Microbial community analysis in R

CascadiaR Conference
June 3, 2017
Lisa Karstens, PhD
Microbes play an important role in human health

We are not alone...

humans are supraorganisms
– Symbiotic relationship with microbes
– Important for health and disease
Early methods microbial investigation relied on culturing techniques

High-throughput sequencing technology has brought huge growth in microbiome investigations.

Metagenome sequencing
*Metagenome* – entire genome of a community

Targeted gene sequencing
*16S rRNA gene* – marker gene for prokaryotes

DNA extraction

- Library prep
- Whole genome sequencing
- PCR amplification of 16S rRNA gene
- Amplicon sequencing
16S sequencing studies generate a lot of data!

One sequencing run can have over 1 million reads from hundreds of samples!
Example workflow processing sequence reads

- Commonly NOT done in R
- Each step is an active area of research
- Many options exist
R offers an ideal environment for processing and analysis

- Reproducible, Organized, Sharable
  - R Studio
    - Interactive, friendly environment
  - R Markdown
    - Documents processing steps
- RData files
  - Sequence data, and sample data, results in one file
Microbiome specific workflows in R

• **dada2**
  - A package to model and correct 16S sequence errors
  - Replaces clustering algorithm for grouping reads


Microbiome specific workflows in R

- **Phyloseq**
  - Tool to import, store, analyze, and graphically display complex phylogenetic sequencing data

Phyloseq object

- OTU table
- Sample data
- Taxonomy table
- Phylogeneic tree

McMurdie and Holmes (2013) [phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data.](https://doi.org/10.1371/journal.pone.0061217)
R offers an ideal environment for processing and analysis

- **Exploratory Data Analysis**
  - Data structure
    - Understand variables
    - Identify (and fix) potential problems
  - Decision making
    - Identify and evaluate transformations, filtering steps
    - Identify data distributions for appropriate analysis
R offers flexibility for data analysis

- Hypothesis testing
  - DeSEQ2
  - MetagenomeSeq
- Multivariate methods
  - Vegan
  - Mixomics

Challenges in microbiome workflows in R

- Requires user to know how to use R
- Appropriate application of packages can be challenging
- Storing raw sequencing data in R is not ideal
- Common processing algorithms are not available

Thank You

Acknowledgements

• Students:
  • Vincent Caruso
  • Eric Leung
  • Mark Klick

• Mentors and Collaborators
  • Shannon McWeeney, PhD
  • Jim Rosenbaum, MD
  • Mark Asquith, PhD
  • Damien A. Fair, PA-C, Ph.D
  • Rahel Nardos, MD, MCR
  • Tom Gregory, MD
  • Rosenbaum Lab
  • Fair Neuroimaging Lab

Funding

K12 HD043488

Collins Medical Trust