Experimental approach, study design, and statistical challenges in microbiome research

Shirin Moossavi, MD PhD(c)

Tehran University of Medical Sciences
4 July 2019
Outline

- Experimental approaches
- Study design
- Data analysis
Experimental approaches

Common untargeted approaches
Culture-enriched molecular profiling
Animal and other experimental models
Common (untargeted) methodologies used to study human microbiota

The choice of the experimental approach depends on the research question.
Amplicon sequencing

- Based on sequencing of a short amplicon of 16S rRNA gene (bacteria) or 18S rRNA/ITS (fungi)
- These genes contain hypervariable regions
  - Function as a molecular clock
  - Useful for phylogenetic classification
- Advantages:
  - Relative easy and inexpensive
  - A good first step profiling method
- Disadvantages
  - Low resolution: can not identify species or strains
  - Does not capture the genetic (functional) variability between different members of the same bacterial group
  - Different primers have biases against different groups of bacteria
Physiology-informed amplicon sequencing

- Amplicon sequencing can inform about functional aspects of interest if we separate bacteria into different functional groups before sequencing.

- **IgA-Seq:** separate the bacteria based on IgA-coating.

- Other physiologically-informed targets to identify for live and/or proliferating bacteria.
Relative vs. absolute quantification

- Differences in bacterial load will not be captured by relative quantification by sequencing.
- Solution: enumeration of equal number bacterial cells from all samples prior to DNA extraction.

Weaknesses of the sequencing approach

- **Biases**
  - Extraction
  - Primer
  - Bioinformatics
  - Depth of sequencing

- ** Doesn’t show viability**

- ** Doesn’t capture phenotypic variation**

Culture-enriched molecular profiling

- Cultivating the sample on different culture media (up to 100)
- Aerobic and anearobic
- Useful for different sample types (e.g. stool, sputum, milk)
- A large proportion of taxa identified by sequencing can be cultured
- Bacterial isolates could be used for downstream functional analysis and strain genotyping
- In low biomass samples (e.g. milk) culture can inform about the potential contaminants

Lau et al. (2016) Genome Medicine;8:72
Applications of culture-dependent approaches

- High throughput drug-bacteria interaction
  - E.g. Digoxin metabolism by *Eggerthella lenta*
- High throughput bacteria-bacteria or bacteria-bacteriophage interaction
- Improve bacterial reference genome catalogue
- Causal study of specific strains or sub-communities and phenotypes of interest

Causal investigation using animal models

- In microbiota research, mouse models are being increasingly used to study the role and functioning of the microbiota and its association with diseases.
- Allow perturbations in gut microbiota to be studied in a controlled experimental setup, and thus help in assessing causality of the complex host-microbiota interactions and in developing mechanistic hypotheses.

- Experimental manipulations of murine models in gut microbiota
  - Host genetic background manipulation (gene knockouts)
  - Gut microbiota composition manipulation (controlled inoculation in germ-free or gnotobiotic mice)
  - Ecosystem interventions including dietary interventions, antibiotic treatment and fecal transplantations.

Nguyen et al. (2015) Disease Models & Mechanisms; 8: 1-16
Important pitfalls to consider

- Anatomical differences in the GI tract due to their diverging diets, feeding patterns, body sizes and metabolic requirements
- Differences in the gut microbiota composition
  - Human: stool sample
  - Mouse: cecal sample
- Microbiome variability between different strains of mice and between different laboratories

Nguyen et al. (2015) Disease Models & Mechanisms; 8: 1-16
Other experimental approaches

May et al. Emerging Topics in Life Sciences (2017) 1:385–400
Study design

General recommendations
Sample size estimation
Sample collection and storage
Sample processing and sequencing
General recommendations

- The success of a microbiome study relies on defining clear scientific questions and objectives.
- The choice of the experimental approach depends on the research question.
- Based on the objective of the study, relevant control groups and sample collection timing and frequency should be selected.
- Measurements (e.g., metabolite concentrations, disease activity score) complementary to the microbiome should be considered to expand the scope of the study and uncover additional insights.
- Quantifying and mitigating possible confounders is a must.
- Ensuring consistency:
  - The same procedures and methods throughout the study.
  - Record the maximum information about participants, samples, and experimental procedures.
- The validation of taxa by the analysis of new samples from independent cohorts is important to show reproducibility and to increase confidence in biological conclusions.
Additional considerations

- Microbiome-specific questions
  - Applicable to all samples: antibiotics, home environment,
  - Sample type specific:
    - Stool: consistency, history of GI disease, diet
    - Biopsy: active lesion, distance from the legion
    - Milk: use of pump
    - Saliva: time of last meal, time of last smoking
    - Vaginal: menstruation cycle
    - Skin: region, time of last shower/bath, personal hygiene practice

- Outcome-specific questions
Sample size estimation

- Effect size and the target population must be considered when deciding on the number of samples to collect

