microbiome profiles of healthy individuals is extremely important in future

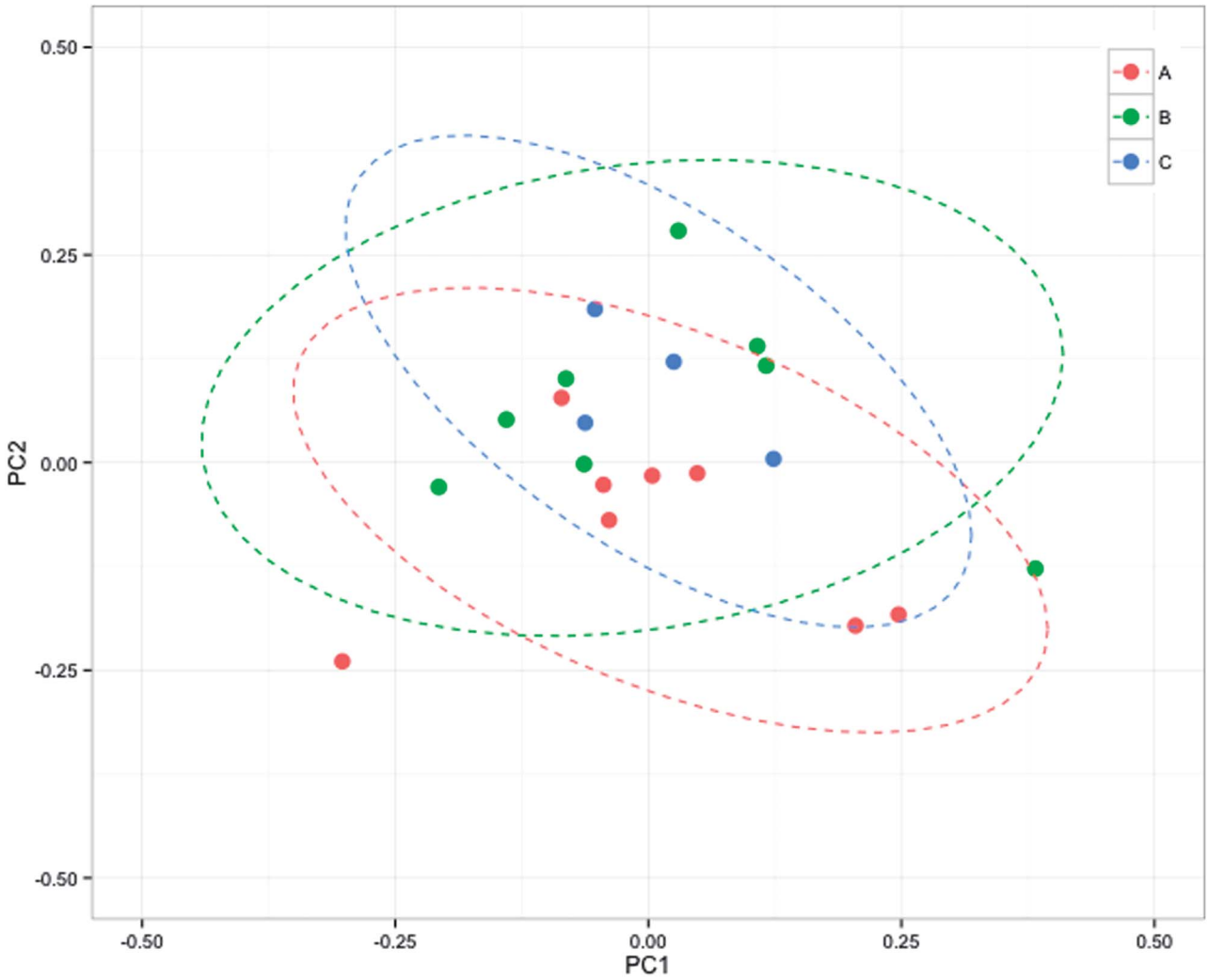

FIGURE 4. Example of a 2-dimensional (2D) principle coordinate analysis. Plot of first 2 principle coordinates (PC) for 20 samples generated using unweighted UniFrac analysis. PC1 (x-axis) accounts for 12.3% of total variation, and PC2 (y-axis) accounts for 10.8% of total variation. Clustering of samples indicates that they are more similar; the sample type can then be used to understand relationships between samples and overall shifts in the microbiome. Here, we show samples classified by 3 different treatment types (A, B, and C) and their subsequent clustering.

