Cell

Leading Edge
Primer

Conducting a Microbiome Study

Julia K. Goodrich,[1,2] Sara C. Di Rienzi,[1,2] Angela C. Poole,[1,2] Omry Koren,[1,2,9] William A. Walters,[3] J. Gregory Caporaso,[4,5] Rob Knight,[6,7,8] and Ruth E. Ley[1,2,*]

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA
[2]Department of Microbiology, Cornell University, Ithaca, NY 14853, USA
[3]Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309, USA
[4]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA
[5]Institute for Genomics and Systems Biology, Argonne National Laboratory, Argonne, IL 60439, USA
[6]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA
[7]BioFrontiers Institute, University of Colorado, Boulder, CO 80309, USA
[8]Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA
[9]Present address: Faculty of Medicine, Bar Ilan University, Ramat Gan 52900, Israel
*Correspondence: rel222@cornell.edu
http://dx.doi.org/10.1016/j.cell.2014.06.037

Human microbiome research is an actively developing area of inquiry, with ramifications for our lifestyles, our interactions with microbes, and how we treat disease. Advances depend on carefully executed, controlled, and reproducible studies. Here, we provide a Primer for researchers from diverse disciplines interested in conducting microbiome research. We discuss factors to be considered in the design, execution, and data analysis of microbiome studies. These recommendations should help researchers to enter and contribute to this rapidly developing field.

Introduction

Many studies have documented differences in the composition of host-associated microbial communities between healthy and disease states (Clemente et al., 2012; Karlsson et al., 2013; Knights et al., 2013). For a growing number of diseases, an altered microbiome is not just a marker of disease, but also actively contributes to pathology (Chassaing et al., 2012). The best empirical direct evidence that microbiomes can drive disease comes from experiments in which the microbiota from diseased donors and controls are "transplanted" into healthy germ-free hosts: if recipients of the disease-associated microbiome display the disease phenotype, the microbiome is considered causal. This approach, pioneered by Jeffrey Gordon and his group (Turnbaugh et al., 2006), has directly demonstrated that the composition of gut microbial communities can alter host metabolism (Koren et al., 2012; Vijay-Kumar et al., 2010), transmit colitis (Garrett et al., 2007), and modulate type I diabetes (Wen et al., 2008). The range of conditions with a host-microbiome interaction component continues to grow and has recently started to include neurological conditions (Collins et al., 2012). Consequently, researchers from a wide array of disciplines are interested in testing whether microbes, and especially gut microbes, are associated with various pathologies, whether they actively participate in disease, and ultimately whether they can present novel targets for therapies. This Primer is intended for non-experts who are considering their first microbiome project and summarizes lessons learned from past successful and unsuccessful projects.

Mammalian microbiome research has a long history (Savage, 1977), recently marked by dramatic increases in scale and scope due to advances in DNA-sequencing technologies and in associated computational methods. Anecdotal descriptions of community composition that set the standard in the recent past have given way to study designs that allow for repeated measurements, error estimates, correlations of microbiota with covariates, and increasingly sophisticated statistical tests (Knight et al., 2012). Today, microbiome data are obtained predominantly in three forms: (1) 16S rRNA gene sequence surveys that provide a view of microbiome membership, (2) metagenomic data used to portray functional potential, and (3) metatranscriptomic data to describe active gene expression. Here, we focus primarily on 16S rRNA gene surveys because they are economical and therefore scale to larger projects. 16S rRNA gene sequence data provide a relatively unbiased characterization of bacterial and archaeal diversity (Box 1 provides a brief overview of methods for characterizing the diversity of microbial eukaryotes and viruses). Regardless of the types of microorganisms targeted or the methodology used to characterize them, choices made at every step, from study design to analysis, can impact results. This Primer highlights resources that address specific technical questions and provides general advice stemming from our collective experience working in the field. Although we focus mainly on the mammalian gut microbiota, many of the same issues apply to microbial communities of other habitats. We have structured the Primer to answer questions that are commonly raised by researchers entering the field (Figure 1).

Animal Studies

The Maternal Effect

A large fraction of microbiome studies are conducted in animals, particularly rodents, as they offer attractive models for human biology and their environmental conditions can be tightly controlled. How animals are bred and raised is the most important source of confounding factors in microbiome studies conducted in animals. Inoculation of mice at birth (the maternal effect) is a major factor shaping the composition of the

Box 1. Archaeal, Viral, and Eukaryotic Diversity

Most studies of the human microbiota describe bacterial diversity, which typically dominates the cellular fraction of the microbiota; but other taxa, including Archaea, fungi, and other microbial eukaryotes, and viruses can be present.

Archaea. Archaeal diversity can be characterized using the commonly employed 515F/806R primer set (and others), and their diversity can be analyzed in the same way as bacterial diversity. The 16S rRNA gene is the most widely used marker gene for the Archaea, and their diversity is represented in reference data sets commonly used for Bacteria.

