Microbial community analysis in R

CascadiaR Conference

June 3, 2017

Lisa Karstens, PhD

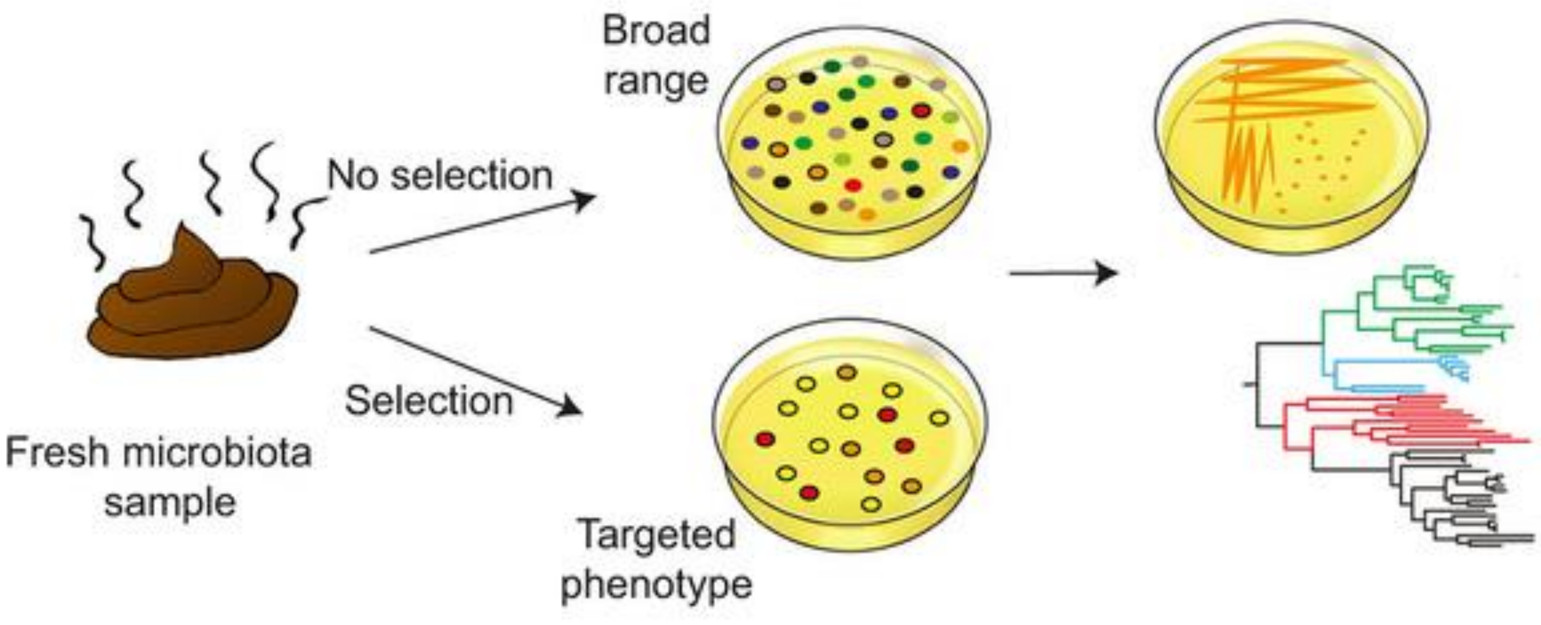
Microbes play an important role in human health

