Best Practice & Research Clinical Gastroenterology xxx (2017) 1–5

Contents lists available at ScienceDirect

Best Practice & Research Clinical Gastroenterology

journal homepage: https://ees.elsevier.com/ybega/default.asp

Microbiome measurement: Possibilities and pitfalls

Djawad Radjabzadeh [a], André G. Uitterlinden [a, b], Robert Kraaij [a, b, *]

[a] Dept. Internal Medicine, Erasmus MC, Rotterdam, The Netherlands
[b] Dept. Epidemiology, Erasmus MC, Rotterdam, The Netherlands

Keywords:
16S rRNA profiling
Gut microbiome
Study design

A B S T R A C T

Microbiome research is an emerging field in medical sciences. Several studies have made headways in understanding the influence of microbes on our health and disease states. Further progress in mapping microbiome populations across different body sites and understanding the underlying mechanisms of microbiome-host interactions depends critically on study design, collection protocols, analytical genetic techniques, and reference databases. In particular, a shift has appeared going from small sample collections to large-scale population studies (with extensive phenotypic information including disease status) which calls for some adaptions. In this review we will focus on gut microbiome profiling using the 16S ribosomal RNA approach in the setting of large-scale population studies, and discuss some novel developments.

1. Introduction

Human microbiome-host interactions are gaining much interest and are driving investment in medical research. This is due to the fact that traditional, cultivation-based approaches to identify microbiota have been replaced by high-throughput DNA-sequencing technologies that allow profiling complete microbiomes of thousands of samples at acceptable costs with considerably less time and labour. Additionally, computational methods, which are used to explore and analyse multidimensional microbiome datasets, have been improved. The importance of the microbiome has been demonstrated extensively in the past years. Many studies have reported associations of differences in the composition and functions of the microbiome of different body sites with various pathologies and traits [1–4]. It has been shown, for example, that the gut microbiome can alter host metabolism [5] and changes in its composition have been reported to have relations with gastrointestinal (GI) [6–8] and non-GI [9,10] diseases, including neurological conditions [11]. Nowadays, large population based reference panels are available that are successful in capturing the significance of biological patterns like age [12], core human microbiome composition [13] or of a disease condition such as Crohn's disease [14]. Two recently published studies of Belgian and Dutch cohorts are good examples of what can be achieved through large population-based studies in terms of understanding the factors important in structuring the gut microbiome [15,16]. These studies together (N = 3948) revealed several markers for gut microbiome composition variation and diversity. They found sixty-nine clinical and questionnaire-based covariates that were associated with microbiome compositional variation with a 92% replication rate. Stool consistency (Bristol Stool Scale) had the largest effect, whereas medication explained the largest total variance. Furthermore, a recent extension of the Human Microbiome Project (HMP), that now comprises 2355 meta-genomes from different body sites, indicated that besides large, also longitudinal panels are needed to capture biological effects [17] Fluctuations in gut microbiome composition were detected over periods of months although the extend was less than the inter-individual differences in composition.

The two main questions in microbiome studies are 'Which microbes are there?' and 'What are they doing?' These questions are addressed by three predominant approaches: 16S ribosomal RNA (16S rRNA) profiling, meta-genomics and *meta*-transcriptomics. 16S rRNA profiling is the most direct and cost-effective approach to obtain phylogenetic profiles. Although meta-genomics performs less well to determine taxonomy, it is not restricted to the domains of Archaea and Bacteria and it will provide the total functional content of the sample. *Meta*-transcriptomics will profile the functions that are actually expressed in the community. Besides these sequencing-based approaches, array-based approaches are now becoming available. In 2016, Affymetrix has introduced the Axiom

* Corresponding author. Dept. Internal Medicine, Erasmus MC, Rotterdam, The Netherlands.
E-mail address: r.kraaij@erasmusmc.nl (R. Kraaij).

https://doi.org/10.1016/j.bpg.2017.10.008
1521-6918/

Microbiome Array that targets over 12,500 species across most domains, including Fungi and Viruses at costs in between 16S rRNA and meta-genomics profiling. The current version determines presence or absence of species, which is in good concordance with 16S rRNA profiling (unpublished results). Further development and extension of this technology and further price reduction will result in an attractive alternative to the sequencing-based profiling approaches. Additionally, *meta*-proteomics and *meta*-metabolomics are approaches, which do not require sequencing and don't provide taxonomic information, but, in combination with one of the other approaches, give more detailed information on actual activities of the microbes. Depending on the type of sample, type of microorganism or the methods that are going to be used to characterize the microbiome in a study, DNA, RNA, protein or metabolites or combinations of these need to be extracted from samples for subsequent phylogenetic or functional profiling. The extracted biomaterials must be translated into readable data (DNA and RNA sequences, or protein and metabolite spectrums) that should pass several bioinformatics steps to obtain clean and error-free datasets. From study design to data analysis, choices made at every step can have high impact on the results.