- Factors to take into account:
  - High degree of inter-individual variability in the overall composition
  - The anticipated depth of sequencing after quality filtering and denoising

- A few methods have been proposed for microbiome sample size estimation (Kelly et al., 2015; La Rosa et al., 2012)

- As a rule of thumb, it is estimated that a sample size of 50-500 is sufficient to detect microbiota shift at significance level of 5% and sequencing depth of 10,000 (FGFP).
Sample collection and storage

- Temporal, spatial, and personal consistency
- Standardizing controllable factors to reduce confounders.
- Negative controls especially for low biomass samples
- Sample collection metadata:
  - Date and time of collection
  - Time to storage at -80 °
  - Sample collection method e.g. manual vs. pump for breastmilk
- Sufficient quantity for future experiments
- Avoid freeze-thaw
Alternative collection and storage could be suitable for 16S rRNA sequencing.

Fecal microbiota composition is separated based on the individual and not the storage method.

A: Immediate freezing at -80 °C; B: 95% ethanol 48h at RT; C: Card 48h at RT.
Sample processing and sequencing

- Choice of the extraction kit
  - Lysis method: enzymatic, mechanical, heat-induced; affects Gram positive and fungi populations
  - Introduction of reagent contaminants

- PCR amplification considerations
  - Should be done in duplicate or triplicate to minimize bias
  - Choice of primer introduces bias against specific groups of bacteria
  - Most commonly used primers target hypervariable regions of 16S rRNA (bacteria), 18S rRNA or the internal transcribed region (ITS) (eukaryotes)
  - Number of PCR cycles: should be kept minimum

- Negative controls in extraction and sequencing

- Positive controls:
  - Mock community: accuracy of extraction and sequencing
  - Biological controls: variability between experiments
  - Spike-in dilution: to identify the limit of detection (helpful for low biomass samples)
Challenges of microbiome data analysis
Choices of experimental procedures and bioinformatics and analysis approaches
Challenges of microbiome data analysis

- **Compositionality:**
  - Depth of sequencing limited by the capacity of the sequencing machine
  - Relative abundances sum to 1
  - It is on in a simplex space (degree of freedom = n-1)

- **Sparsity**
  - Contains a lot of zero

- **Relative vs. absolute abundance**
  - Misrepresentation of the reality if we don’t consider bacterial load

- **Reproducibility & repeatability**
  - A major challenge in microbiome research
Wet lab researcher degrees of freedom

Different choices at each step can affect the results
Bioinformatic researcher degrees of freedom

Different choices at each step can affect the results

Greg Gloor, ISAPP 2019
Further considerations

Sex and gender as effect modifying factors
Theory-driven vs. data-driven analysis
Data quality control
Further considerations

- Consider sex as a biological factor and gender as a sociocultural factor

- Importance of theory-driven vs. data-driven approach

Moossavi et al. Cell Host Microbe. 2019; 25, 324–335
Importance of quality control prior to analysis when there are multiple batches

- **Contaminant Removal**
  - decontam
  - Comparison of prevalence between batches

- **Sequencing Accuracy Assessment**
  - Mock community
  - Biological controls

- **Repeatability & Reproducibility of Results**
  - Repeatability in the 1st batch
  - Reproducibility in the other batch(es)

Moossavi et al. (in preparation)
Let’s face it: microbiome usually has small contribution to diseases

- Importance of paying attention to the effect sizes
  - Exception: role of antibiotic-induced microbiota disruption in pathobiont bloom e.g. *Clostridioides difficile* colitis or yeast vaginosis

- Importance of system level interrogation of multiple molecular and clinical features to understand the causal mechanism(s)
  - Deep phenotyping

- Importance of epistemological understanding of cause and effect
  - e.g. If event A happens as a result of exposure B in 10% of cases, would B be the cause of A?
  - COMPLEX diseases are COMPLEX and we need to take this COMPLEXITY into account
  - Microbiome dynamic is to some degree stochastic (random); we need to take this into account as well.
Acknowledgement

Azad Lab
Meghan Azad, Kelsey Fehr
Faisal Atakora, Kozeta Miliku
Michelle La, Deborah Chan
Sarah Turner, Reham Aliani

Gut Microbiome Laboratories
Ehsan Khafipour
Martin Nyachoti

CHILD Study
Allan B. Becker
Piushkumar J. Mandhane
Stuart E. Turvey
Padmaja Subbarao
Theo J. Moraes
Diana L. Lefebvre
Malcolm R. Sears

University of California San Diego
Lars Bode
Bianca Robertson

University of Alberta
Catherine Field
Sue Goruk

Queen's University
Qingling Duan
Jihoon Choi
Amirthagowri Ambalavanan

University of British Columbia
Stuart E. Turvey
Brett Finley
Hind Sbihi
Rozlyn Boutin

University of Toronto
Jeffery Brook

McMaster University
Michael Surette
Hooman Derakhshani