We are not alone…

humans are supraorganisms

- Symbiotic relationship with microbes
- Important for health and disease

Early methods microbial investigation relied on culturing techniques

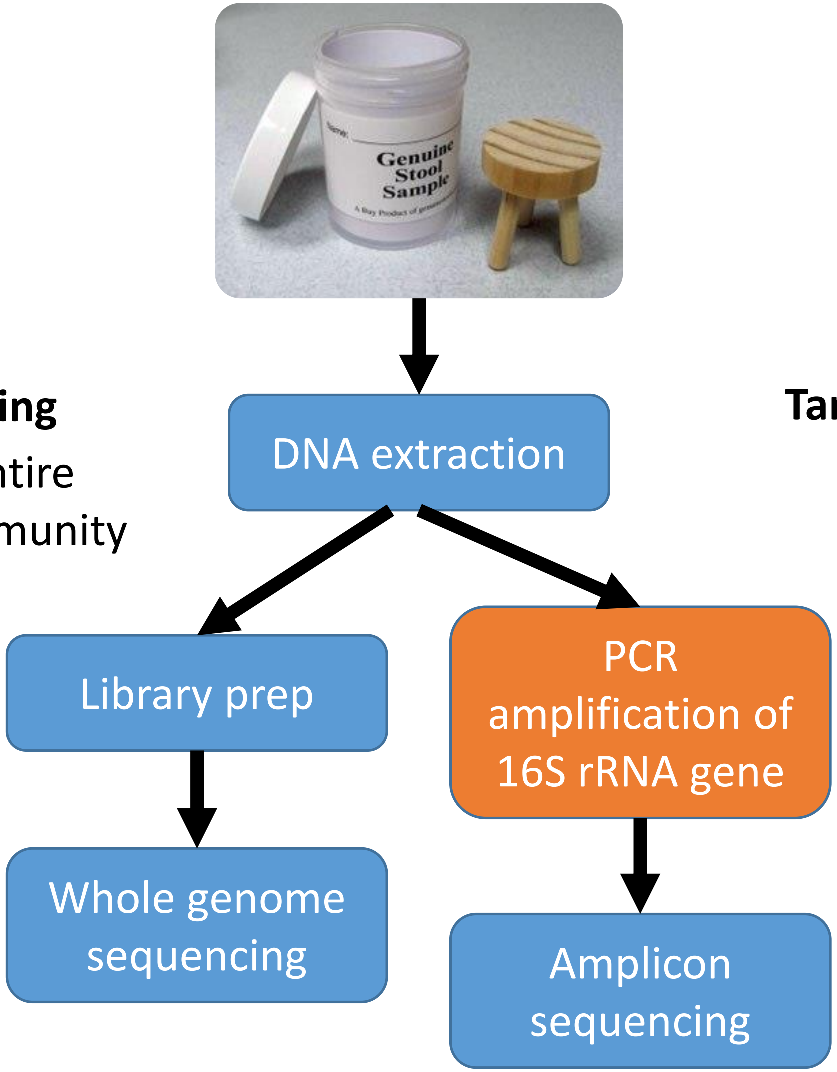


http://www.biotechniques.com/news/Culture-Wars-Identifying-Unculturable-Bacteria-in-the-Human-Gut/biotechniques-364663.html#.WQe0IMa1tPY