research and clinical interpretation.[8] Further insights into what contributes to this normal variation is an important area of work.[69,70]

FURTHER APPLICATION

We are at a point of transition—from observation and description to functional analysis and inferring causality. Understanding the host gene expression and functional impacts of microbiome in health and disease may help build evidence to infer the role that bacteria play in disease.[16] The understanding of the immunoregulatory role that microbes are presumed to play is extremely exciting, and we are at the start of a hugely important period in our understanding.

Currently, it seems unlikely that 16S sequencing will be used in the in diagnosis of acute infectious disease, in which accuracy, antibiotic sensitivity, and reliability are paramount, but there have been previous attempts.[71] In the future improved technology may allow the application of microbiome sequencing for routine sequencing, potentially identifying specific genes related to antibiotic resistance and allowing rapid diagnosis and treatment of infections to become routine practice.

Understanding broad microbial shifts and dysbiosis in immunodeficient individuals may help to prevent some infections associated with this group of conditions.[72]

CONCLUSION

The human microbiome seems to play a key role in maintaining health and also in the pathogenesis of many chronic diseases. Over the next 5 years, it will become increasingly important for clinicians to understand the impact of the microbiome on normal development, disease, and treatment. The meteoric rise of genomic medicine to the brink of routine clinical use can only be seen as

