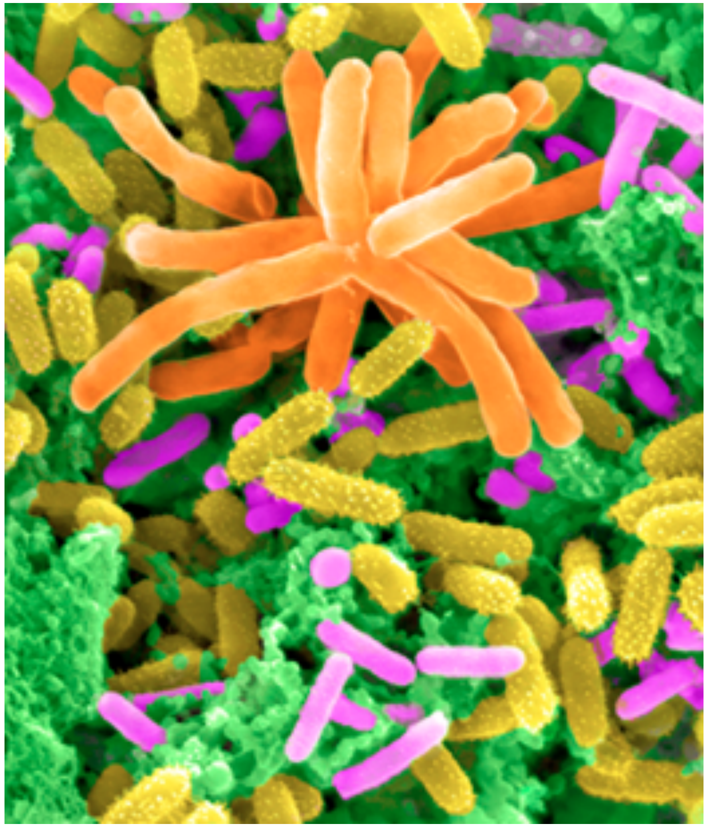
# Shiny-phyloseq: Web Application for Interactive Microbiome Analysis with Provenance Tracking



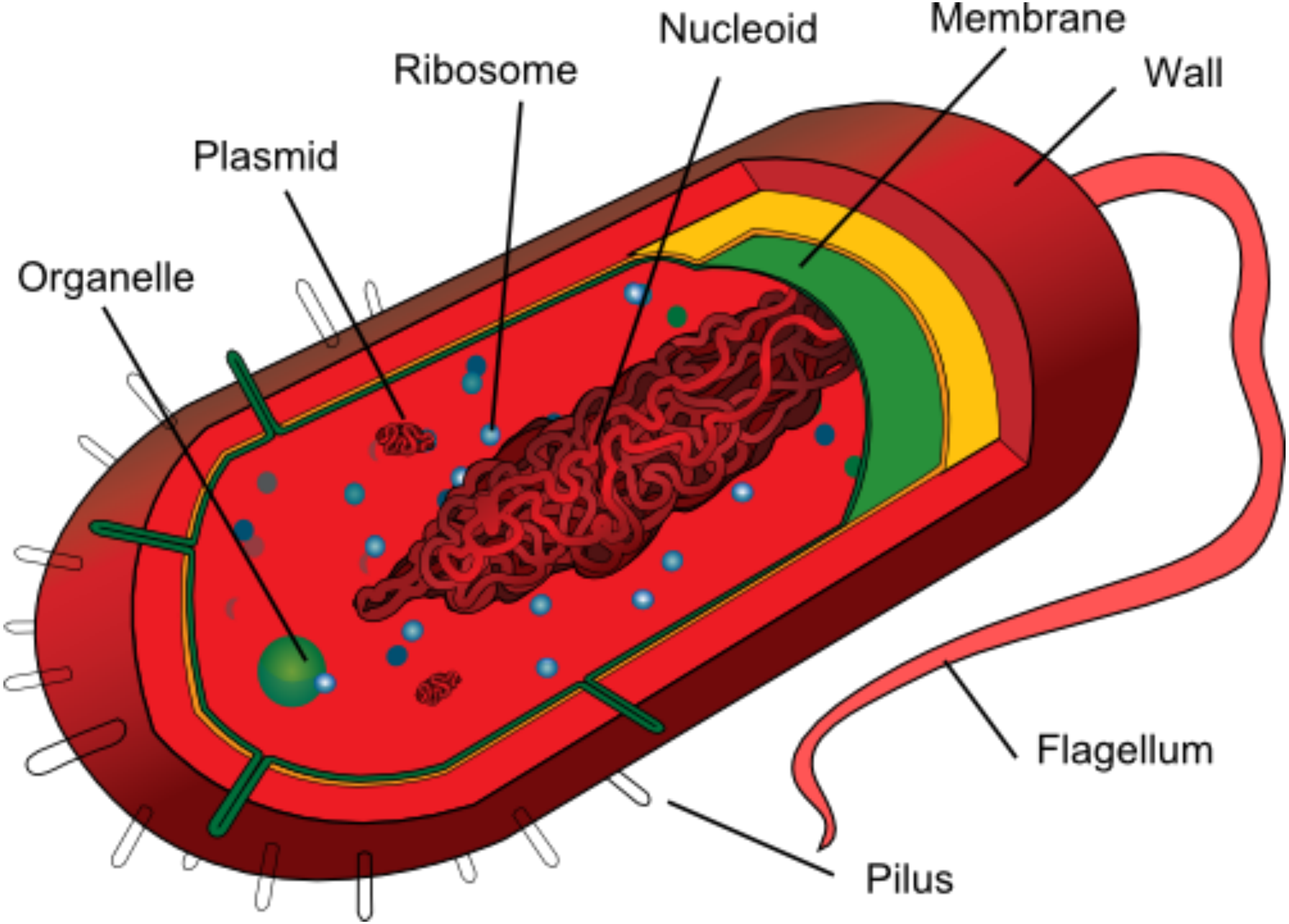
**Paul J. McMurdie**
Research Associate
Prof Susan Holmes Group
Statistics Department
Stanford University

