

# Methods in Microbiota Research: Coding Lab

**Ivan Vujkovic-Cvijin**

*(Ee-vahn Vooykoveech Tsveeyeen)*

Assistant Professor

F. Widjaja Inflammatory Bowel Disease Institute

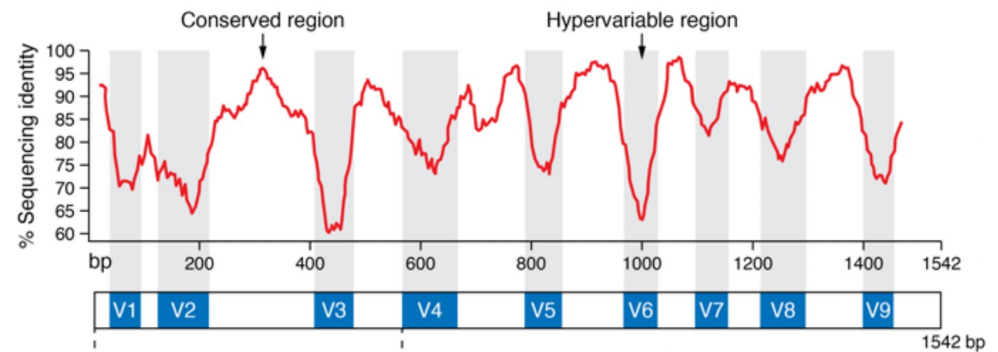
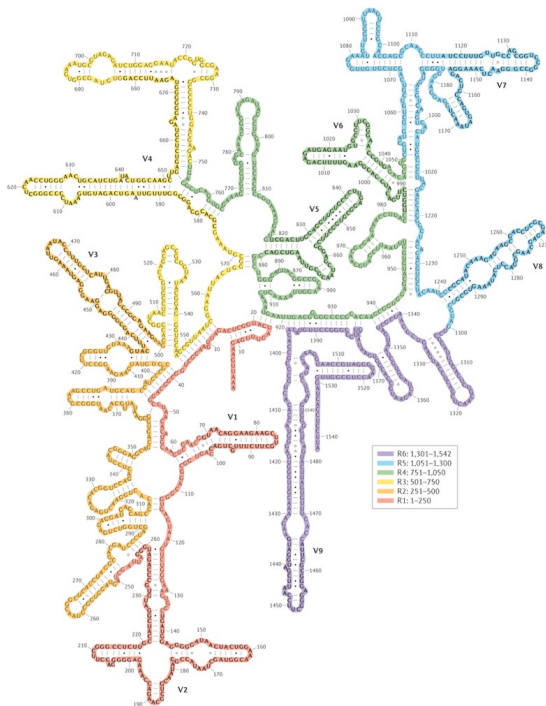
Department of Biomedical Sciences

## Notes prior to beginning

- Google is a coder's best friend
  - Almost every error message has a solution on some online forum
- Statistical modeling/testing in microbiome science is a constantly evolving area
  - Precedent does not always justify usage – methods may be outdated
  - No consensus on best methods – more clarity on what not to do
- I am self-taught
  - If I can do it, so can you!