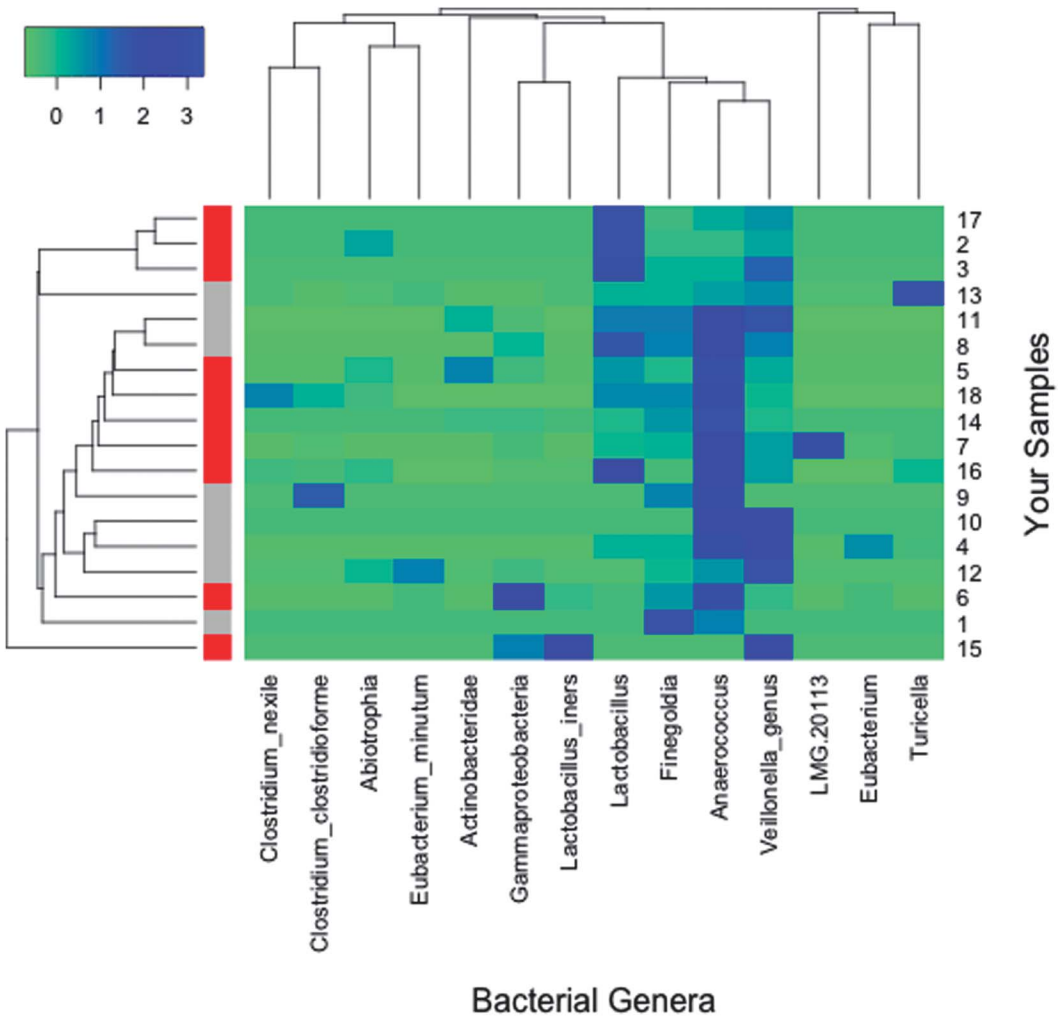

FIGURE 5. Generated example of a microbiome heat map. Simple heat map of 18 samples and 14 bacterial genera. The color (heat) is representative of the relative abundance of that bacterial genera within the sample. The trees represent how similar a sample is to another sample (y-axis) or how related a bacterial species is to another species (x-axis).

a blueprint for microbiome researchers; the ability for physicians to understand and interpret 16S data is vital to this growth and to the transition to a generation of microbiome-literate clinicians.

REFERENCES

1. Turnbaugh PJ, Ley RE, Hamady M, et al. The human microbiome project. *Nature*. 2007;449:804–810.
2. Wylie KM, Weinstock GM, Storch GA. Emerging view of the human virome. *Transl Res.* 2012;160:283–290.
3. Cui L, Morris A, Ghedin E. The human mycobiome in health and disease. *Genome Med.* 2013;5:63.
4. Gevers D, Kugathasan S, Denson LA, et al. The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe.* 2014;15:382–392.
5. Findley K, Grice EA. The skin microbiome: a focus on pathogens and their association with skin disease. *PLoS Pathog.* 2014;10:e1004436.
6. Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature*. 2012;486:222–227.
7. Grice EA, Segre JA. The skin microbiome. *Nat Rev Microbiol.* 2011;9:244–253.
8. Aagaard K, Petrosino J, Keitel W, et al. The human microbiome project strategy for comprehensive sampling of the human microbiome and why it matters. *FASEB J.* 2013;27:1012–1022.
9. Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev.* 2004;68:669–685.
10. Lane DJ, Pace B, Olsen GJ, et al. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A.* 1985;82:6955–6959.
11. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol.* 2006;72:5069–5072.
12. Tamboli CP, Neut C, Desreumaux P, et al. Dysbiosis in inflammatory bowel disease. *Gut*. 2004;53:1–4.
13. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7:335–336.
14. Caron DA, Countway PD, Savai P, et al. Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. *Appl Environ Microbiol.* 2009;75:5797–5808.
15. EMBL-EBI WT. Available at: https://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-you-will-learn/what-next-generation-dna-2015. Accessed May 1, 2016.

