REVIEW ARTICLES

The human gut microbiome: current knowledge, challenges, and future directions

MANEESH DAVE, PETER D. HIGGINS, SUMIT MIDDHA, and KEVIN P. RIOUX

ROCHESTER, MINN; ANN ARBOR, MICH; AND CALGARY, ALBERTA, CANADA

The Human Genome Project was completed a decade ago, leaving a legacy of process, tools, and infrastructure now being turned to the study of the microbes that reside in and on the human body as determinants of health and disease, and has been branded "The Human Microbiome Project." Of the various niches under investigation, the human gut houses the most complex and abundant microbial community and is an arena for important host–microbial interactions that have both local and systemic impact. Initial studies of the human microbiome have been largely descriptive, a testing ground for innovative molecular techniques and new hypotheses. Methods for studying the microbiome have quickly evolved from low-resolution surveys of microbial community structure to high-definition description of composition, function, and ecology. Next-generation sequencing technologies combined with advanced bioinformatics place us at the doorstep of revolutionary insight into the composition, capability, and activity of the human intestinal microbiome. Renewed efforts to cultivate previously "uncultivable" microbes will be important to the overall understanding of gut ecology. There remain numerous methodological challenges to the effective study and understanding of the gut microbiome, largely relating to study design, sample collection, and the number of predictor variables. Strategic collaboration of clinicians, microbiologists, molecular biologists, computational scientists, and bioinformaticians is the ideal paradigm for success in this field. Meaningful interpretation of the gut microbiome requires that host genetic and environmental influences be controlled or accounted for. Understanding the gut microbiome in healthy humans is a foundation for discovering its influence in various important gastrointestinal and nutritional diseases (eg, inflammatory bowel disease, diabetes, and obesity), and for rational translation to human health gains. (Translational Research 2012;160:246–257)

Abbreviations: GI = gastrointestinal; IBD = inflammatory bowel disease; IBS = irritable bowel syndrome; TRFLP = terminal restriction fragment length polymorphism; UC = ulcerative colitis

From the Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, Minn; Division of Gastroenterology and Hepatology, Department of Internal Medicine, University of Michigan Medical Center, Ann Arbor, Mich; Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minn; Department of Medicine, Division of Gastroenterology and Hepatology, Department of Microbiology and Infectious Diseases, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada.

The authors have no financial disclosures relevant to this article.

Submitted for publication December 6, 2011; revision submitted May 8, 2012; accepted for publication May 8, 2012.

Reprint requests: Kevin P. Rioux, University of Calgary, 1863 Health Sciences Centre, 3330 Hospital Drive NW, Calgary, Alberta, Canada T2N 4N1; e-mail: kprioux@ucalgary.ca.

1931-5244/$ - see front matter

doi:10.1016/j.trsl.2012.05.003

Humans live in a biosphere where microbes are ubiquitous and have existed and evolved over 3.8 billion years.[1] The intestinal tract of humans harbors a complex microbial community estimated to contain approximately 100 trillion cells, exceeding the number of human cells by a factor of 10.[2-4] This complex community of microbes in the alimentary tract (bacteria, archaea, eukarya, and viruses) is also known as the "gastrointestinal (GI) microbiota."[5,6] Akin to the human genome, which is the sum of all human genes, the term "microbiome" refers to all of the microbiota in a defined microbial community, which are usually differentiated by their genetic elements.[7] Metagenomics refers to the functional and compositional analysis of an assemblage of microbes based on molecular study of their collective genomes.[8,9]

The gut microbiota performs a variety of beneficial functions (Table I), and given their importance in human health and disease, the Human Microbiome Project was launched by the National Institutes of Health to (1) characterize the microbial communities of various niches of the human body (eg, the nasal passages, oral cavity, skin, urogenital system, and GI tract), (2) to determine whether individuals share a core human microbiome, and (3) to explore whether changes in the human microbiome cause or correlate with human disease.[10,11] The distal GI tract contains the most abundant and diverse communities of microbes, in continuous interplay with the human host resulting in both local (mucosal and luminal) and systemic (metabolic and nutritional) effects. Therefore, of all microniches being explored by human microbiome researchers, the GI tract holds the most promise for discovery of important new concepts and understanding of the human "superorganism" and translation to clinical biomarkers and therapies.[12]

There are excellent reviews that have focused on the temporal and spatial development of the human gut microbiota[13] and their role in health and disease.[14] The goal of this article is to highlight the current state of knowledge of the human gut microbiome, discuss technical and practical limitations of scientific devices and tools, and map out some of the knowledge gaps in this challenging field that will be solved by thoughtful scientific approaches, advanced sequencing and computational technology, and persistence.

TECHNIQUES FOR ANALYZING MICROBIOTA

Studies in the 1970s using anaerobic culture–based techniques identified more than 400 to 500 distinct bacterial species in the human gut.[15] Studying the human GI microbiota by cultivation methods has many drawbacks: It produces selective growth of some organisms and thus distorts composition of the natural community.

Table I. Biological effects of the gut microbiota on human host

Development of innate and adaptive immunity
Intestinal epithelial integrity
Energy source
Vitamin biosynthesis, bile salt transformation, catabolism of dietary glycans (eg, cellulose and pectins)
Barrier to colonization by microbial pathogens
Xenobiotic metabolism

Moreover, approximately 60% to 80% of gut microbes simply cannot be grown by conventional in vitro techniques.[16,17] In 1977, Woese and Fox[18] described a technique for molecular characterization of bacterial phylogeny based on ribosomal RNA sequence analysis. In particular, the 16S rRNA is a molecule that is universally present in bacteria and has highly conserved domains flanking hypervariable sequences that can be used to distinguish bacterial groups. In the early 1980s, various culture-independent molecular techniques based on the 16S rRNA became available, and during the last 3 decades they have been used extensively to assess the structure of complex or fastidious prokaryotic communities.[1] Examples of these techniques are terminal restriction fragment length polymorphism (TRFLP), denaturing gradient gel electrophoresis, and fluorescent in situ hybridization.[19] These bacterial community profiling or "fingerprinting" techniques first involve isolating bacterial DNA from environmental or biological samples. The DNA is amplified by polymerase chain reaction using universal primers that target conserved regions of the 16S rRNA gene, and the resulting amplicons contain variable regions that discern the constituent members of bacterial communities by electrophoretic or hybridization techniques.

Cloning and then sequencing of the 16S rRNA gene in an automated capillary sequencer is a higher-resolution method of studying bacterial phylogeny.[20] This technique uses Sanger sequencing to produce a long read (~800 base pairs) of the 16S rRNA gene, which enables identification of bacteria at a higher-level phylogenetic resolution (ie, genera and species). This relies on robust bioinformatics tools, such as the Ribosomal Database Project.[21]

In the wake of the Human Genome Project, next-generation sequencing technologies have emerged that have increased the depth and speed of phylogenetic coverage and decreased the cost through massively parallel sequencing methods. There are various commercial platforms for next-generation sequencing. One of these newer techniques is pyrosequencing that identifies nucleotides by the amplitude of light signals generated when luciferin is converted to oxyluciferin during