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

Metagenome sequencing

Metagenome – entire genome of a community



DNA extraction

Library prep

Whole genome sequencing

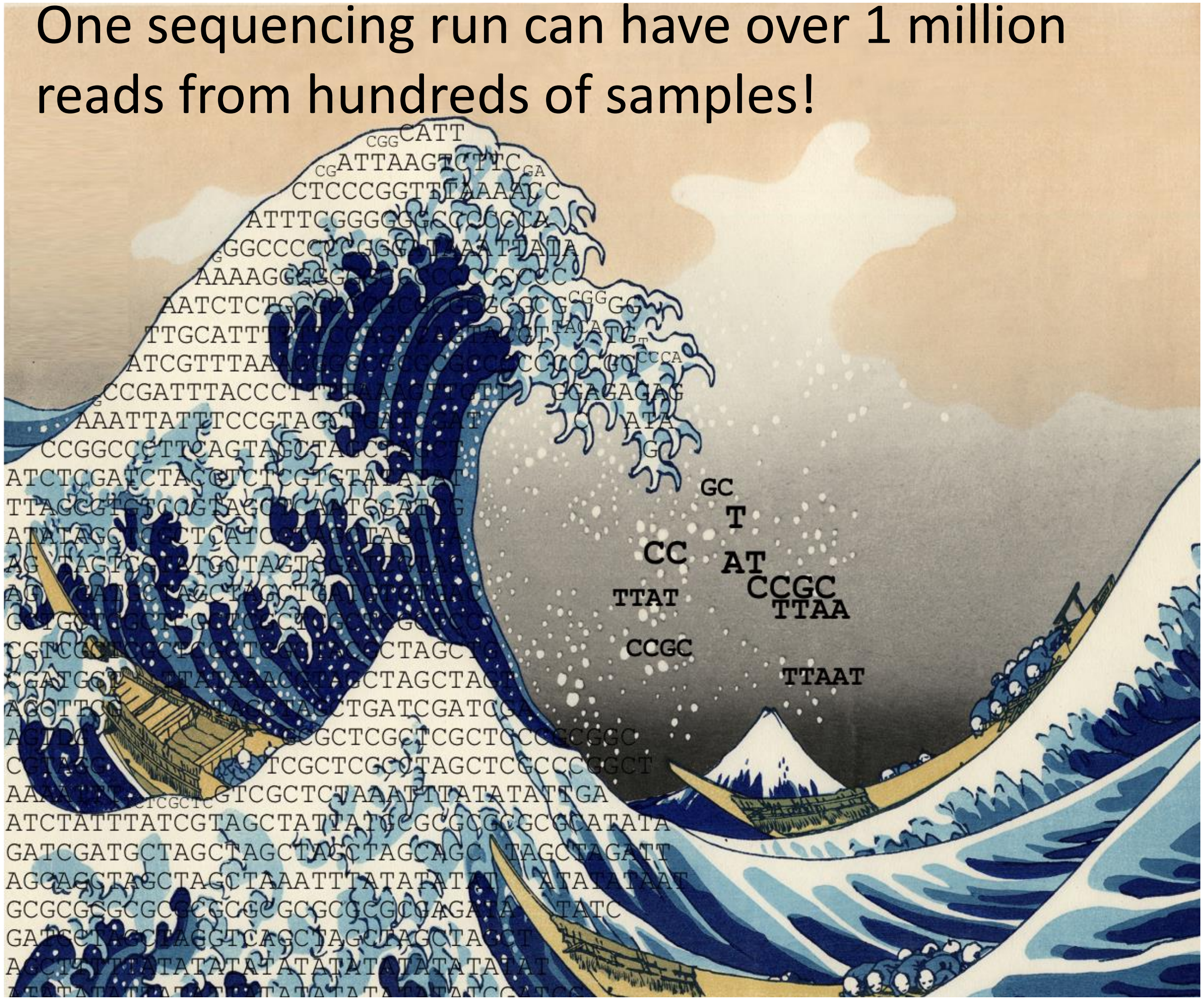
PCR amplification of 16S rRNA gene

Amplicon sequencing

Targeted gene sequencing

16S rRNA gene – marker gene for prokaryotes

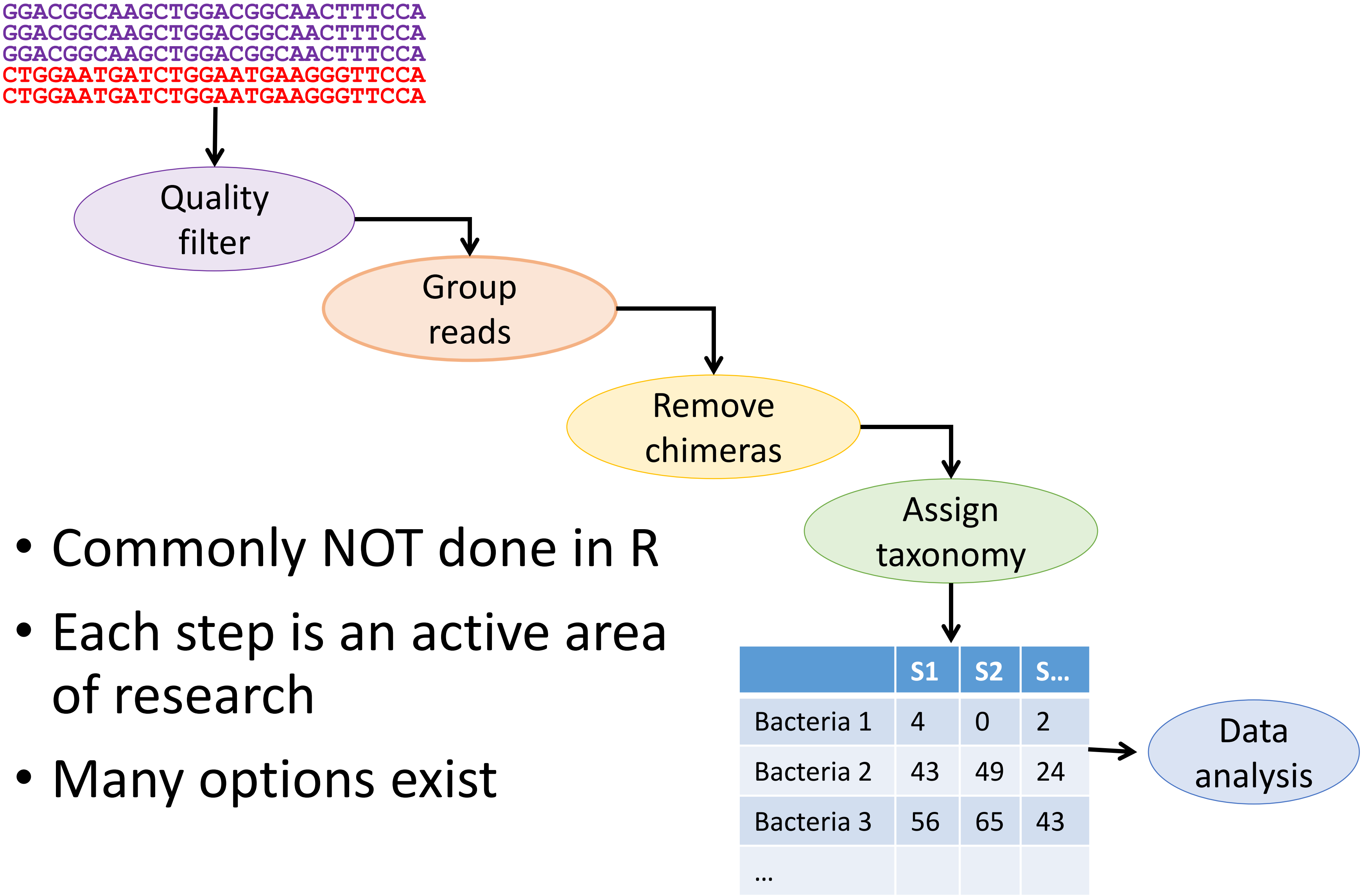
16S sequencing studies generate a lot of data!



One sequencing run can have over 1 million reads from hundreds of samples!

Original image: https://commons.wikimedia.org/wiki/Commons:Valued_image_candidates/The_Great_Wave_off_Kanagawa.jpg, modified by Lisa Karstens

Example workflow processing sequence reads



GGACGGCAAGCTGGACGGCAACTTTCCA
GGACGGCAAGCTGGACGGCAACTTTCCA
GGACGGCAAGCTGGACGGCAACTTTCCA
CTGGAATGATCTGGAATGAAGGGTTCCA
CTGGAATGATCTGGAATGAAGGGTTCCA

