Shiny-phyloseq: Web Application for Interactive Microbiome Analysis with Provenance Tracking

Paul J. McMurdie
Research Associate
Prof Susan Holmes Group
Statistics Department
Stanford University
Overview

- Intro to Microbiome Research
- phyloseq - a microbiome BioC package
- (RNA-Seq methods solve a microbiome problem)
- Shiny-phyloseq: a shiny interface to phyloseq
What are microbes?

Cell structure

(they don’t all look like this)
What are microbes?

What is a microbiome?

The totality of microbes in a defined environment, especially their genomes and interactions with each other and surrounding environment.

- A population of a single species/strain is a culture, extremely rare outside of lab, some infections

- A microbiome is a mixed population of different microbial species (microbial ecosystem)
Why study **microbiomes**?

Earth Microbiome Project:

- Oceans, soils, waterways
- Wastewater Treatment
- Deep-Sea Hydrothermal Vent
- Cow Rumen
- Human Microbiomes
Human Body Sites, HMP

>10 times more microbial cells than human cells

Entire human microbiome weighs less than 2 kg, at most
Fecal Transplants
(Clostridium difficile infection)

Nature Rev Gastroenterology & Hepatology
Why is microbiome research new?

Bias for cultivable microbes, especially pathogens

- Culture-based methods fail to detect most microbes
- Microbes are easy to miss (except pathogens)
- Most microbes are NOT pathogens (even the human-associated)

Availability of tools limited to last 3 decades

- PCR, fast & cheap DNA sequencing, microarrays, etc
- Discovery of culture-independent techniques - 16S-rRNA
How do we query microbiomes??

16S rRNA

ribosome

ribosome in action
How do we query microbiomes??

- Universal (e.g. 16S rRNA) Gene census
- Shotgun Metagenome Sequencing
- Transcriptomics (shotgun mRNA)
- Proteomics (protein fragments)
- Metabolomics (excreted chemicals)
Microbiome data
heterogeneity and processing
phyloseq

**OTU Abundance**
- `otu_table`

**Sample Variables**
- `sample_data`

**Taxonomy Table**
- `taxonomyTable`

**Phylogenetic Tree**
- `phylo`

**Reference Seq.**
- `refseq`

**data structure & API**

- **Accessors:**
  - `get_taxa`
  - `get_samples`
  - `get_variable`
  - `nsamples`
  - `ntaxa`
  - `rank_names`
  - `sample_names`
  - `sample_sums`
  - `sample_variables`
  - `taxa_names`
  - `taxa_sums`

- **Processors:**
  - `filter_taxa`
  - `merge_phyloseq`
  - `merge_samples`
  - `merge_taxa`
  - `prune_samples`
  - `prune_taxa`
  - `subset_taxa`
  - `subset_samples`
  - `tip_glom`
  - `tax_glom`

**constructor:**
- `phyloseq`
  - `otu_table`
  - `sample_data`
  - `tax_table`
  - `phy_tree`
  - `refseq`

**optional**
- `data`

**import**
- `phyloseq`

**import packages:**
- `OTU Abundance` from `otu_table`
- `Sample Variables` from `sample_data`
- `Taxonomy Table` from `taxonomyTable`
- `Phylogenetic Tree` from `phylo`
- `Reference Seq.` from `refseq`
Side Note: BioC tools for microbiome

dgeR, DESeq(2), metagenomeSeq perform better than popular alternatives in differential abundance detection:

McMurdie and Holmes (2014) PLoS Comp Biol
DOI: 10.1371/journal.pcbi.1003531

http://joey711.github.io/waste-not-supplemental/
## Acknowledgements

<table>
<thead>
<tr>
<th>Name</th>
<th>Contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Susan Holmes</td>
<td>Postdoc Advisor, Mentor, Co-author</td>
</tr>
<tr>
<td>Holmes Group</td>
<td>Helpful advice and feedback</td>
</tr>
<tr>
<td>Wolfgang Huber</td>
<td>Helpful advice and feedback re: DESeq(2)</td>
</tr>
<tr>
<td>BioC and CRAN</td>
<td>Support, Feedback, Distribution of phyloseq and biom</td>
</tr>
<tr>
<td>RStudio</td>
<td>Shiny, RStudio IDE</td>
</tr>
<tr>
<td>Hadley Wickham</td>
<td>ggplot2, reshape2, plyr R packages</td>
</tr>
</tbody>
</table>
Shiny-phyloseq

Live Demo

How to Run:
install.packages("shiny")
shiny::runGitHub("shiny-phyloseq", "joey711")

http://joey711.github.io/shiny-phyloseq/
End.

Questions?