# 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?

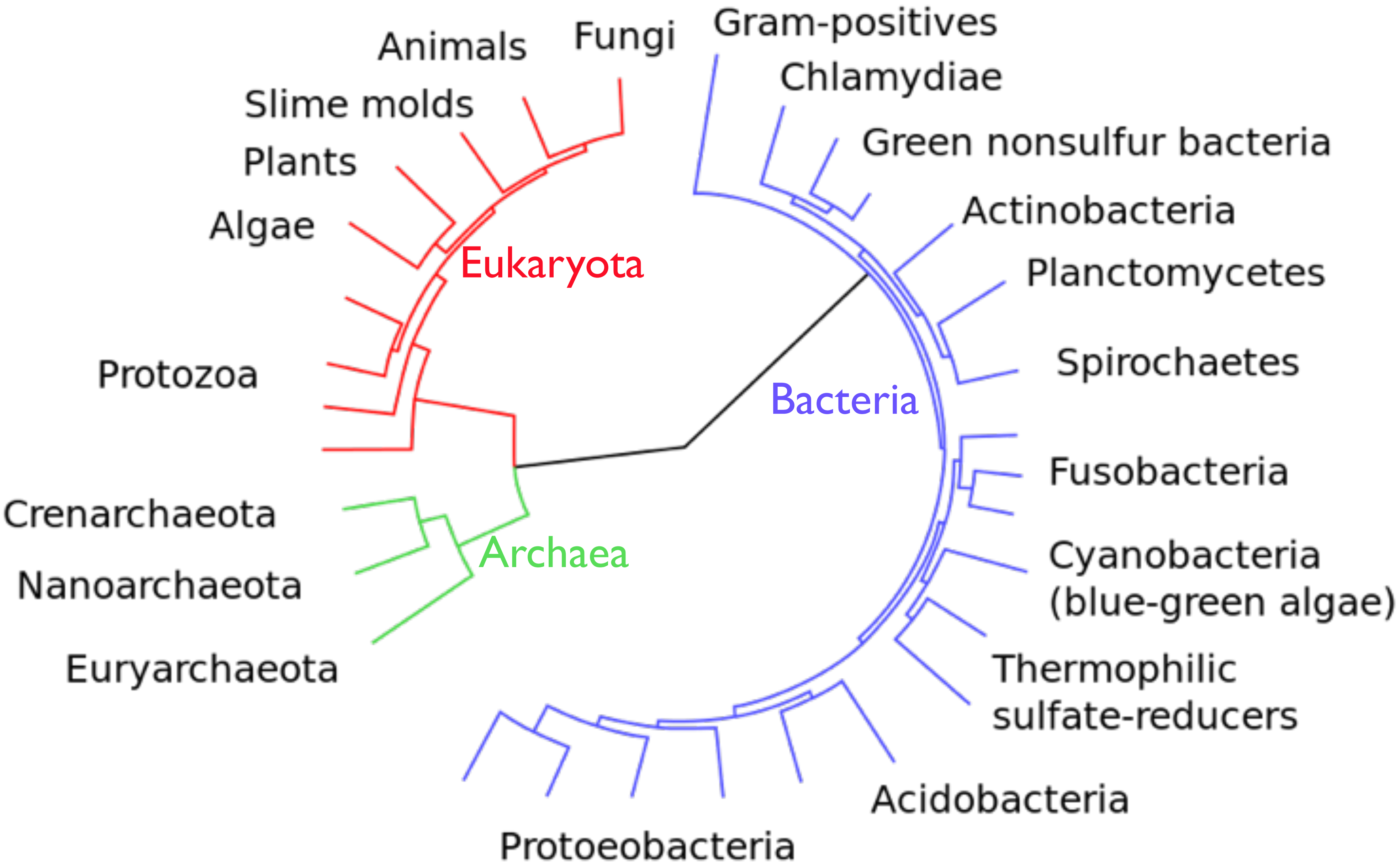
Cell structure



(they don’t all look like this)

# What are microbes?

## Ancestry of Life



http://en.wikipedia.org/wiki/Tree_of_life_(biology)

# What is a microbiome?

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

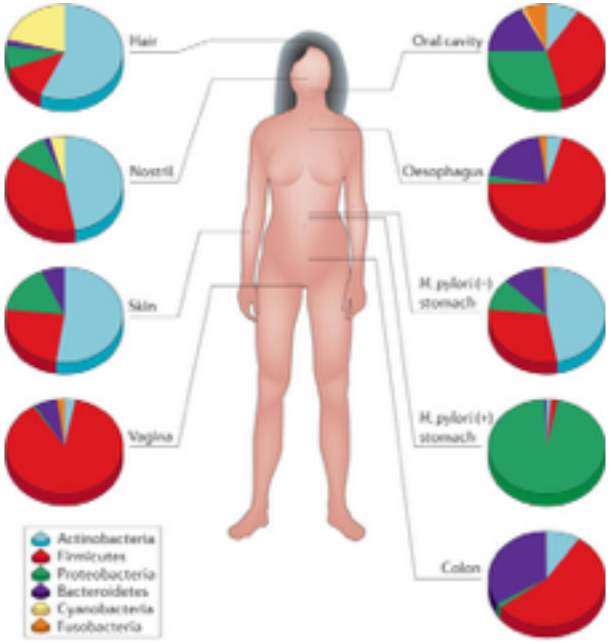
- A population of a single species/strain is a culture, extremely rare outside of lab, some infections
- A microbiome is a mixed population of different microbial species (microbial ecosystem)

# Why study microbiomes?

Wastewater Treatment

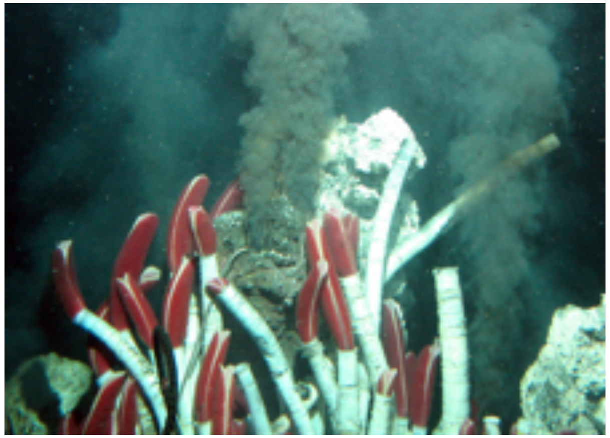
Earth Microbiome Project:

