Computational and Structural Biotechnology Journal 13 (2015) 390–401

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Mini Review

Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome

Shirley Bikel [a,c], Alejandra Valdez-Lara [a,c], Fernanda Cornejo-Granados [a,c], Karina Rico [a], Samuel Canizales-Quinteros [a], Xavier Soberón [b], Luis Del Pozo-Yauner [b], Adrián Ochoa-Leyva [a,c,*]

[a] Unidad de Genómica de Poblaciones Aplicada la Salud, Facultad de Química, UNAM, Instituto Nacional de Medicina Genómica (INMEGEN), México, D.F. 14610, Mexico
[b] Instituto Nacional de Medicina Genómica (INMEGEN), México, D.F., Mexico
[c] Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de Mexico, Avenida Universidad 2001, Cuernavaca C.P. 62210, Mexico

ARTICLE INFO

Article history:
Received 11 March 2015
Received in revised form 1 June 2015
Accepted 4 June 2015
Available online 9 June 2015

Keywords:
Metagenomics
Metatranscriptomics
Viromics
Human microbiome
Systems-level
Bioinformatics

ABSTRACT

The advances in experimental methods and the development of high performance bioinformatic tools have substantially improved our understanding of microbial communities associated with human niches. Many studies have documented that changes in microbial abundance and composition of the human microbiome is associated with human health and diseased state. The majority of research on human microbiome is typically focused in the analysis of one level of biological information, i.e., metagenomics or metatranscriptomics. In this review, we describe some of the different experimental and bioinformatic strategies applied to analyze the 16S rRNA gene profiling and shotgun sequencing data of the human microbiome. We also discuss how some of the recent insights in the combination of metagenomics, metatranscriptomics and viromics can provide more detailed description on the interactions between microorganisms and viruses in oral and gut microbiomes. Recent studies on viromics have begun to gain importance due to the potential involvement of viruses in microbial dysbiosis. In addition, metatranscriptomic combined with metagenomic analysis have shown that a substantial fraction of microbial transcripts can be differentially regulated relative to their microbial genomic abundances. Thus, understanding the molecular interactions in the microbiome using the combination of metagenomics, metatranscriptomics and viromics is one of the main challenges towards a system level understanding of human microbiome.

Contents

* Corresponding author at: Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de Mexico, Avenida Universidad 2001, Cuernavaca C.P. 62210, Mexico.
E-mail addresses: adriano8a@gmail.com, aochoa@ibt.unam.mx (A. Ochoa-Leyva).

http://dx.doi.org/10.1016/j.csbj.2015.06.001

1. Introduction

The human body is inhabited by a high diversity of bacteria and archaea, as well as fungi, protozoa and viruses. These microbes inhabit in several niches within the human body sections and are collectively known as the human microbiota, whereas their collective genomes form the human metagenome [1]. The advance in high-throughput technologies used to analyze the components of the microbiome have substantially improved our knowledge of the microbial communities associated to human niches [2,3]. The decrease in cost of sequencing using high-throughput technologies has enabled large-scale studies of the human microbiome, revealing high interindividual variability of the microbiota composition and in different body sites within individuals [4–7].

Changes in abundance and composition in the fecal microbiota (dysbiosis) has been observed in patients with several human diseases, ranging from inflammatory bowel disease and obesity to diabetes and neurological disorders [8–17]. The fundamental objectives of human microbiome research is to study the structure and dynamics of microbial communities, the relationships between their members (microorganisms, viral particles and host) and their potential association with health and disease. The study of interactions between the DNAs, RNAs, and viruses that are present in the microbiome, are the main interest of metagenomics, metatranscriptomics, and viromics, respectively. For studying the microbial community of the human microbiome using high throughput sequencing technologies, there are several types of large scale analyses: the 16S profiling analysis which is based in sequencing the hypervariable regions of the 16S rRNA gene and the shotgun analysis which is based in direct sequencing of the total DNA (metagenome) and/or total RNA (metatranscriptome). In addition, the viral component of the microbiome (virome) can be also analyzed by sequencing the total viral particles.