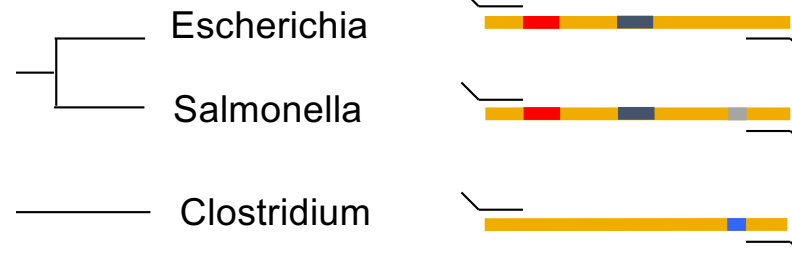
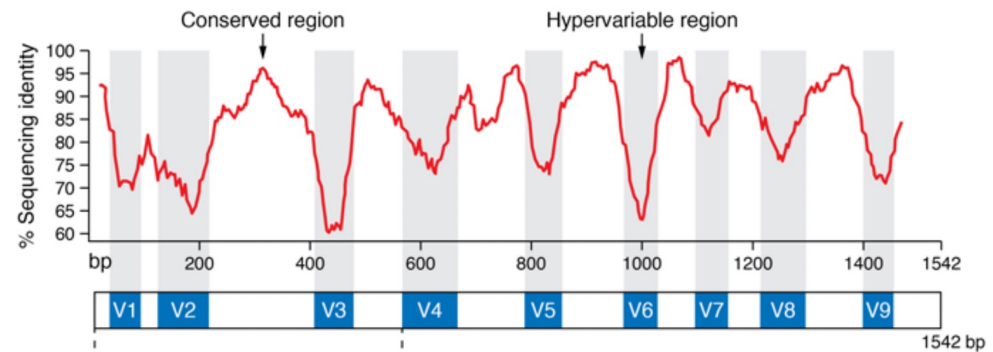
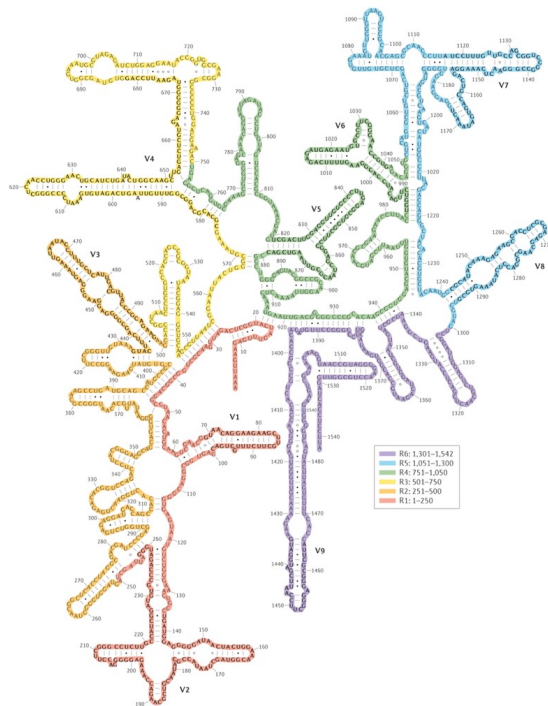
# Tools of the Trade

The bacterial 16S ribosomal RNA gene as a phylogenetic barcode



# Tools of the Trade

## The bacterial 16S ribosomal RNA gene as a phylogenetic barcode



- Amplification and high-throughput sequencing allows quantification of microbiota constituents

# 16S rRNA sequencing analysis: dada2

Read 1 C T A A G A C C G G A T A G G T A

Read 2 C T A A G A C C G G A T A G G T A

Read 3 C T A A G A C C G G **C** T A G G T A

Read 4 C T A A G A C C G G A T A G G T A

Read 5 C T A A G A C C G G A T A G G T A

Illumina 'quality scores'

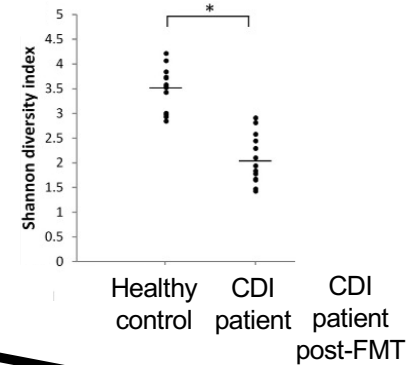
low high



Quality Score Q(X)	Error Probability P(~X)
Q40	0.0001 (1 in 10,000)
Q30	0.001 (1 in 1,000)
Q20	0.01 (1 in 100)
Q10	0.1 (1 in 10)

## Prototypical microbiome-mediated disease: *Clostridioides difficile* infection

- *C. difficile* infection is associated with a loss of diversity of the endogenous microbiota
- Is generally treated with antibiotics targeting *C. difficile* (e.g. vancomycin)
- Antibiotic treatment often does not work, condition becomes chronic (~35%)
- Restoration of the endogenous microbiota via fecal microbial transplantation (FMT) cures up to 94% of recurrent patients (van Nood, NEJM, 2013)