16. Haberman Y, Tickle TL, Dexheimer PJ, et al. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J Clin Invest.* 2015;125:1363.
17. Quince C, Ijaz UZ, Loman N, et al. Extensive modulation of the fecal metagenome in children with Crohn's disease during exclusive enteral nutrition. *Am J Gastroenterol.* 2015;110:1718–1729.
18. Wright EK, Kamm MA, Teo SM, et al. Recent advances in characterizing the gastrointestinal microbiome in Crohn's disease: a systematic review. *Inflamm Bowel Dis.* 2015;21:1219–1228.
19. Bonamichi-Santos R, Aun MV, Agondi RC, et al. Microbiome and asthma: what have experimental models already taught us? *J Immunol Res.* 2015;2015:614758.
20. Huang YJ, Nariya S, Harris JM, et al. The airway microbiome in patients with severe asthma: associations with disease features and severity. *J Allergy Clin Immunol.* 2015;136:874–884.
21. Shen N, Clemente JC. Engineering the microbiome: a novel approach to immunotherapy for allergic and immune diseases. *Curr Allergy Asthma Rep.* 2015;15:39.
22. Williams MR, Gallo RL. The role of the skin microbiome in atopic dermatitis. *Curr Allergy Asthma Rep.* 2015;15:65.
23. Yang L, Chaudhary N, Baghdadi J, et al. Microbiome in reflux disorders and esophageal adenocarcinoma. *Cancer J.* 2014;20:207–210.
24. Xu W, Yang L, Lee P, et al. Mini-review: perspective of the microbiome in the pathogenesis of urothelial carcinoma. *Am J Clin Exp Urol.* 2014;2: 57–61.
25. Pushalkar S, Ji X, Li Y, et al. Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. *BMC Microbiol.* 2012;12:144.
26. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–124.
27. Leach ST, Mitchell HM, Eng WR, et al. Sustained modulation of intestinal bacteria by exclusive enteral nutrition used to treat children with Crohn's disease. *Aliment Pharmacol Ther.* 2008;28:724–733.
28. Chumpitazi BP, Cope JL, Hollister EB, et al. Randomised clinical trial: gut microbiome biomarkers are associated with clinical response to a low FODMAP diet in children with the irritable bowel syndrome. *Aliment Pharmacol Ther.* 2015;42:418–427.
29. Langille MG, Zaneveld J, Caporaso JG, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol.* 2013;31:814–821.
30. Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. 2012;490:55–60.
31. Hartstra AV, Bouter KE, Bäckhed F, et al. Insights into the role of the microbiome in obesity and type 2 diabetes. *Diabetes Care*. 2015;38:159–165.
32. Stewart CJ, Marrs ECL, Nelson A, et al. Development of the preterm gut microbiome in twins at risk of necrotising enterocolitis and sepsis. *PLoS One*. 2013;8:e73465.
33. Morrow AL, Lagomarcino AJ, Schibler KR, et al. Early microbial and metabolomic signatures predict later onset of necrotizing enterocolitis in preterm infants. *Microbiome*. 2013;1:13.
34. Torrazza RM, Neu J. The altered gut microbiome and necrotizing enterocolitis. *Clin Perinatol.* 2013;40:93–108.
35. Zhou Y, Shan G, Sodergren E, et al. Longitudinal analysis of the premature infant intestinal microbiome prior to necrotizing enterocolitis: a case-control study. *PLoS One*. 2015;10:e0118632.
36. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol.* 1975;94: 441–448.
37. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860–921.
38. Land M, Hauser L, Jun SR, et al. Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics.* 2015;15:141–161.
39. Loman NJ, Pallen MJ. Twenty years of bacterial genome sequencing. *Nat Rev Microbiol.* 2015;13:787–794.
40. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456:53–59.
41. Nossa CW, Oberdorf WE, Yang L, et al. Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. *World J Gastroenterol.* 2010;16:4135–4144.
42. Benítez-Páez A, Portune KJ, Sanz Y. Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION portable nanopore sequencer. *Gigascience*. 2016;5:4.
43. Case RJ, Boucher Y, Dahllöf I, et al. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol.* 2007;73:278–288.
