Annals of Epidemiology 26 (2016) 373–379

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

The Microbiome and Epidemiology

Characterization of the gut microbiome in epidemiologic studies: the multiethnic cohort experience

Benjamin C. Fu MPH [a,b], Timothy W. Randolph PhD [a], Unhee Lim PhD [c], Kristine R. Monroe PhD [d], Iona Cheng PhD, MPH [e,f], Lynne R. Wilkens DrPH [c], Loïc Le Marchand MD, PhD [c], Meredith A.J. Hullar PhD [a,1], Johanna W. Lampe PhD, RD [a,b,*,1]

[a] Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA
[b] Department of Epidemiology, University of Washington, Seattle
[c] Epidemiology Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu
[d] Department of Preventive Medicine, University of Southern California, Los Angeles
[e] Cancer Prevention Institute of California, Fremont
[f] Stanford Cancer Institute, Stanford, CA

ARTICLE INFO

Article history:
Received 24 November 2015
Accepted 26 February 2016
Available online 8 March 2016

Keywords:
Microbiome
Gut bacteria
Stool
Multiethnic Cohort

ABSTRACT

Purpose: The development of next-generation sequencing and accompanying bioinformatics tools has revolutionized characterization of microbial communities. As interest grows in the role of the human microbiome in health and disease, so does the need for well-powered, robustly designed epidemiologic studies. Here, we discuss sources of bias that can arise in gut microbiome research.
Methods: Research comparing methods of specimen collection, preservation, processing, and analysis of gut microbiome samples is reviewed. Although selected studies are primarily based on the gut, many of the same principles are applicable to samples derived from other anatomical sites. Methods for participant recruitment and sampling of the gut microbiome implemented in an ongoing population-based study, the Multiethnic Cohort (MEC), are also described.
Results: Variation in methodologies can influence the results of human microbiome studies. To help minimize bias, techniques such as sample homogenization, addition of internal standards, and quality filtering should be adopted in protocols. Within the MEC, participant response rates to stool sample collection were comparable to other studies, and in-home stool sample collection yields sufficient high-quality DNA for gut microbiome analysis.
Conclusions: Application of standardized and quality controlled methods in human microbiome studies is necessary to ensure data quality and comparability among studies.

Introduction

Spurred by the development and implementation of novel sequencing technologies and bioinformatic techniques, the last decade has seen great strides in human microbiome research. Furthermore, substantial reductions in sequencing costs have facilitated incorporation of microbiome research into large-scale epidemiologic studies. Major efforts to characterize microbial communities [1,2] have helped to elucidate the rich and diverse microbial landscape in and on the human body, as well as the substantial variation across individuals.

In addition to identifying which microbes are present, the functional capacity of the microbiome can be characterized with metagenomic sequencing which provides a snapshot of the genetic composition of microbial genomes. Bacteria respond rapidly to changes in their environment not only in abundance but also in composition of the metabolically active fraction of the microbial community. Metatranscriptomics (i.e., microbial rRNA or mRNA sequencing) can be used to measure ribosomal and actively transcribed mRNA to gain insight into gene expression patterns. Although still fairly new, the utility of these methods continues to grow as more genes become annotated. Metaproteomic (proteins from the microbial community) and metabolomic approaches are also increasingly used as a way to examine the products of

* Corresponding author. Public Health Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M4-B402, Seattle, WA 98109. Tel.: +1-206-667-6580; fax:+1-206-667-7850.
E-mail address: jlampe@fredhutch.org (J.W. Lampe).
[1] These authors contributed equally, shared senior authorship.

http://dx.doi.org/10.1016/j.annepidem.2016.02.009
1047-2797/

microbial metabolism [3]. The bacterial component chosen to represent the microbiome (DNA, RNA, protein, or metabolite) needs to be considered early as it will influence the experimental design and timing of metadata gathered from the human population.

With the increasing wealth of microbiome data being generated from these techniques, there is a need to better understand clinically meaningful differences, confounding factors, and causality when studying the microbiome in relation to disease risk. Epidemiology is well equipped to tackle these issues by building on the knowledge and analytical tools that have been developed in assessing microbial exposures and multifactorial diseases. Here, we provide a review of potential sources of variation and bias that can be introduced and should be considered when studying the gut microbiome in large, prospective, population-based studies, as well as describe some of our experience collecting fecal samples as part of a Multiethnic Cohort (MEC) study in Hawaii and California. Although our focus is on the gut microbial community (GMC) and its function, aspects of sample collection, preservation, processing, and analysis are also relevant to other anatomical sites.