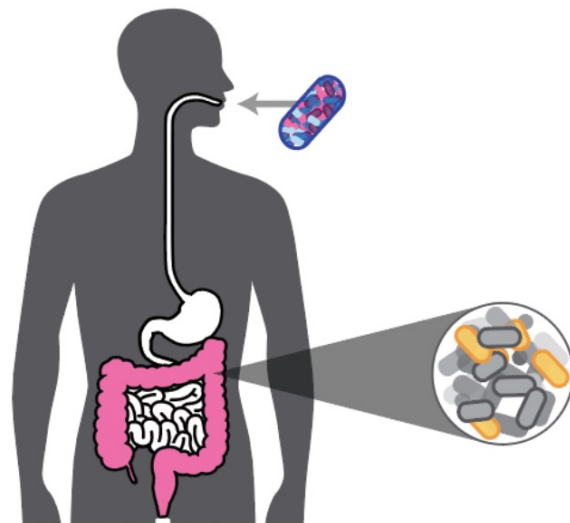
Weingarden et al. Am J Phys 2015

Donor fecal microbiome

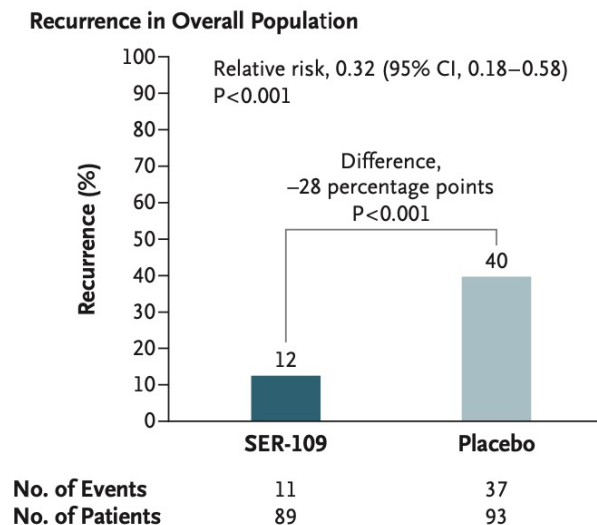


**Question:** How can we identify gut taxa critical for recurrent CDI cure?

# Precision microbiota restoration prevents recurrent CDI



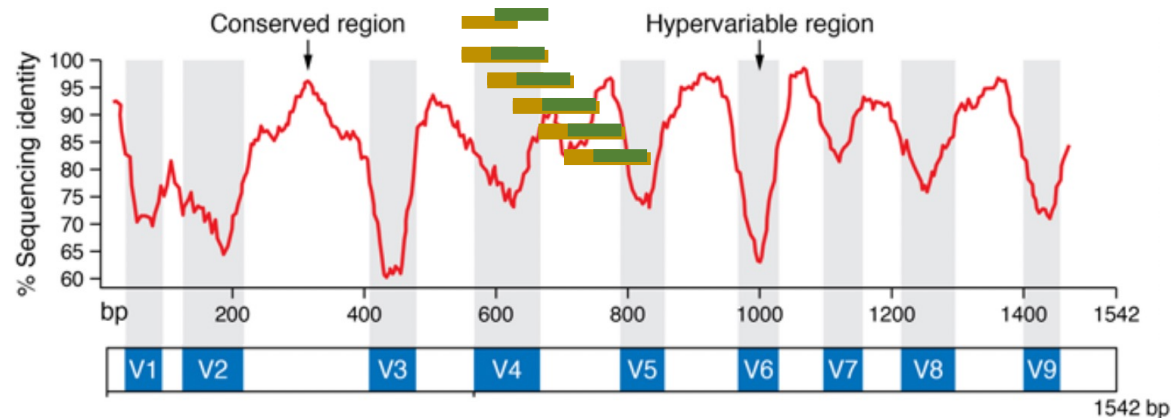
3 Seres Therapeutics, Inc. © 2021  
Henn et al Gastroenterology, 2021



Feuerstadt et al. NEJM. 2022

- Seres Therapeutics identified gut taxa that could help prevent CDI, bottled them up, and made a therapeutic
- **Question:** How can we identify these key gut taxa? Let's get coding...