Oceans, soils, waterways

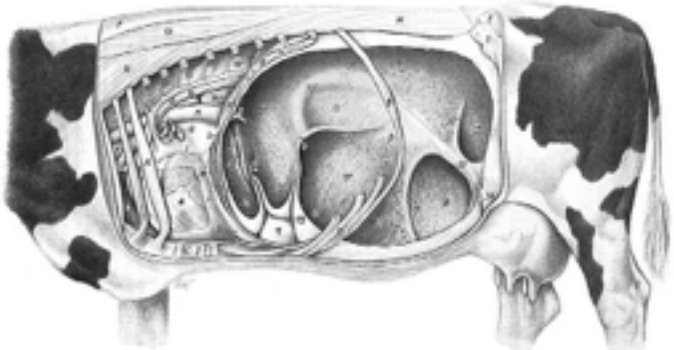


Nature Reviews | Genetics

Human Microbiomes

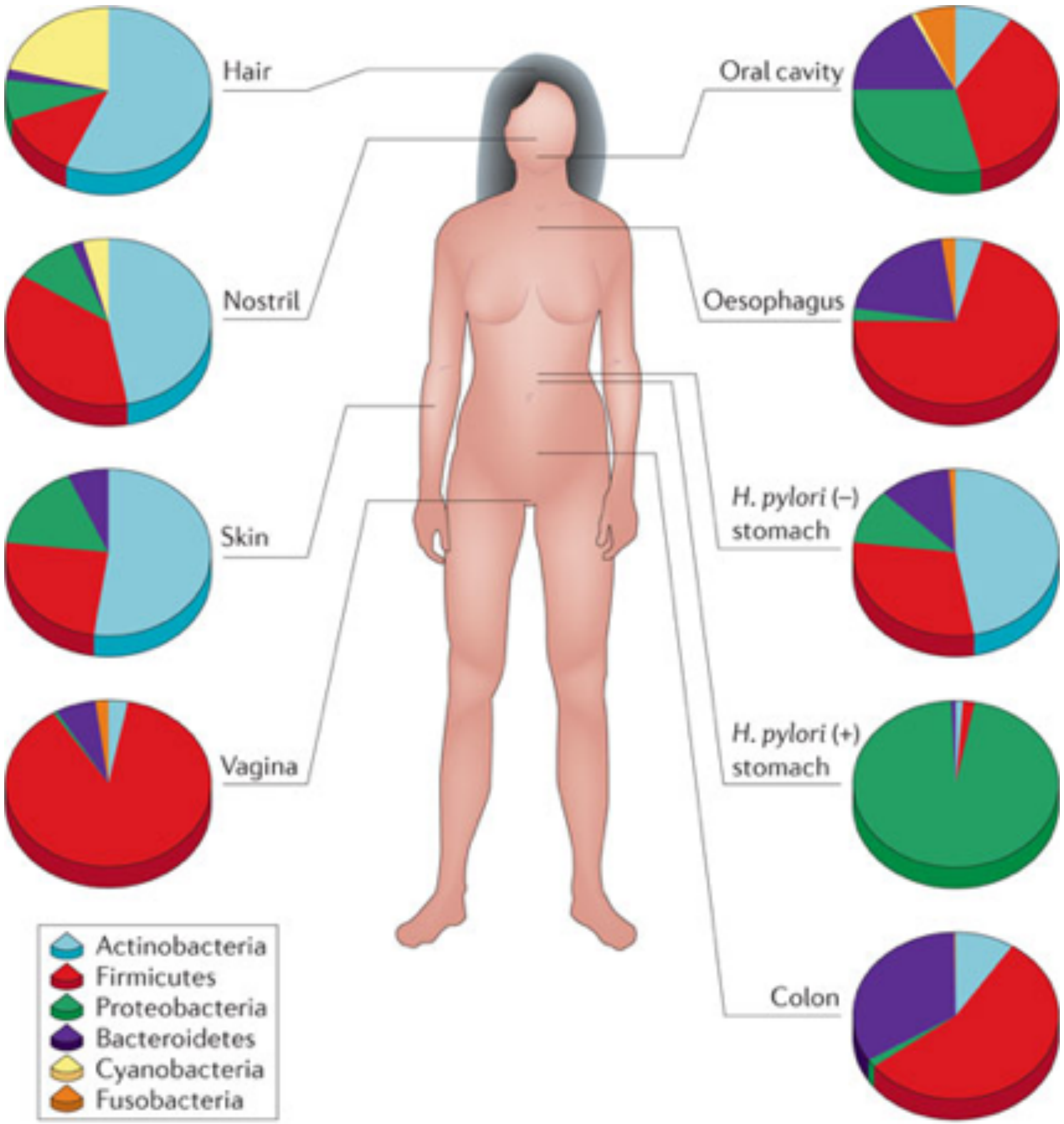


Deep-Sea Hydrothermal Vent



Cow Rumen

# Human Body Sites, HMP



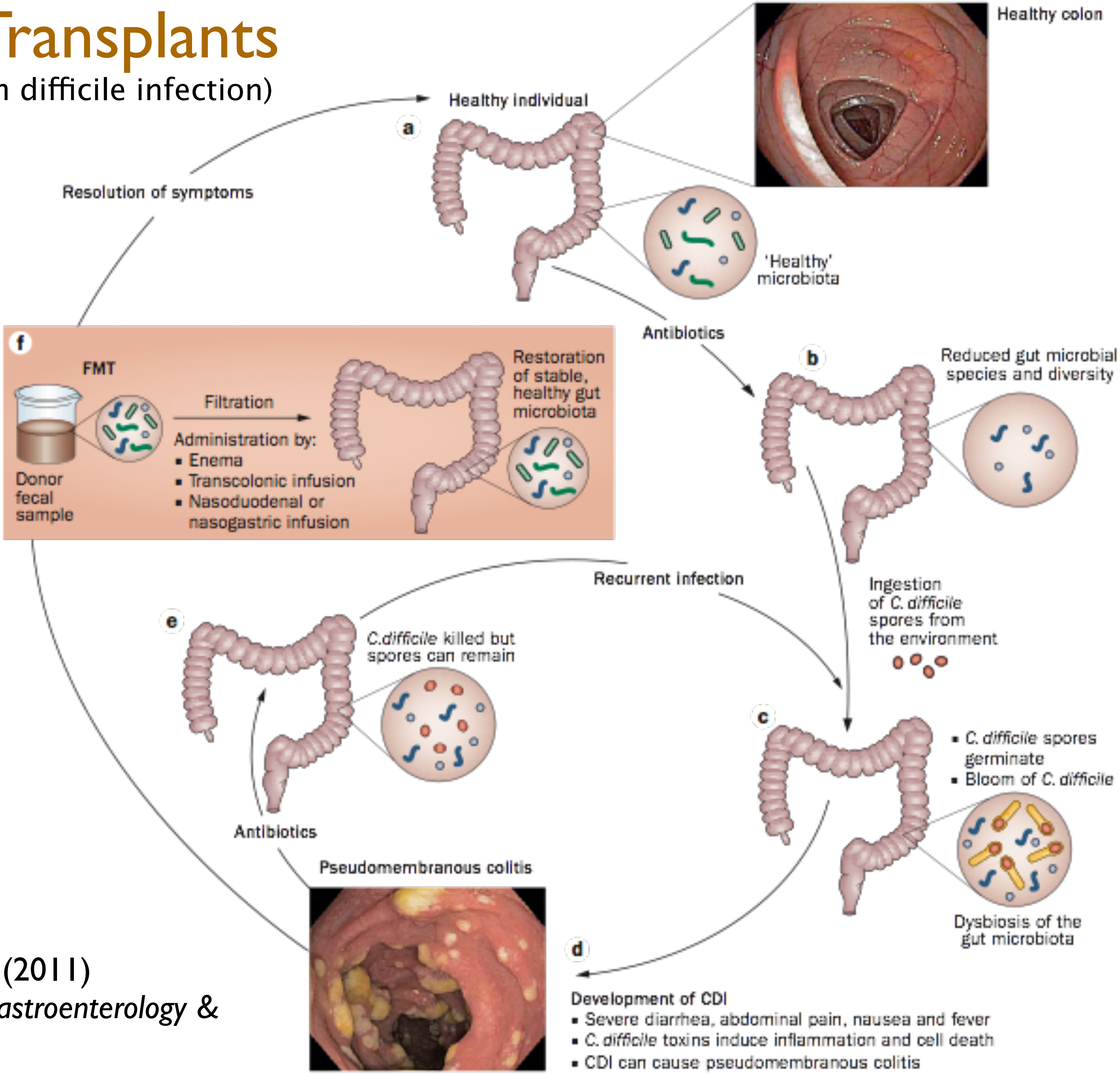
Nature Reviews | Genetics

>10 times more microbial cells than human cells

Entire human microbiome weighs less than 2 kg, at most

