Seminars in Fetal & Neonatal Medicine 21 (2016) 368–372

Contents lists available at ScienceDirect

Seminars in Fetal & Neonatal Medicine

journal homepage: www.elsevier.com/locate/siny

Review

The National Institutes of Health Human Microbiome Project

Lita M. Proctor*

Coordinator, Human Microbiome Project, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

Keywords:
Microbiome
DNA sequence
Metagenomics
Metabolomics
Metatranscriptomics
Resources

SUMMARY

This overview describes the impetus for and the goals of the National Institutes of Health (NIH)'s Human Microbiome Project (HMP) and the research resources available through the HMP. As the HMP also serves as a catalyst for human microbiome research at the NIH, NIH Institutes and Centers support for this field is also briefly addressed.

Published by Elsevier Ltd.

1. Introduction

The conceptual and technological foundations for the study of the human microbiome were forming over three decades ago in the ecology field. Microbial ecologists recognized that most micro-organisms in nature were not amenable to culture and so developed molecular approaches to the study of microbial communities. Further, these scientists also recognized that these micro-organisms interacted in communities so that culture-based approaches, even if available, would not capture these interactions. An early and broadly adopted method, based on the three-domain system for biological classification [1], was the use of the 16S ribosomal RNA gene as a phylogenetic (i.e. taxonomic) marker for interrogating microbial diversity in nature [2]. With the growth of nonculture-based, genome-enabled technologies to study environmental micro-organisms and communities, some medical microbiologists began to apply the 16S approach on human specimens and found far greater microbial diversity associated with the human body than expected, even in well-studied regions such as the oral cavity [3–5]. A National Academy of Sciences report coined the term “metagenomics” to describe both the technical approaches and this young field, the study of microbial communities and their interactions through molecular analysis [6].

In the infectious diseases field, recognition was growing that many diseases could not satisfy Koch's postulates as the pathogenesis of many of these diseases appeared to involve multiple micro-organisms. The term ‘polymicrobial diseases’ was coined to describe those diseases with multiple infectious agents [7]. We now recognize that many of these formerly classified polymicrobial diseases, such as aacuired immune deficiency syndrome (AIDS)-related opportunistic infections, periodontal diseases, and respiratory diseases, are associated with multiple microbial members, i.e. the microbiome. In an essay on the history of microbiology and infectious disease, Lederberg [8], who coined the term “microbiome,” called for “a more ecologically-informed metaphor” to understand the relationship between humans and their endogenous resident microbial communities.

The field of immunology was also undergoing a revolution with the recognition that the innate and adaptive immune systems not only evolved to eliminate specific pathogens but were also intimately involved in shaping the composition of the commensal intestinal microbiota [9–11]. Recognition was also growing that the resident microbiota were involved in regulating gut development and function [12,13].

Finally, another key catalyst for the study of the human microbiome was the publication of the first drafts of the human genome sequence. Relman and Falkow [14] noted on this occasion that a “second human genome project” should be undertaken to produce a comprehensive inventory of microbial genes and genomes associated with the human body. Led by Davies [15], this research community renewed the call for the study of the human-associated micro-organisms in development and in health and disease. Early metagenomic studies using the next-generation sequencing methods, which were just appearing, examined one of the most complex of human microbiomes, the large colon of the gastointestinal tract [16,17] and demonstrated the tremendous complexity and metabolic potential of the gut microbiome. By the middle of the last decade, sequencing costs began to fall sufficiently to provide the opportunity for NIH to consider extensive surveys of the microbial communities associated with the human body.

* Address: Human Microbiome Project, National Human Genome Research Institute, National Institutes of Health, 5635 Fisher Lane, Suite 4076, Bethesda, MD 20852, USA. Tel.: +1 (301) 496 4550.
E-mail address: lita.proctor@nih.gov.

http://dx.doi.org/10.1016/j.siny.2016.05.002
1744-165X/Published by Elsevier Ltd.

2. The NIH Human Microbiome Project: a community resource

The micro-organisms that live in and on the human body are more than a collection of microbes. The microbial communities that encompass all microbial life (bacteria, archaea, eukaryotic viruses, bacteriophage and eukaryotic microbes such as fungi and protozoa), their genes and genomes and their collective metabolic properties have co-evolved with us as a hominid species. These communities can be found on every epithelial surface of the body and interact with us to assist in digestion and detoxification, support immunity, protect against invading pathogens and maintain our overall health. These microbes interact with each other and with the human host and it is these interactions which have developed over millennia that have led to what we know as the modern human microbiome. At trillion of cells, thousands of species, and at least 20 million unique microbial genes (compared to the 20,000–25,000 human genes in the genome), the global human microbiome constitutes the largest genetic component of the human superorganism. To catalyze this new field of science, in 2008 the NIH Common Fund Office formally launched the Human Microbiome Project (HMP) (http://commonfund.nih.gov/hmp/). This program was designed as a community resource project in two phases and these are described below. Both phases were designed to provide methods, tools, protocols, data and other research resources to support this emerging field.