# Pre-processing: Chimera filtering

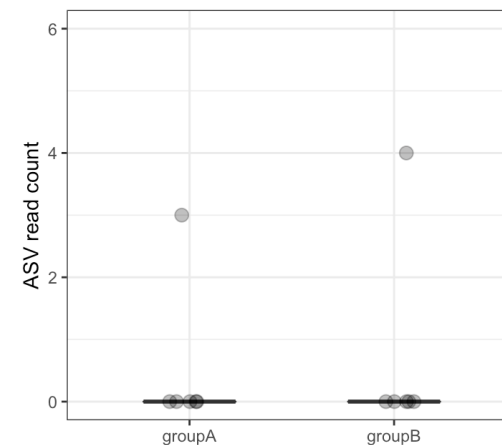
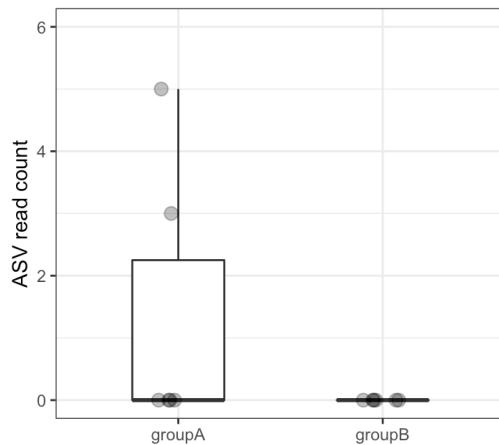


- Chimeras are identifiable if one portion is identical to a more abundant 'parent' sequence and the rest is identical to another 'parent' sequence
  - 'removeBimeraDenovo' function in dada2



Pre-processing:  
removing taxa with low read counts

Mean fold  
difference  
A/B  
= infinity



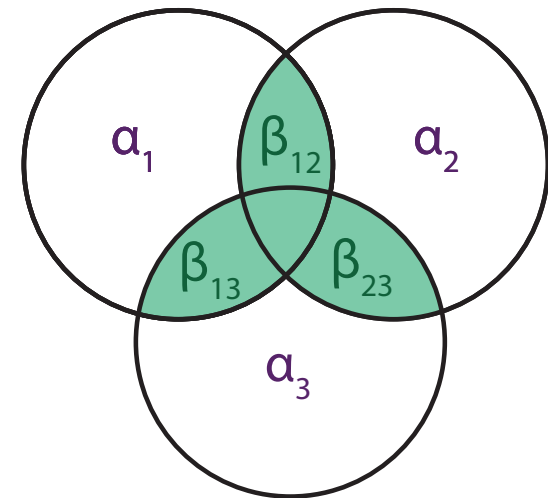
Mean fold  
difference  
A/B  
= 0.75

- Random re-sampling (e.g. multiple rarefactions or re-sequencing) of samples causes wide variation in abundance of taxa with low read counts
- Fold changes between groups can be very unstable and unreliable
  - Example solution: filter taxa that are below 0.001% abundance threshold

# Comparative metrics: alpha vs. beta diversity

## Alpha diversity:

- The number of different taxa present in a sample (richness) and/or the relative distribution of these taxa (evenness)
  - Examples: chao1 (richness), Shannon or inverse Simpson (richness and evenness), Pielou's (evenness)
- Each sample gets its own alpha diversity quantification



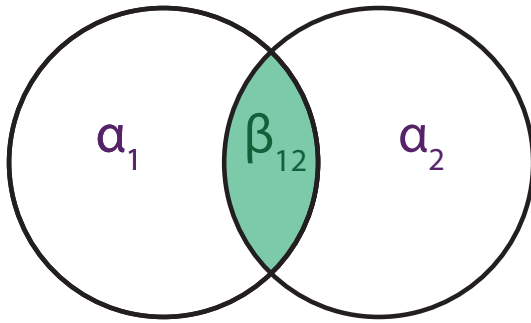
## Beta diversity:

- The ecological similarity *between two samples*
  - Similarity in common taxon membership (ex. Canberra, Bray-Curtis), phylogenetic similarity of taxa within a community (ex. Unweighted UniFrac)
- Each *pair of samples* gets its own beta diversity quantification