# Fecal Transplants

(Clostridium difficile infection)



Borody, et al (2011)
*Nature Rev Gastroenterology & Hepatology*

# Why is microbiome research new?

Bias for cultivable microbes, especially pathogens

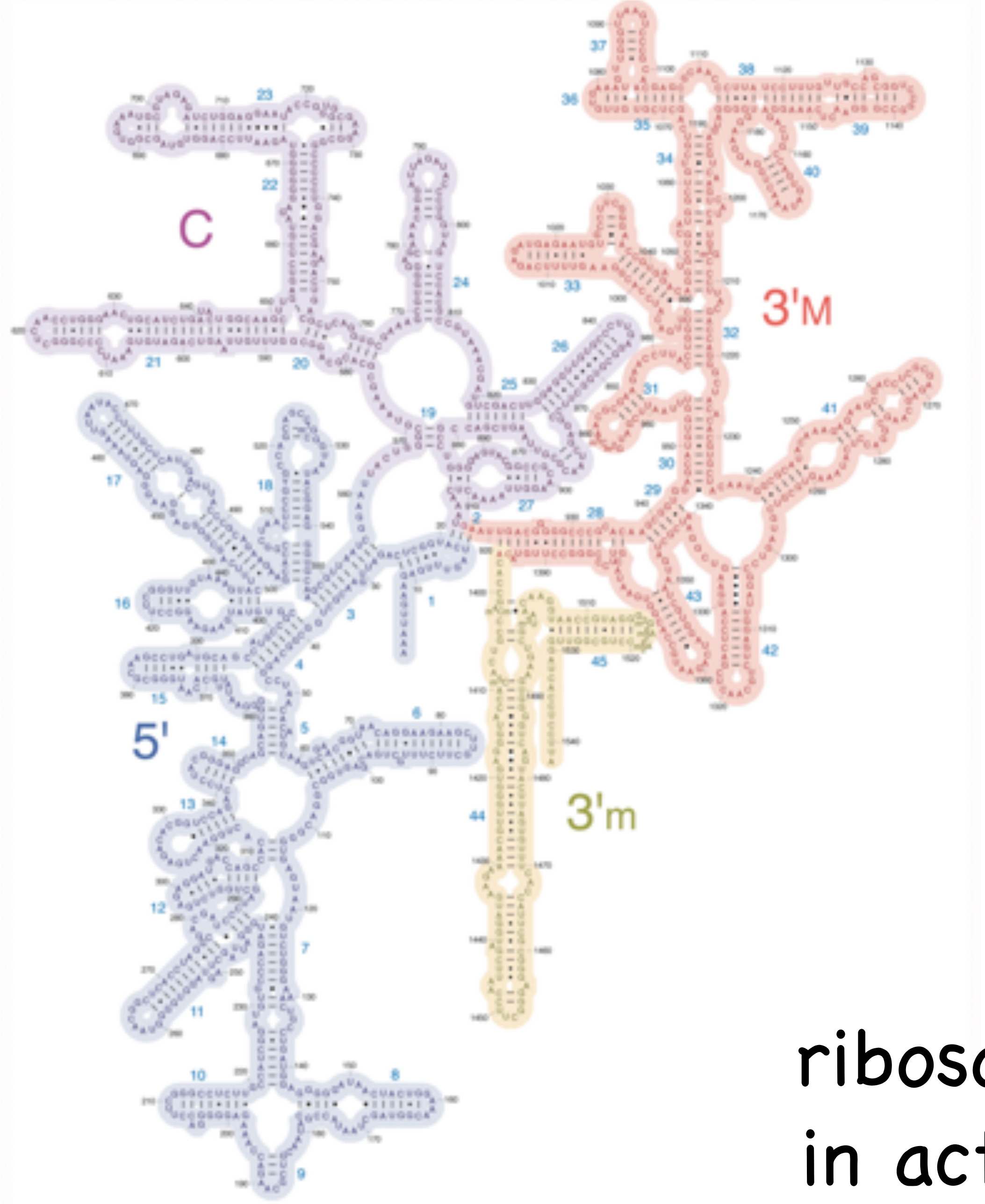
- Culture-based methods fail to detect most microbes
- Microbes are easy to miss (except pathogens)
- Most microbes are NOT pathogens (even the human-associated)