The first phase of HMP, begun in 2008, involved the development of clinical protocols, computational tools, 16S and whole genome shotgun metagenomics datasets, reference microbial genome sequences and a microbial strains collection as community resources. In addition, three programs in technology development, computational tools development and in the ethical, legal and social implications of microbiome research were created to support the field. A Data Analysis and Coordination Center (DACC) supported the sequencing and data analysis efforts of the 300-plus member HMP Research Network Consortium and served as a portal to the primary data sets, the derived data, the reference genome sequences catalog, the computational tools and the other resources (http://www.hmpdacc.org).

A major focus of phase one was to complete a survey of microbial communities in humans. The major scientific objectives were to evaluate the composition of microbiomes: (i) of healthy adults to determine whether there was a characteristic microbiome associated with a health phenotype, and (ii) of cohorts with putative microbiome-associated diseases to determine whether there were characteristic microbiomes associated with these diseases. These latter studies were known as Demonstration Projects. These two key cohort studies are highlighted in the next sections.

3. The healthy adult cohort study

Most human microbiome studies focus on a specific region or organ system of the body. The largest study to date of the microbiomes of five major body regions of healthy adults (airway, skin, oral cavity, gastrointestinal tract, and vagina) was undertaken by HMP. All five major body regions were sampled simultaneously multiple times over multiple years. As each major body region had multiple sites (i.e. habitats), a total of 18 habitats were sampled at each time, and as the volunteers were clinically verified to be free of overt disease in all of the body regions, this study was known as the healthy adult cohort study.

Extensive exclusion criteria for selection of healthy volunteers were used, based on a combination of health history, use of antibiotics, probiotics or immunomodulators, as well as physical examination of each body region of each volunteer (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000228.v3.p1). Three hundred 18–40-year-old subjects, with an equal number of men and women, were enrolled; 20% self-identified as a racial minority, and 11% as Hispanic. Of the 300 volunteers in this study, >90% were sampled twice and 33% were sampled a third time over ~2 years. Among the 18 habitats sampled, all were directly sampled except for the gut tract, for which stool (i.e. feces) served as a proxy. The body regions of all subjects were clinically examined before each sampling and verified to be free of disease; further, these subjects could not use antibiotics, probiotics or immunomodulators or bathe or brush their teeth prior to each sampling. Blood was collected for whole genome sequencing of the study subjects, a project which is currently underway.

Two types of DNA sequence analysis were completed for this study. 16S sequencing was performed on all samples and these data were used to develop microbial community diversity profiles. A subset of the samples were also analysed by whole genome shotgun metagenomic sequence analysis, which involves sequencing the total DNA in a sample. Not only were these metagenomic sequence data used to develop microbial community diversity profiles, but also new tools were developed to support analyses of the metabolic potential of these communities through metabolic pathway reconstructions [18]. In addition, ~3000 microbial strains of nonpathogenic isolates collected from all regions of the body were sequenced to create a microbial reference genome sequence catalog. These genome sequences were used as scaffolding for organizing the metagenomic sequences and as a reference for predicting the potential metabolic pathways that are present in the metagenomes.

All of the primary data were deposited in NIH's public archives for sequence data under the National Center for Biotechnology Information (NCBI) Bioproject system. All HMP 16S sequence data can be found in Bioproject 48489 (http://www.ncbi.nlm.nih.gov/bioproject/48489). All HMP whole genome shotgun metagenomic data can be found in Bioproject 43017 http://www.ncbi.nlm.nih.gov/bioproject/43017). All HMP reference genome sequence data can be found in Bioproject 28331 (http://www.ncbi.nlm.nih.gov/bioproject/28331). Protected data, such as human sequence data and specific clinical metadata, are deposited in NCBI's Genotype and Phenotype controlled access SRA database (dbGaP, http://www.ncbi.nlm.nih.gov/gap). Access to these protected data are regulated through Data Access Committees at each NIH Institute; the procedure for requests for these data can be found at dbGaP (https://dbgap.ncbi.nlm.nih.gov). The clinical protocol from this study was also published [19].

The HMP Consortium published two landmark papers in 2012 which included an initial analysis of the data collected in the healthy cohort study [20,21]. These results described the range of microbial community diversity and predicted metabolic profiles that can be found among clinically verified healthy adults in a western population. These studies demonstrated that each body region supported its own unique microbial community so that, for example, the microbial composition of all GI tract microbiomes was more similar between subjects than were the microbiomes within any one subject. From an ecological viewpoint, each body site is a unique microbial habitat.

One other major result which emerged from these studies may seem confusing at first but speaks to fundamental properties of microbial communities. Unlike community composition, metabolic pathway reconstructions of the metagenomics data for any one habitat or body region were much more constant. That is, although microbial taxonomic composition varied among healthy individuals, the collective metabolic functions of these microbial communities were remarkably similar across any one body region of these subjects. In other words, the metabolic capabilities of the