	S1	S2	S…
Bacteria 1	4	0	2
Bacteria 2	43	49	24
Bacteria 3	56	65	43
…			

- 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

Data table

	S1	S2	S…
Bacteria 1	4	55	78
Bacteria 2	50	32	32
Bacteria 3	20	2	4
…			

Sample data table

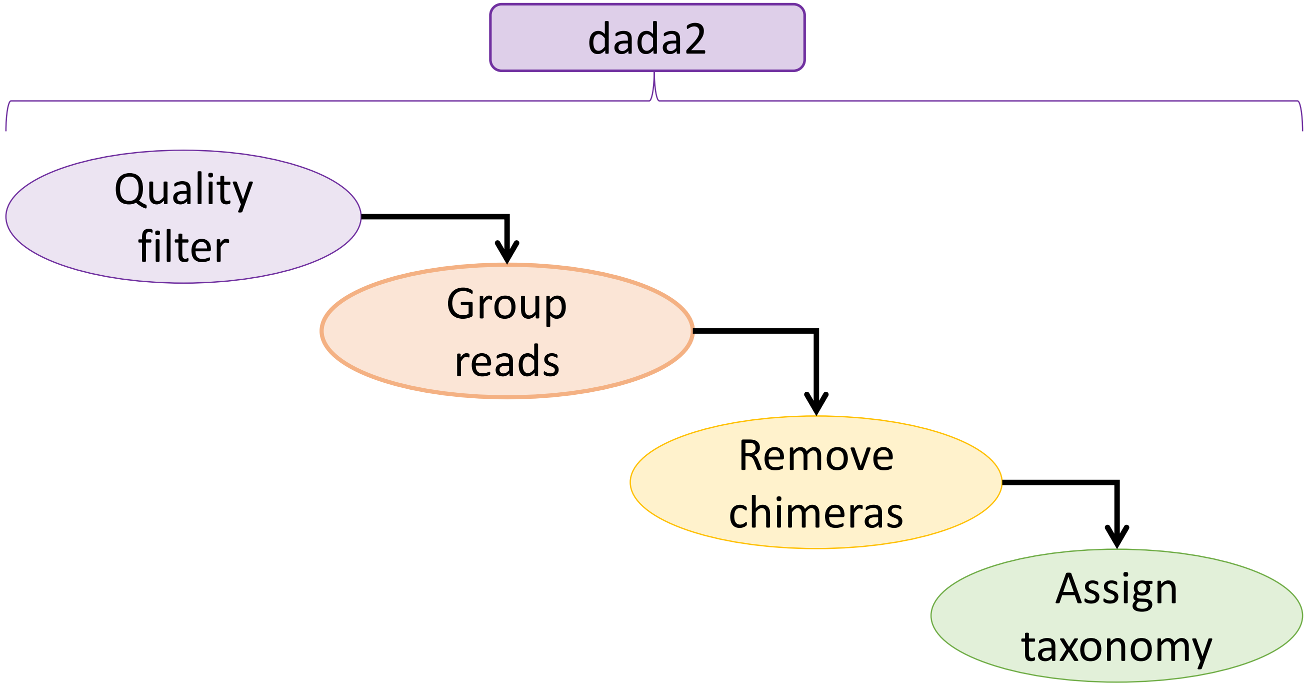
	S1	S2	S…
Diagnosis	D	C	D
BMI	28	25	23
Genotype	0	1	0
…			

Results table

	Change	p
Bacteria 1	2.5	.005
Bacteria 2	5	.05
Bacteria 3	1.3	.96
…		

Microbiome specific workflows in R

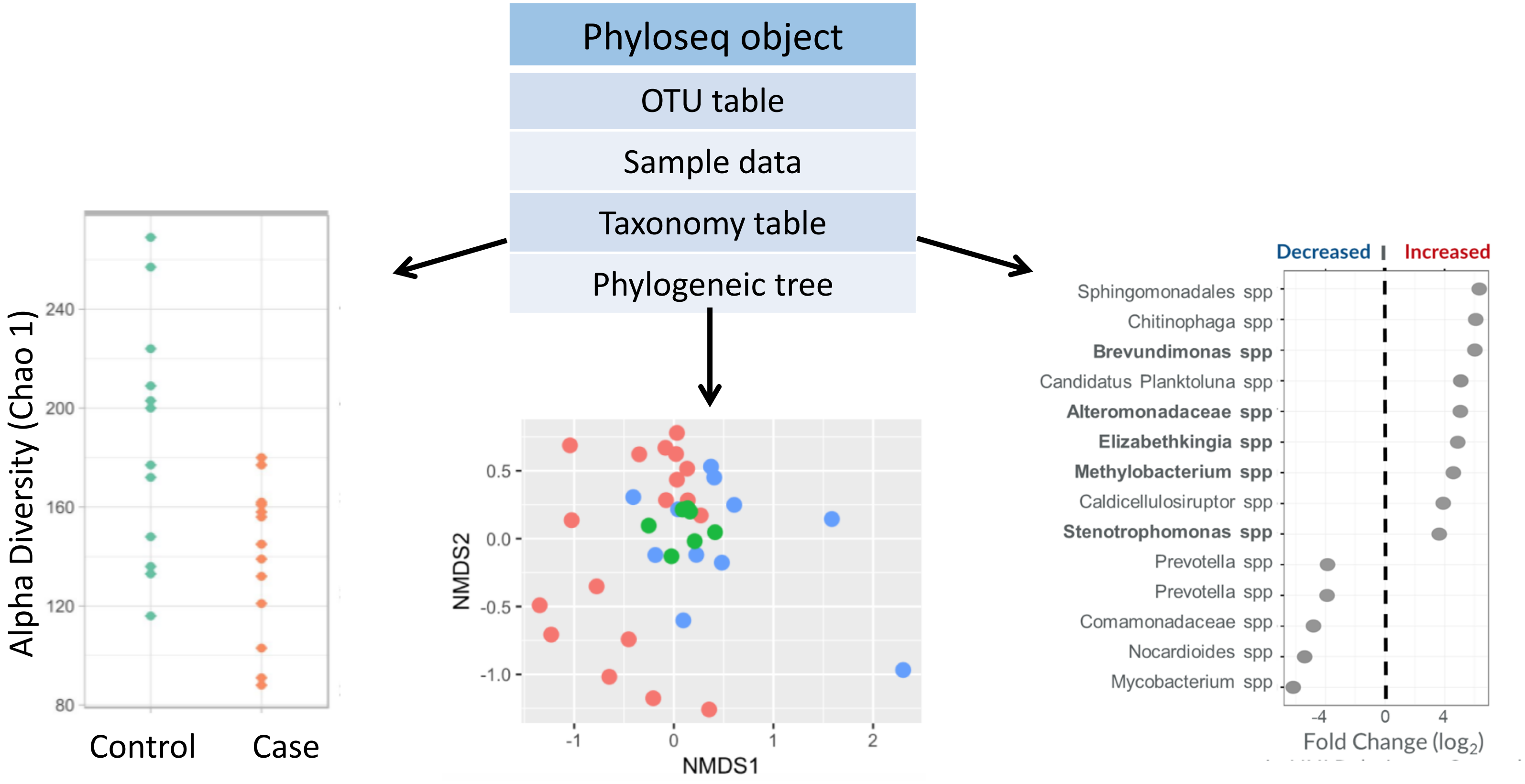
- dada2
  - A package to model and correct 16S sequence errors
  - Replaces clustering algroithms for grouping reads