Availability of tools limited to last 3 decades

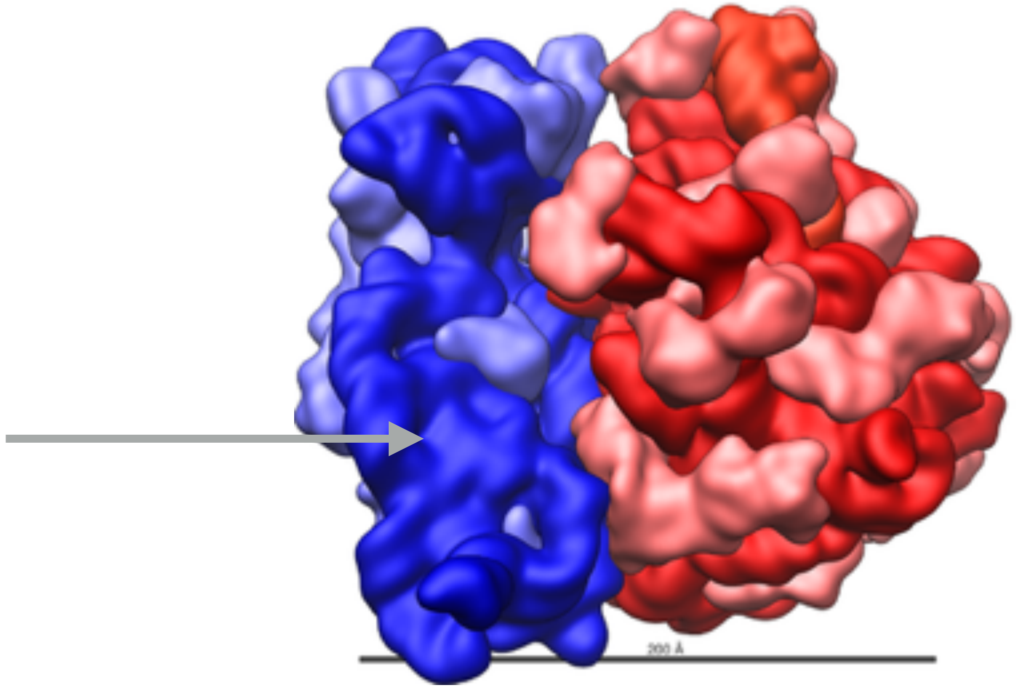
- PCR, fast & cheap DNA sequencing, microarrays, etc
- Discovery of culture-independent techniques - **16S-rRNA**