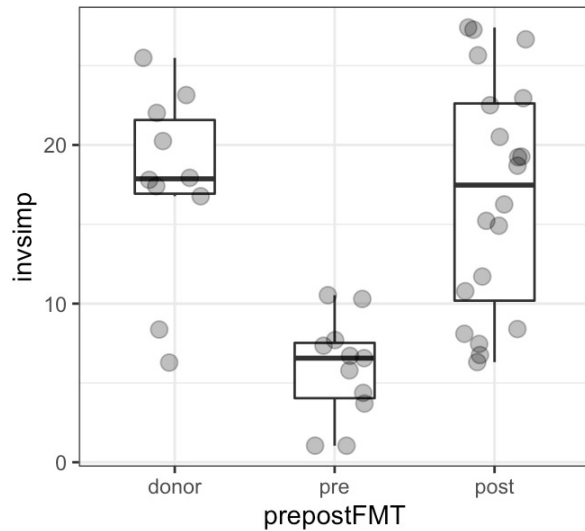
Beta diversity matrix

Subject ID	1	2	3	...
1	0	0.22	0.9	
2	0.22	0	0.47	
3	0.9	0.47	0	
...				

# Comparative metrics: alpha vs. beta diversity



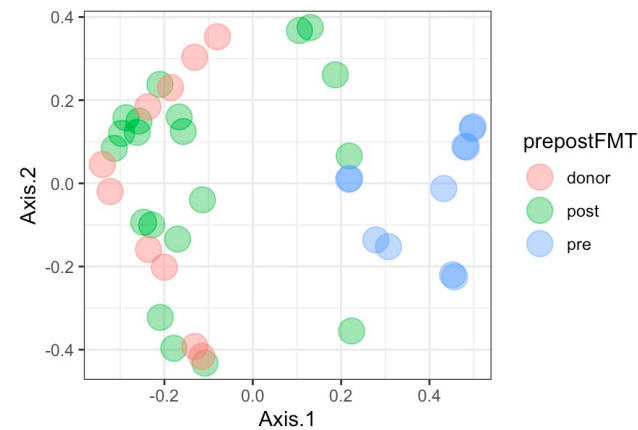
Inverse Simpson  
(Alpha diversity)



Beta diversity matrix

Subject ID	1	2	3	...
1	0	0.22	0.9	
2	0.22	0	0.47	
3	0.9	0.47	0	
...				

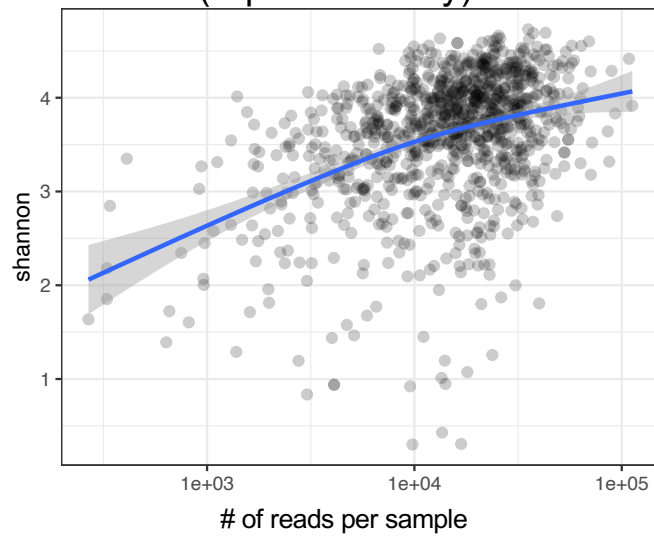
Principal Coordinates Analysis (PCoA)  
(Based on beta diversity matrix)