Callahan BJ, McMurdie PJ, *et al* **DADA2: High-resolution sample inference from amplicon data.** Nature Methods 2016, **13:** 581
Callahan BJ, Sankaran K, *et al.* **Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses**
F1000Research 2016, **5**:1492

Microbiome specific workflows in R

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



Phyloseq object

OTU table

Sample data

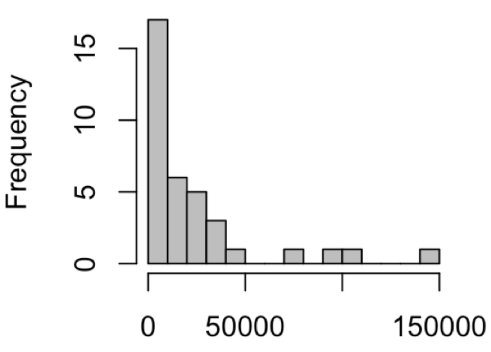
Taxonomy table

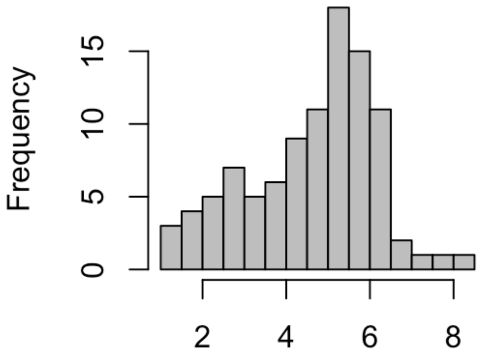
Phylogeneic tree

McMurdie and Holmes (2013) **phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data.** PLoS ONE. 8(4):e61217

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

DeSEQ2: Love MI, et al. (2014). Genome Biology, 15, 550.
MetagenomeSeq: Paulson JN et al. (2013) Nature Methods. 10. 1200

Vegan: Dixon P (2003) Journal of Vegetation Science. 14 (6) 927
Mixomics: Le Cao KA et al. (2016) PLoS ONE 11(8): e0160169

http://www.themeasurementstandard.com/wp-content/uploads/2015/09/one-size-does-not-fit-all.jpg

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

https://www.tibco.com/blog/wp-content/uploads/2012/08/EnterpriseSocialChallenges.jpg

OHSU

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

NIH

Oregon bircWh

Building Interdisciplinary Research Careers in Women's Health

K12 HD043488

Collins Medical Trust