# How do we query microbiomes??

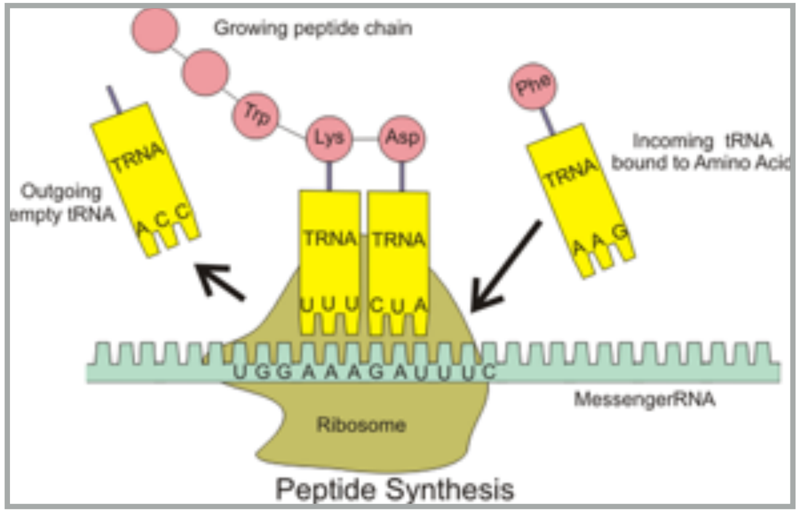
16S rRNA



ribosome

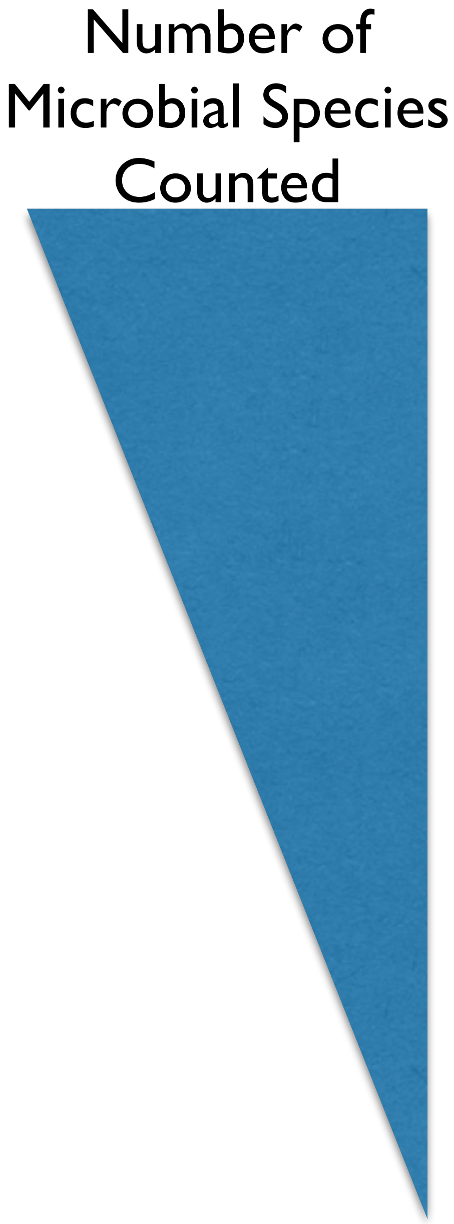


ribosome in action



# How do we query microbiomes??

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



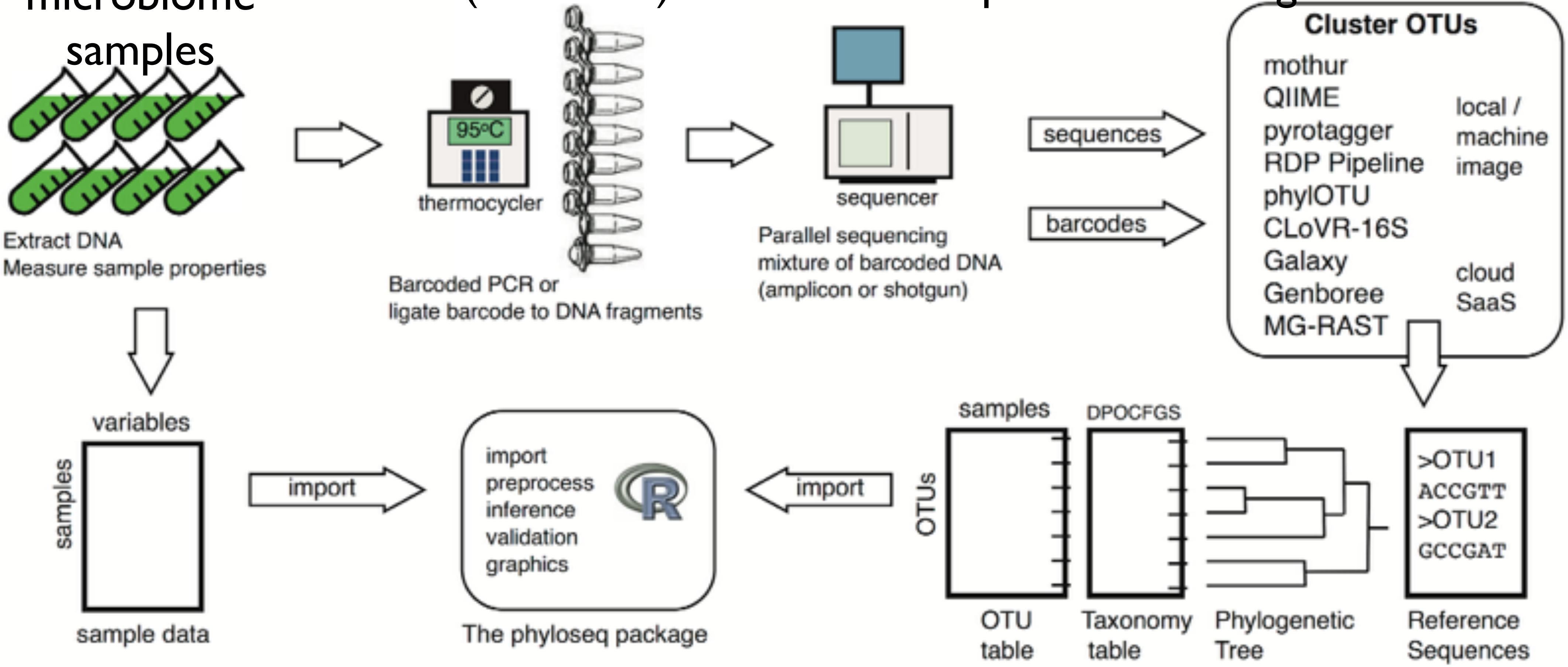
Number of
Microbial Species
Counted

# phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data

microbiome samples

amplify 16S rRNA (barcoded)

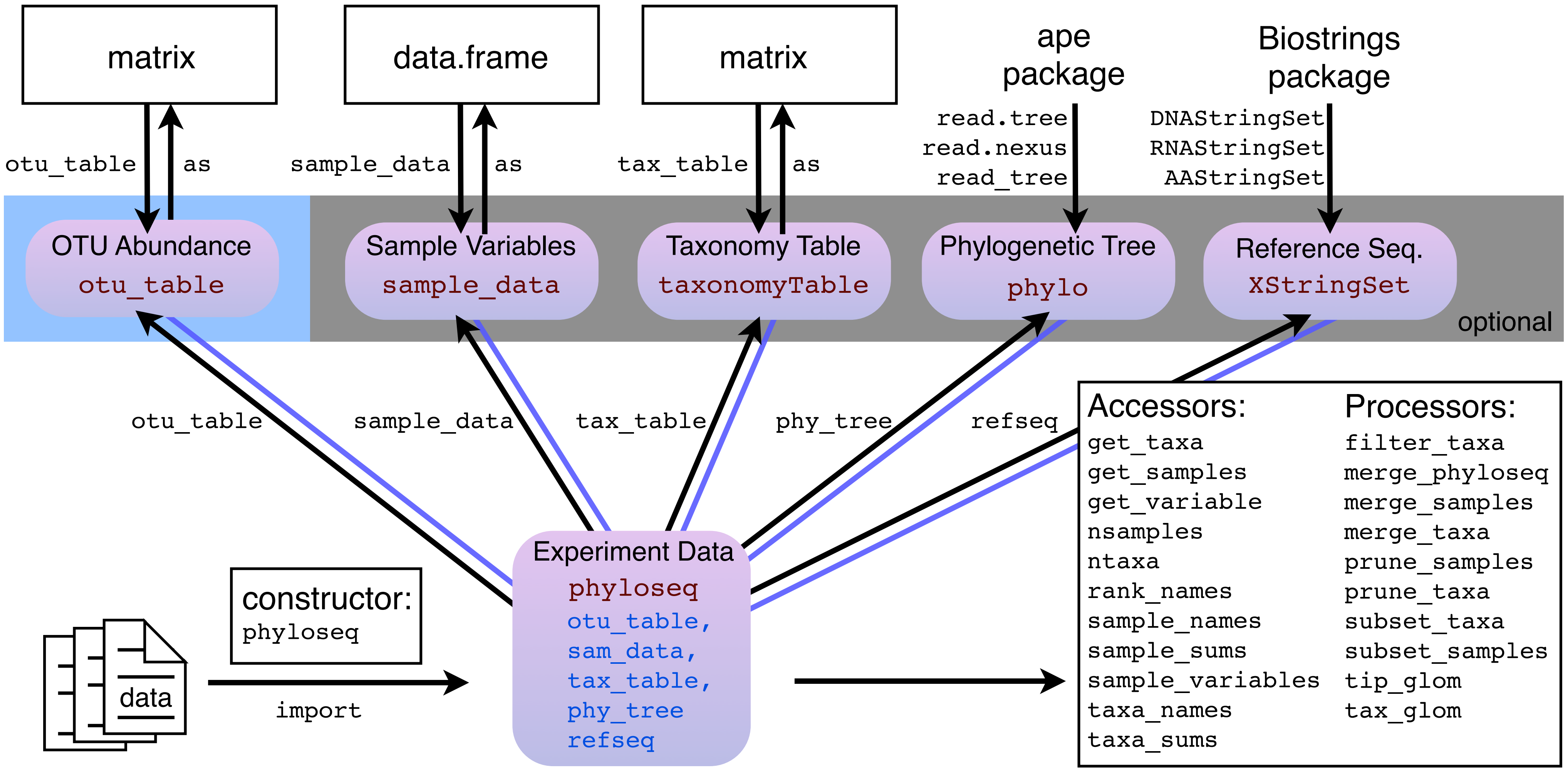
demultiplex and species clustering



Microbiome data
heterogeneity and processing

# phyloseq

## data structure & API



matrix

data.frame

matrix

ape package

Biostrings package

otu_table
as

sample_data
as

tax_table
as

read.tree
read.nexus
read_tree

DNAStringSet
RNAStringSet
AAStringSet

OTU Abundance
`otu_table`

Sample Variables
`sample_data`

Taxonomy Table
`taxonomyTable`

Phylogenetic Tree
`phylo`

Reference Seq.
`XStringSet`

optional

otu_table

sample_data

tax_table

phy_tree

refseq

Experiment Data
`phyloseq`
`otu_table,`
`sam_data,`
`tax_table,`
`phy_tree`
`refseq`