Where to sample the gut microbiome

Studies of the gut microbiome often use stool samples, for which collection is noninvasive and can be carried out privately by study participants. Biopsies are a second option, especially in studies of colorectal cancer risk or other situations when a colonoscopy or sigmoidoscopy is indicated. However, the gut mucosal and luminal microbiomes are not necessarily comparable [4,5]; several studies have shown differences between stool and biopsy samples [6–9], and microbial populations have been reported to differ also in biopsies collected at various locations along the gastrointestinal tract [6,10,11]. Swabs have been used as well, although both rectal swabs and swabs of fresh stool may differ from the previous two methods in terms of microbial composition and DNA yield [12–14].

Two other less common collection methods for interrogating the fecal and mucosal microbiomes, respectively, are fecal occult blood tests (FOBT) and formalin fixed paraffin-embedded tissue from surgical resection or biopsy. Stool collected with FOBT cards was reported to be similar to that collected directly into a storage tube, either with or without an RNA-stabilizing agent, in terms of microbial community structure and taxa distribution [15]. Another recent study found FOBT to have optimal stability and reproducibility compared with seven fecal sampling methods [16]. Formalin fixed paraffin-embedded tissue is often collected in clinical settings and has been used for bacterial identification in a variety of diseases [17–19], but faces sample quality issues related to low DNA yields, fragmentation, and sequence artifacts [20,21].

Sampling method is an important consideration in understanding disease etiology, as disease states may have varying effects on microbial communities depending on anatomical site. For example, Bajaj et al. [8] reported no differences in the stool microbiota of patients with and without hepatic encephalopathy, whereas significant differences in microbes associated with colonic mucosal biopsies were present between the two groups. Ultimately, the choice of sampling approach should be driven by the hypothesis being tested. For example, study of direct effects of microbes on the gut mucosa likely warrants the use of colon tissue for characterization of the gut microbiome, whereas studies of microbial metabolism of dietary constituents may be better served by analysis of luminal contents or stool.

Sampling the stool microbiome

Clinical analysis of stool is commonplace [22,23], and collection methods can be readily applied to research settings. Various groups, including the Human Microbiome Project [24], have developed protocols that allow participants to collect a stool sample in the privacy of their own homes. An issue with using stool is that the considerable variation in the gut environment may lead to an uneven distribution of microbes in the stool sample. The spatial distribution of gut microbes, both longitudinally and radially, is influenced by factors such as increasing pH levels from the proximal to distal colon [25] and higher concentrations of oxygen near the mucosa relative to the lumen [26]. Homogenization of whole stool is one approach for obtaining a more uniform sample; this has been shown to reduce the variation in both the amount of DNA extracted [27] and the relative abundance of bacterial taxa [28,29].

Stool collection and transport

Several factors related to specimen handling may also influence the quantity and quality of nucleic acids present in the stool samples and therefore impact the microbiome data generated.

Temperature

Studies have investigated the effect of temperature either during sample transport or storage on microbial community structure, finding variation due to temperature to be less than that due to interindividual differences [13,30,31]. However, storing samples at room temperature for more than 24 hours without preservative can have significant effects on bacterial community composition and RNA fragmentation [32,33]. In terms of specific microbes, Tedjo et al. [13] found no taxa to be associated with storage method, while Rubin et al. [34] reported that only one of 2781 taxa significantly differed across three temperatures (−20°C, 4°C, room temperature). Gorzelak et al. [29] observed changes in Firmicutes levels within 3 days and Bacteroidetes by 14 days for storage at −20°C, whereas Fouhy et al. [35] showed significant differences in two genera (*Faecalibacterium*, *Leuconostoc*) when comparing fresh and flash-frozen samples.

Storage solution

Submersion of samples in a nucleic acid storage solution also aids in preservation and greatly adds to the ease of sampling in the home environment. The product commonly used for human microbiome studies is RNAlater (Ambion). Several studies have shown RNAlater to be an effective storage reagent for preserving RNA or DNA [36–38]. In cluster analyses, comparing frozen samples to those in RNAlater only, samples nearly always grouped by individual when either DNA- or RNA-based methods were used [15,33,38,39]. However, some studies have found RNAlater to reduce yield and purity of bacterial DNA [15,29,40], and one found alterations in several bacterial phyla over 72 hours of storage compared with frozen samples [41].

Laboratory considerations

Storage time

For large cohort studies, the capacity to store specimens without extensive processing is an important cost consideration. Studies have found interindividual variation in microbial composition to be greater than variation due to storage time regardless of storage length. Clustering of repeated samples within individual has been seen over periods of up to 14 days [42], 6 months [43], and over 2 years [38]. Although few studies have assessed changes in individual taxa over storage time, Lauber et al. [42] have reported that the relative abundance of some microbes significantly changed over