Microbial Eukaryotes. Characterization of fungal communities, in particular, is an active research area. In principle, the bioinformatics pipeline is the same for eukaryotic marker genes as for bacterial marker genes (Iliev et al., 2012). However, the lack of a standard marker gene and reference database means that the bioinformatics protocols are not as standardized as for 16S rRNA gene analysis. For fungi, although several marker gene options exist, the internal transcribed spacer (ITS) region of the 16S rRNA gene is generally preferred for obtaining high taxonomic resolution. The UNITE database (Abarenkov et al., 2010) is often used for ITS sequence-based analyses of fungal sequences. However, the ITS region is not amenable to alignments across distinct fungal taxa, so ITS-based fungal community studies frequently do not make use of phylogenetic metrics for alpha- and beta-diversity comparisons. One strategy that is being explored is using the 18S rRNA gene and ITS in conjunction to define fungal phylogenetic trees. Moreover, the 18S rRNA gene can, in principle, be used to analyze eukaryotic communities in the same manner that 16S rRNA genes are used. A reference database containing many eukaryotic sequences, such as SILVA (Quast et al., 2013), should be used for such analyses. One should confirm that the region of the 18S gene amplified discriminates between the taxa studied and should be aware that the 18S rRNA gene is not sufficient to characterize the eukaryotic phylogeny: trees built from 18S sequence alone will likely be of questionable utility.

Viruses. Characterizing the human virome requires a different approach because, unlike for cellular life, no gene or genomic region is homologous across all viruses. The current approach for studying these communities is to isolate virus-like particles (VLPs) using size fractionation and to sequence those using metagenomics (Caporaso et al., 2011a; Handley et al., 2012; Hurwitz et al., 2013; Reyes et al., 2010). Alternatively, viruses can be characterized using DNA microarrays (Jack et al., 2009; Palacios et al., 2007).

microbiota and leads to a sharing of suites of bacteria between littermates and their mothers that differentiates them from members of other families and can persist over several generations (Ley et al., 2005). The maternal effect determines the specific suite of microbes available to colonize a host. Subsequently, the individual host and host diet shapes the relative abundances of these taxa (Ley et al., 2005; Rawls et al., 2006).

Mitigating the Maternal Effect

The maternal effect is particularly problematic when it confounds the experimental effect (see Figure 2 for an example). Because littermates and even co-caged unrelated animals can share microbiotas due to coprophagy and other modes of transmission, randomization of treatments across litters/cages becomes an important aspect of experimental design. When the goal is to compare the effect of different genotypes on the microbiome, the options range from the use of germ-free mice gavaged with the same inoculum to the use of mixed-genotype litters.

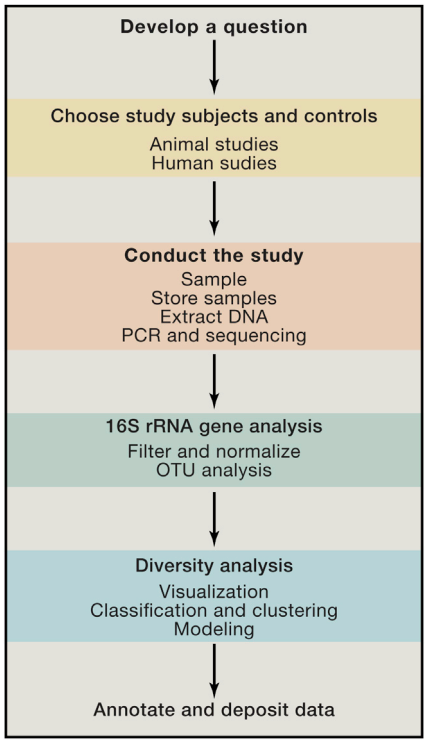

Figure 1. Conducting a Microbiome Study
The sequential steps of conducting a microbiome study are diagramed, mirroring the sections of this Primer.

When these options are not available, alternate approaches include embryo transfers so that mixed genotypes are born together, cross-fostering, and cohousing post weaning. The last two options may be the least effective, as microbiotas will be at least partly assembled. In large studies with multiple litter/cage replicates, the variance in the data that is attributable to the maternal effect can be accounted for in statistical models (Benson et al., 2010).

Using the Maternal Effect to Maximize a Phenotype

In some cases, animals of different genotypes are maintained separately in order to maximize the maternal effect and obtain a strong microbial phenotype (Vaishnava et al., 2011; Vijay-Kumar et al., 2010). Separately maintained mice can then be cohoused to demonstrate the spread of a microbiota between adults (Lawley et al., 2012; Ridaura et al., 2013). Conversely, mice can be housed individually to minimize cross-contamination and maintain individual microbiotas (Ley et al., 2005).

Environment Matters: Microbiotas Vary Greatly among Facilities

Mouse microbiotas can differ significantly between facilities even if they have identical genotypes (Friswell et al., 2010). Environmental conditions can differ between facilities—for instance, the water acidity, food, bedding, and so on can differ. But it appears that different colonies harbor their own populations of microbes as well. One striking example of this facility effect is that of the segmented filamentous bacteria (SFB), which have been reported more common in mice obtained from one common vendor (JAX) than another (Taconic) (Ivanov et al., 2009).