constructor:
`phyloseq`

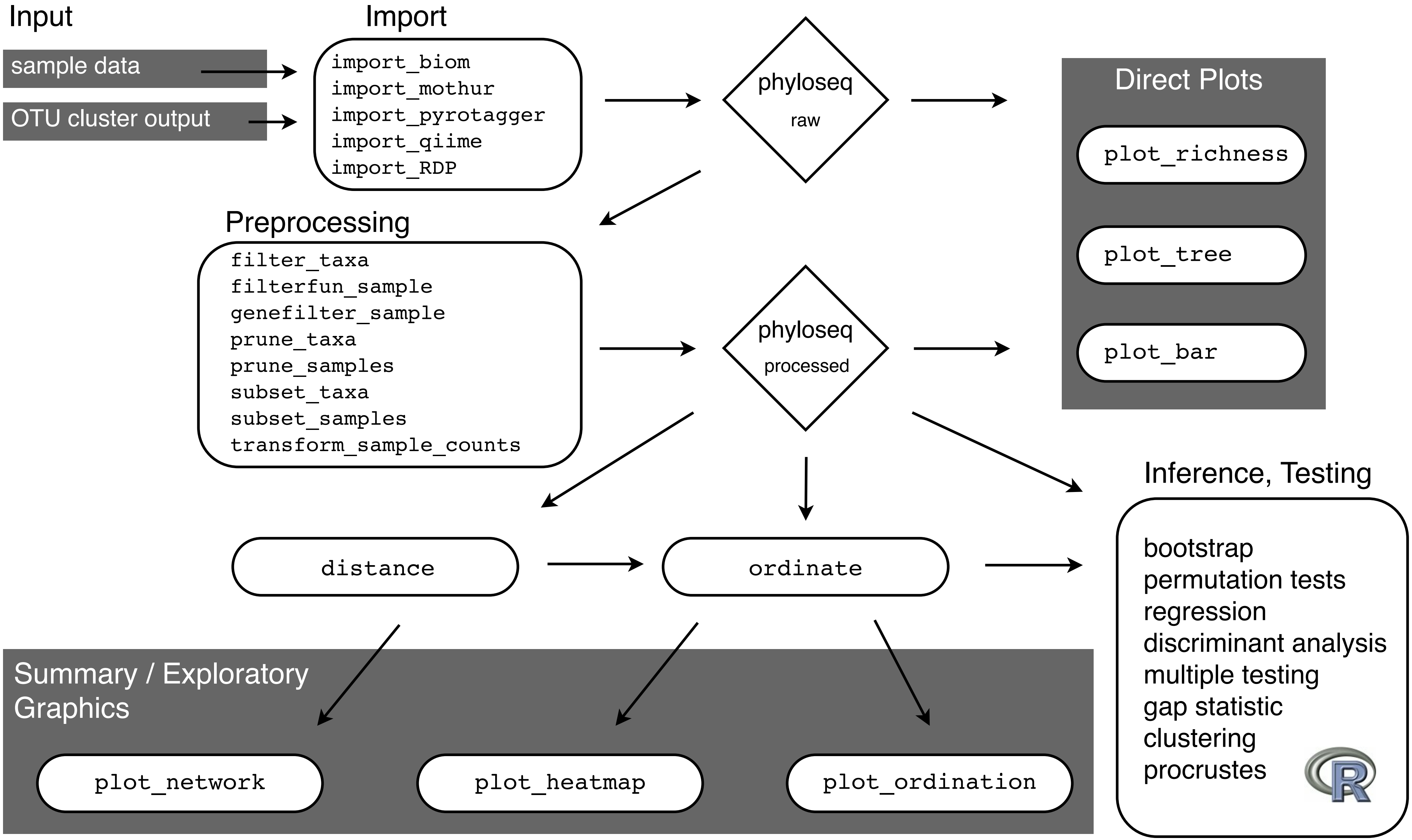
data

import

| Accessors: | Processors: |
|---|---|
| `get_taxa` | `filter_taxa` |
| `get_samples` | `merge_phyloseq` |
| `get_variable` | `merge_samples` |
| `nsamples` | `merge_taxa` |
| `ntaxa` | `prune_samples` |
| `rank_names` | `prune_taxa` |
| `sample_names` | `subset_taxa` |
| `sample_sums` | `subset_samples` |
| `sample_variables` | `tip_glom` |
| `taxa_names` | `tax_glom` |
| `taxa_sums` | |

# phyloseq

work flow



## Input

sample data

OTU cluster output

## Import

`import_biom`
`import_mothur`
`import_pyrotagger`
`import_qiime`
`import_RDP`

phyloseq
raw

## Preprocessing

`filter_taxa`
`filterfun_sample`
`genefilter_sample`
`prune_taxa`
`prune_samples`
`subset_taxa`
`subset_samples`
`transform_sample_counts`

phyloseq
processed

## Direct Plots

`plot_richness`

`plot_tree`

`plot_bar`

`distance`

`ordinate`

## Inference, Testing

bootstrap
permutation tests
regression
discriminant analysis
multiple testing
gap statistic
clustering
procrustes