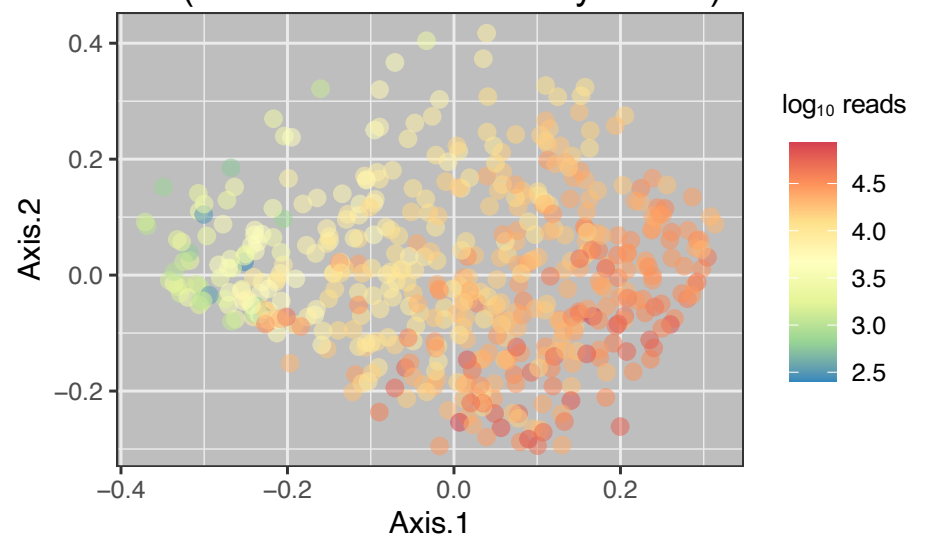
Pre vs. post  
PERMANOVA  
P = 0.00001

# Alpha and beta diversity are skewed by sequencing read depth

Shannon diversity  
(Alpha diversity)



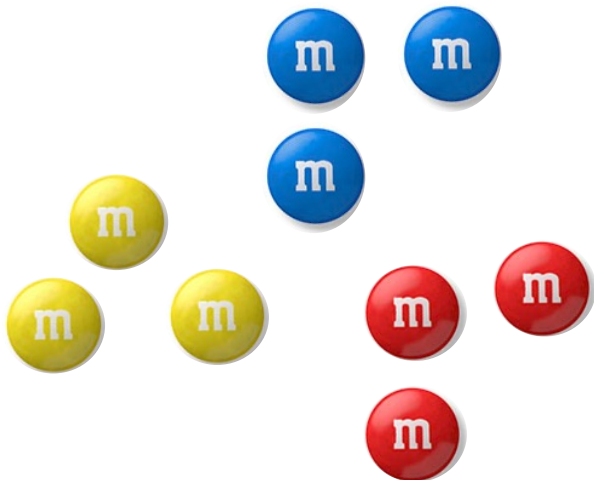
Principal Coordinates Analysis (PCoA)  
(Based on beta diversity matrix)



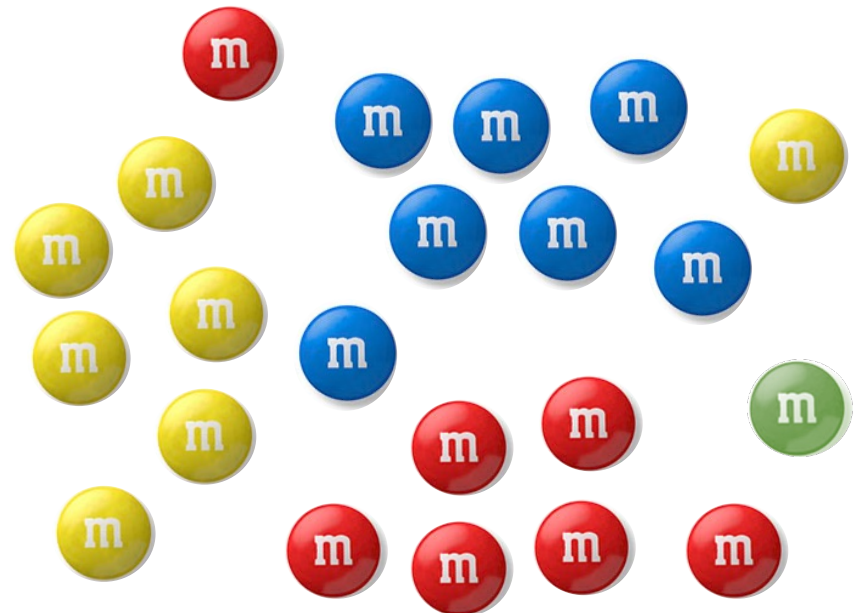
# Rarefaction: random selection of 'n' reads per sample

Randomly sub-sample 9 candies from each handful  
(rarefy to 9 candies)

Handful #1



Handful #2



Let's get coding