This review addresses these specific technical issues in microbiome research. We focus on 16S rRNA profiling of the human gut microbiome, since the majority of microorganisms are located in the gut and most studies are addressing this community by means of 16S rRNA profiling.

2. Possibilities and pitfalls in 16S rRNA profiling of the human gut microbiome

In general (Fig. 1), 16S rRNA gut microbiome profiling starts with the collection of stool samples and sample metadata. A small amount of the sample is homogenized in a homogenizing buffer and after a 'bead-beating' step - a process by which cells are disrupted by physical forces of glass-beads - microbial DNA is isolated. In a subsequent PCR step, selected hyper-variable regions of the 16S rRNA gene are amplified. The 16S rRNA gene encodes the 16S ribosomal RNA that is part of the small subunit of ribosomes, which can be found in Bacteria and Archaea (Eukaryotes have 18S rRNA). The gene consists of nine hyper-variable regions (V1 – V9) flanked by conserved regions. The conserved regions allow designing primers that target as many bacterial species as possible, while the variable regions are used for species identification [18]. The amplicons are purified and sequenced. The resulting sequences are clustered into Operational Taxonomic Units (OTUs) and taxonomy is assigned to each OTU using 16S rRNA databases to generate an OTU table. This table, together with the information collected in the beginning and other phenotype data, is used in downstream analyses.

2.1. Evaluation of biological factors affecting the gut microbiome

Apart from sampling microbiota from an individual it is important to also collect phenotypic information such as age, sex, ethnicity, antibiotic use, dietary patterns etc. Applying microbiomics in well-characterized epidemiological study populations ensures availability of such data. It is crucial that study designs consider the effect sizes of these biological factors. They become important when the effect of the phenotype being associated with gut microbiome variation is confounded by these factors. Some of these biological factors like diet, medication, diseases, body mass index (BMI), age and ethnicity have been reported to (strongly) influence the microbiome composition of the gut [15,16,19]. The effect sizes of some of these biological confounders, such as antibiotics, proton pump inhibitors and metformin, are so strong that they can dramatically alter gut microbiome composition and should be considered as exclusion criterion or call for stratified analyses [20–22]. During final analyses, matching groups in categorical data on known biological confounders can control for confounding. If matching is satisfactory, no further adjustment for the confounding factors is needed. However, matching is often not successful, since the phenotypic effect being investigated is strongly confounded by these biological factors and finding the same range of data within the categorical data being investigated is not possible. When the phenotype in the study is continuous, the matching method should be replaced by other methods like linear regression, in which the result can be adjusted for the confounding effects.

2.2. Evaluation of technical factors affecting the gut microbiome

Technical factors can also influence the composition of the gut microbiome and this influence can be larger than biological influences. The technical sources include sample collection and storage protocols, DNA extraction, and selection of the hypervariable region(s), PCR methods, sequencing platform and bioinformatic analysis methods.

2.3. Sample collection and storage techniques

For gut microbiome, the best procedure for stool sample collection is to homogenize the stool sample immediately after defecating followed by immediate sample preparation using part of the homogenate [23]. However, in most large cohort studies this approach will not be feasible. This is due to the fact that in large cohort studies sample production at home is necessary and collecting samples from the participant's home is not practical and economically favoured. Therefore, deviations from this protocol need to be made and optimized. It should be realized that when the molecular activity of the microbiota will be studied (e.g. RNA expression, proteins or metabolites), in which samples may experience strong effects from the environmental fluctuations, direct storage at −80 °C or direct sample preparation is needed. DNA, on the other hand, is more stable and allows for less stringent approaches such as collection at ambient temperature for a short period of time [24–28]. Alternatively, samples could be stored at the participant's home freezer (either at 4 °C or −20 °C) and subsequently collected by the researchers. Modern freezers, however,

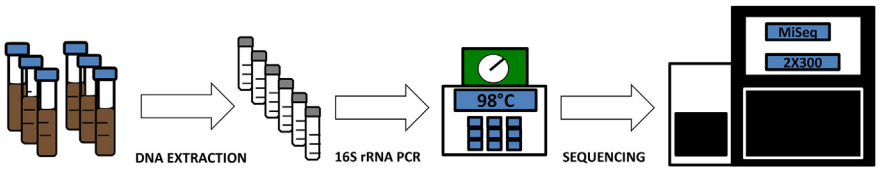

Fig. 1. Overview of the wetlab part of 16S ribosomal RNA profiling. DNA is extracted from the faecal samples and a part of the 16S ribosomal RNA gene is amplified and sequenced. A phylogenetic profile is generated in subsequent bioinformatic analyses.

Please cite this article in press as: Radjabzadeh D, et al., Microbiome measurement: Possibilities and pitfalls, Best Practice & Research Clinical Gastroenterology (2017), https://doi.org/10.1016/j.bpg.2017.10.008