## Summary / Exploratory Graphics

`plot_network`

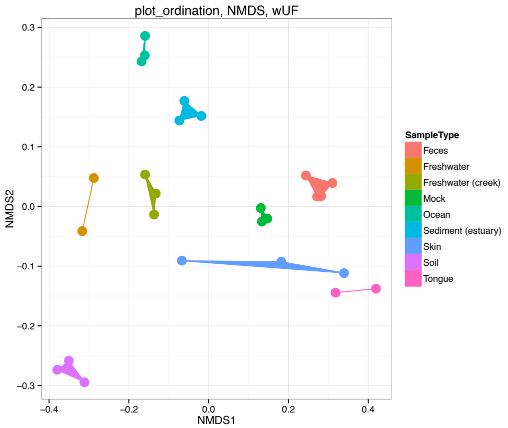
`plot_heatmap`

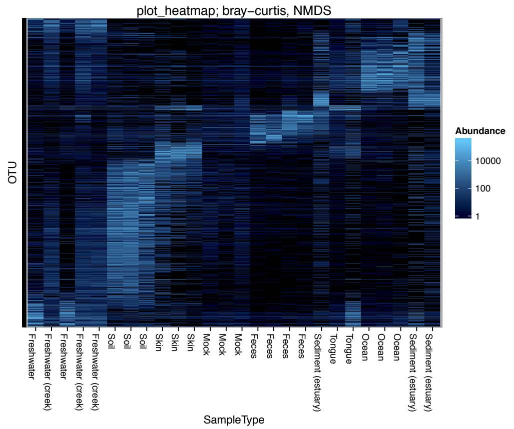
`plot_ordination`

# phyloseq

graphics

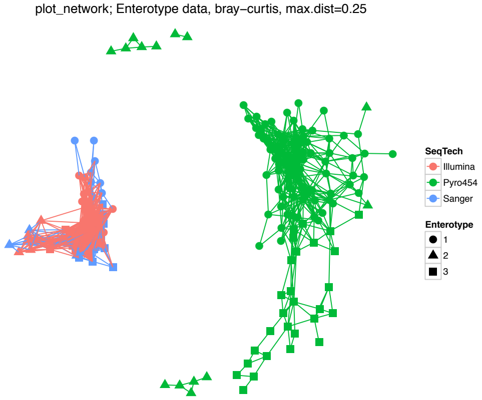
plot_ordination()

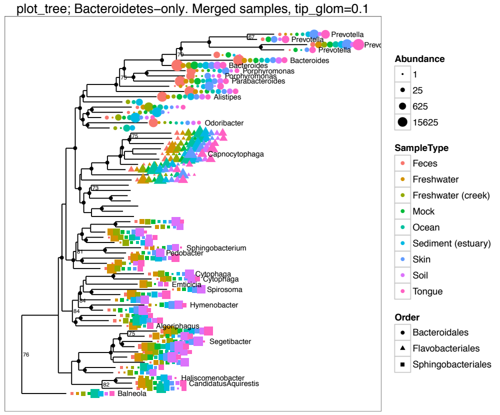




plot_heatmap()

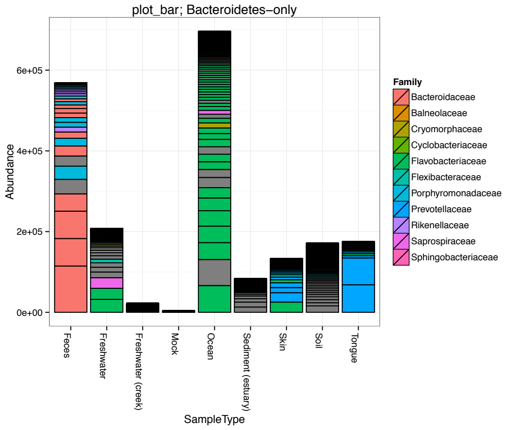
plot_network()

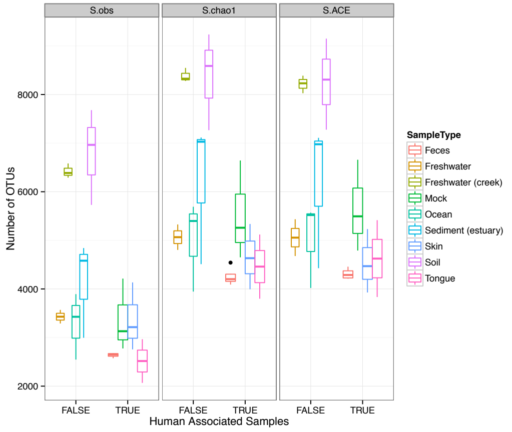




plot_tree()

plot_bar()



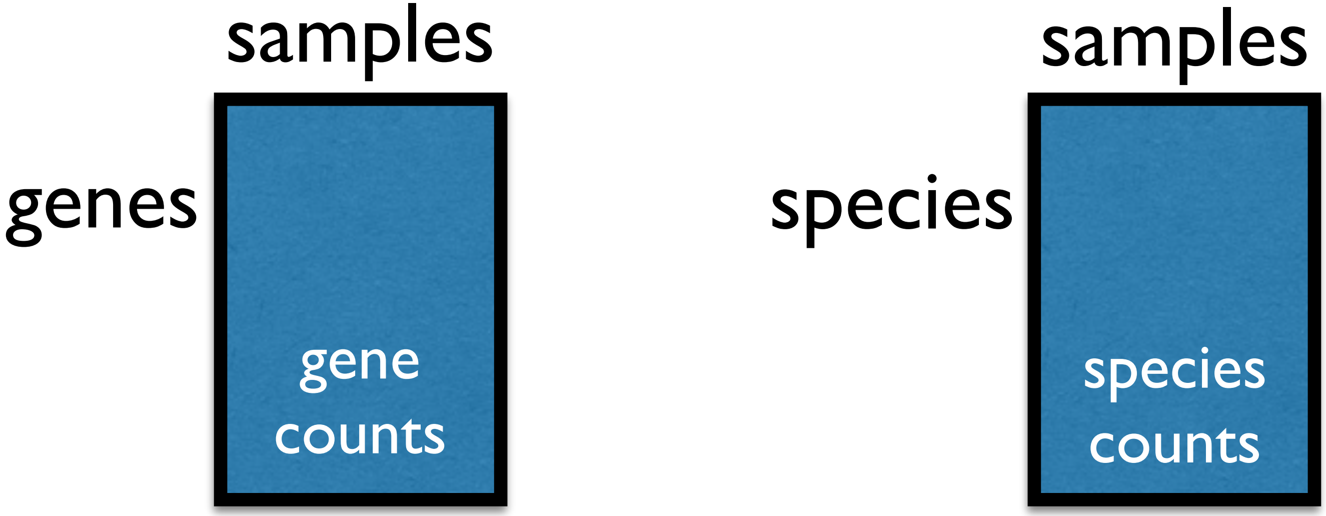


plot_richness()

# Side Note: BioC tools for microbiome

edgeR, 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

| | |
|---|---|
| Susan Holmes | Postdoc Advisor, Mentor, Co-author |
| Holmes Group | Helpful advice and feedback |
| Wolfgang Huber | Helpful advice and feedback re: DESeq(2) |
| BioC and CRAN | Support, Feedback, Distribution of phyloseq and biom |
| RStudio | Shiny, RStudio IDE |
| Hadley Wickham | ggplot2, reshape2, plyr R packages |

# Shiny-phyloseq

# Live Demo

How to Run:

```
install.packages(“shiny”)
shiny::runGitHub(“shiny-phyloseq”, “joey711”)
```

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

# End.
# Questions?