In the last decade, many studies using the sequencing of the 16S rRNA gene to characterize the microbiota composition have been conducted, however, this analysis mainly identifies the abundance and diversity of bacteria and archaea in the sample. Although, there is a computational approach to predict the metagenome functional composition by the 16S rRNA gene sequences [18]. The metagenomic analysis also identifies the abundance and diversity of microbial community, but additionally can identify the gene content and inferred functional potential of proteins encoded in the genomes of the microbial community. The metatranscriptomic analysis allows the identification of expressed transcripts in the microbiome. The transcript numbers can also be used to compare the gene expression profiles between microbial communities. In addition, for comparative study, metatranscriptomic data must be paired with metagenomic data in order to analyze if the transcript abundance is reflecting changes in community composition [19,20]. The use of high throughput sequencing technologies to analyze human metagenomes has also revealed the existence of many bacteriophages in metagenomes [21]. Interestingly, it has been proposed that bacteriophages may have a role in shaping the diversity and composition of the oral and gut bacteria [22,23]. The involvement of phages in microbial dysbiosis may indirectly contribute to the disease. In this regard, a model suggesting that virome may contribute to the intestinal inflammation and bacterial dysbiosis was recently reported in the human gut microbiome [24]. However, studies that involve metagenomic or metatranscriptomic combined with viromic analysis are still necessary to understand the molecular interactions within the human microbiome and their relevance in health and diseased states.

The microbiome has been conceptualized as a dynamic ecological community consisting of multiple taxa each potentially interacting with each other, the host and the environment [25]. Hereafter, we use the term microbiome, to refer to the microbial communities and viruses in conjunction with the environment they inhabit, interacting as a system. In the first section of this review, an overview of the different sequencing, experimental, and bioinformatic procedures that have been used to study the human microbiome are discussed. In the second section, the recent advances combining metagenomic, metatranscriptomic and viromic analyses to identify the molecular dialog within the microbiome are dicussed. The metabolomic and metaproteomic analyses are not in the focus of this review.

2. Sequencing and bioinformatic strategies to study the human microbiome

2.1. 16S rRNA gene profile analysis

The small ribosome subunit 16S gene (16S rRNA gene) is used as a housekeeping genetic marker to study bacterial phylogeny and taxonomy as it is highly conserved between different species of bacteria and archea. In addition to highly conserved regions, the 16S rRNA gene contains hypervariable regions that are used to identify between different bacteria. Furthermore, some bacteria have a different copy number of the 16S rRNA gene, often existing as a multigene family, or operons. Hence, the 16S rRNA gene sequencing has become typically used to identify and quantify bacterial taxa present within a microbiome sample. 16S rRNA profiling relies on using PCR 'universal' primers targeted at the conserved regions and designed to amplify a range of different microorganism as wide as possible. The amplified fragments (amplicons) of the gene correspond to selected short-hypervariable regions ranging from V1 to V9, making it faster and cheaper to sequence with high throughput technologies than many other bacteria genes (Figs. 1 and 2). Two of the most significant limitations of 16S rRNA sequencing that should be considered before starting a sequencing project are: (1) the introduction of biases by selection of the 16S rRNA hypervariable regions and (2) the introduction of biases by PCR primer design, which may select for or against particular groups of microorganisms (Fig. 2). To minimize the biases introduced by primer design, the primers include degenerated bases and can be used at lower hybridization temperatures to capture more microbial diversity. Other problems using the PCR is that bacterial contamination of reagents may be affecting the results [26] and that the 16S rRNA gene is also present in different copy numbers in bacterial genomes influencing the apparent relative abundance of a microorganism [26].

The sequence fragments obtained by high throughput sequencing technologies are typically called sequence reads. Longer read lengths (1000 bp), as the ones obtained by 454/Roche technology, can span multiple hypervariable regions of the 16S rRNA gene, increasing the number of the microorganisms that can be identified at species level. Although, this technology is cost prohibitive and it will be discontinued in 2016. However, short-read length sequences spanning only one hypervariable region has sufficient resolution for the accurate taxonomic assignments [27,28]. The optimal community clustering confirmed with sequence reads of this length is an important advance in amplicon design because sequencing only one hypervariable region is more cost-effective (Table 1). There is a large amount of PCR primers to amplify different hypervariable regions of 16S rRNA gene for sequencing in the short read sequencing platforms. Although, the compatibility of the fragment length should be according with sequencing platform read length capacity. There are many studies that target different regions of the 16S rRNA gene, for example V3–V5 [29], V1–V2 [30], V1–V3 [26], V4–V5 [31], and V8–V9 [32]. There is an active discussion about the hypervariable region that should be sequenced to perform a microbial diversity analysis. For example, the V6 region is not optimal for sequencing analyses that are directed for taxonomic assignment and community clustering, as opposed to sequence reads spanning the V2 and V3 regions [33,2]. The most informative 16S rRNA gene region to amplify may also depend of the analyzed environment, for example, in a study for the diagnosis of pathogenic bacteria Chakravorty at al., showed that in a mix of 110 different bacterial species including common blood borne pathogens, CDC-defined agents and environmental microflora, the V2 and V3 were most suitable for distinguishing all