44. Weisburg WG, Barns SM, Pelletier DA, et al. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol.* 1991;173:697–703.
45. Yarza P, Yilmaz P, Pruesse E, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol.* 2014;12:635–645.
46. Claesson MJ, Wang Q, O'Sullivan O, et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* 2010;38:e200.
47. Hayashi H, Sakamoto M, Benno Y. Evaluation of three different forward primers by terminal restriction fragment length polymorphism analysis for determination of fecal bifidobacterium spp. in healthy subjects. *Microbiol Immunol.* 2004;48:1–6.
48. Chakravorty S, Helb D, Burday M, et al. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods.* 2007;69:330–339.
49. Kuczynski J, Lauber CL, Walters WA, et al. Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet.* 2012;13:47–58.
50. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 2009;75:7537–7541.
51. Majaneva M, Hyytiäinen K, Varvio SL, et al. Bioinformatic amplicon read processing strategies strongly affect eukaryotic diversity and the taxonomic composition of communities. *PLoS One*. 2015;10:e0130035.
52. Chao A. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*. 1987;43:783–791.
53. Li K, Bihan M, Yooseph S, et al. Analyses of the microbial diversity across the human microbiome. *PLoS One*. 2012;7:e32118.
54. Bray JR, C JT. An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr.* 1957;27:325–349.
55. Lozupone CA, Hamady M, Kelley ST, et al. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol.* 2007;73:1576–1585.
56. Hughes JB, Hellmann JJ, Ricketts TH, et al. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol.* 2001;67:4399–4406.
57. Kelder T, Stroeve JH, Bijlsma S, et al. Correlation network analysis reveals relationships between diet-induced changes in human gut microbiota and metabolic health. *Nutr Diabetes.* 2014;4:e122.
58. Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 2012;13:R79.
59. Zaura E, Brandt BW, Teixeira de Mattos MJ, et al. Same exposure but two radically different responses to antibiotics: resilience of the salivary microbiome versus long-term microbial shifts in feces. *MBio*. 2015;6: e01693–15.
60. Jernberg C, Löfmark S, Edlund C, et al. Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology*. 2010;156(pt 11):3216–3223.
61. Jernberg C, Löfmark S, Edlund C, et al. Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *ISME J.* 2007;1:56–66.
62. De Filippo C, Cavalieri D, Di Paola M, et al. 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.* 2010;107:14691–14696.
63. Koenig JE, Spor A, Scalfone N, et al. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A.* 2011; 108(suppl 1):4578–4585.
64. Neu J, Rushing J. Cesarean versus vaginal delivery: long-term infant outcomes and the hygiene hypothesis. *Clin Perinatol.* 2011;38:321–331.
65. Greenblum S, Carr R, Borenstein E. Extensive strain-level copy-number variation across human gut microbiome species. *Cell*. 2015;160:583–594.
66. Harley IT, Karp CL. Obesity and the gut microbiome: striving for causality. *Mol Metab.* 2012;1:21–31.

67. Schloissnig S, Arumugam M, Sunagawa S, et al. Genomic variation landscape of the human gut microbiome. *Nature*. 2013;493:45–50.
68. Hildebrand F, Nguyen TL, Brinkman B, et al. Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol.* 2013;14:R4.
69. Dominianni C, Sinha R, Goedert JJ, et al. Sex, body mass index, and dietary fiber intake influence the human gut microbiome. *PLoS One*. 2015; 10:e0124599.
70. David LA, Maurice CF, Carmody RN, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. 2014;505: 559–563.
71. McKenney PT, Pamer EG. From hype to hope: the gut microbiota in enteric infectious disease. *Cell*. 2015;163:1326–1332.
72. Zevin AS, McKinnon L, Burgener A, et al. Microbial translocation and microbiome dysbiosis in HIV-associated immune activation. *Curr Opin HIV AIDS*. 2016;11:182–190.