Introduction to metagenetics analysis with QIIME2

Websites
NEOF: https://neof.org.uk/
NERC: https://nerc.ukri.org/
CGR:
https://www.liverpool.ac.uk/genomic-research/

Twitter
NEOF: @NERC_EOF
NERC: @NERCscience
CGR: @CGR_UoL

Upcoming workshops

https://neof.org.uk/training/

- Metabarcoding for diet analysis and environmental DNA
  - 28th February & 2nd March 2023
- Microbial shotgun metagenomics
  - 21st & 23rd March 2023
- Eukaryote genome assembly
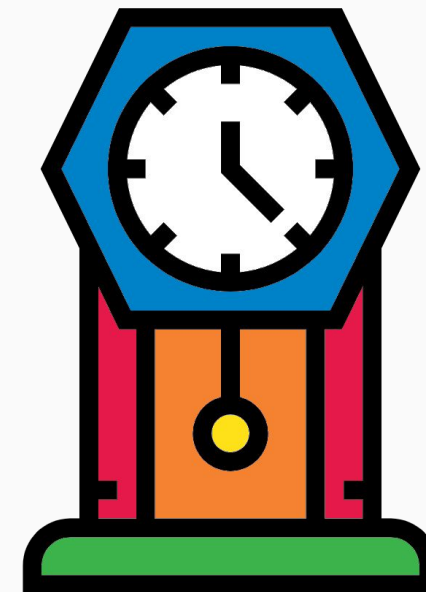 - 18th & 20th April 2023
- More!

Format & Schedule

This intro

Bookdown
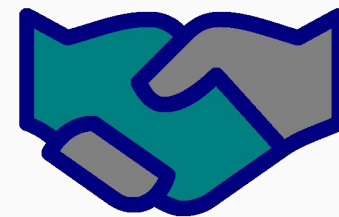
Theory

Practice

Exercises

MCQs

Optional materials

Work at your own pace on your own time

Introduction

- Why 16S rRNA?
- What is QIIME2?
- QIIME2 Workflow
- Quality Control prior to QIIME2 analysis
- DADA2
- Sequence table and taxonomy classification
- Biodiversity
  - Alpha, Beta and Gamma
- Biomarker detection

- 16S ribosomal RNA
- Prokaryotic ribosome
- Phylogeny reconstruction
- Slow rate of evolution
- Functional constancy
- ~1,500 bp long
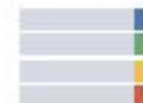- 9 variable regions
- Flanked by conserved regions
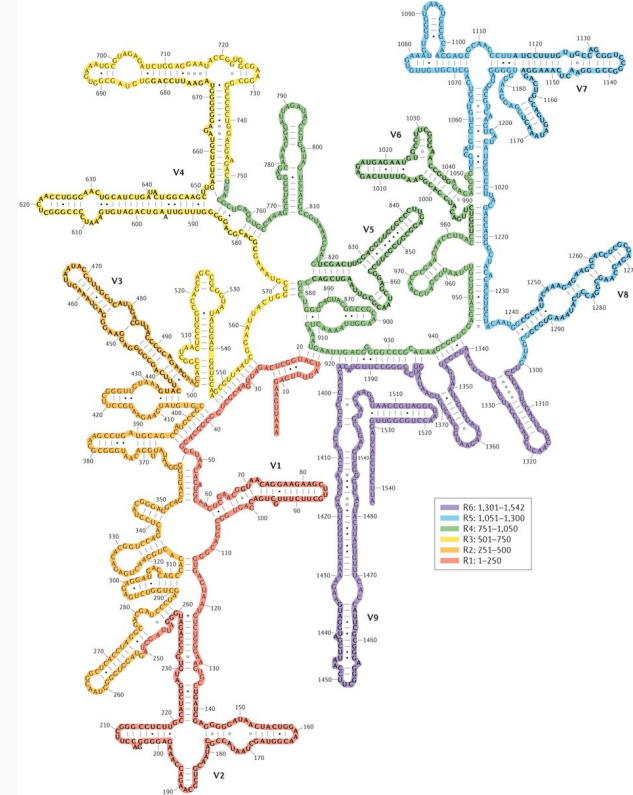
Nature Reviews | Microbiology

Yarza, Pablo, et al. "Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences." *Nature Reviews Microbiology* 12.9 (2014): 635-645.
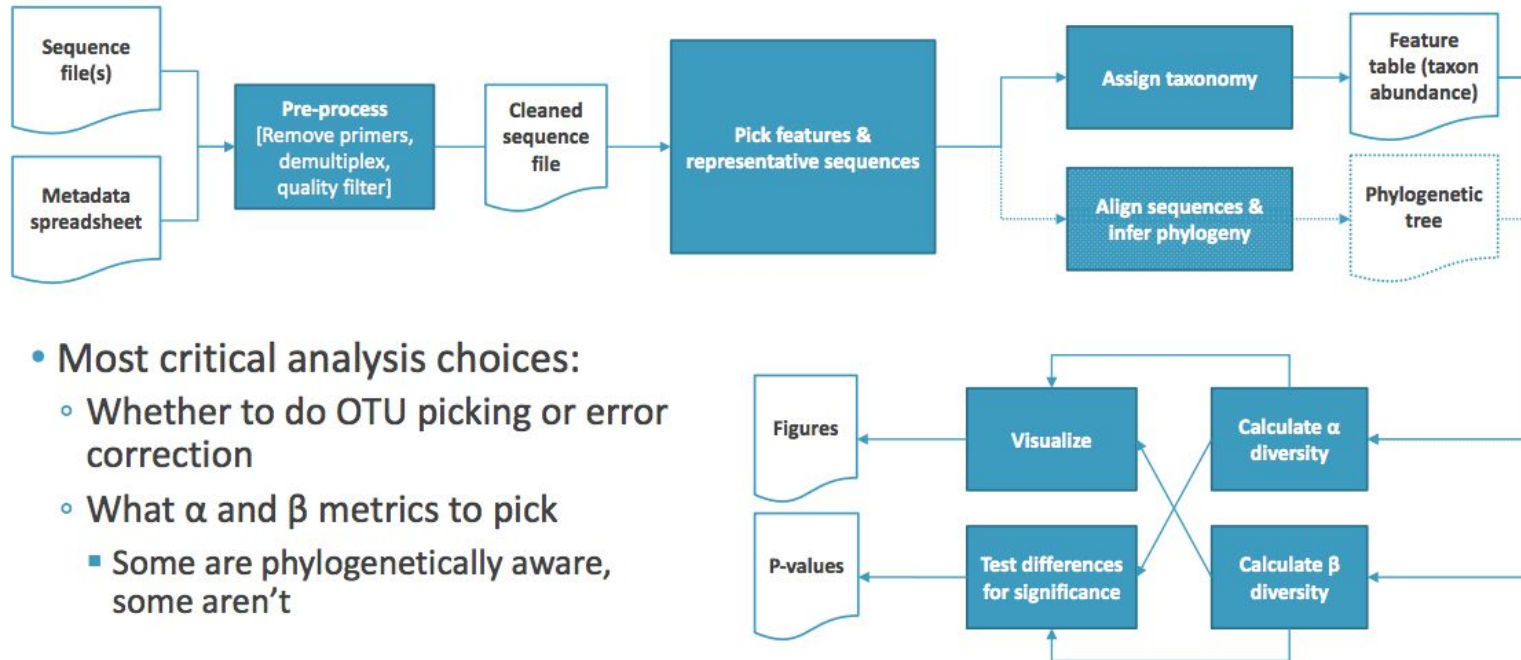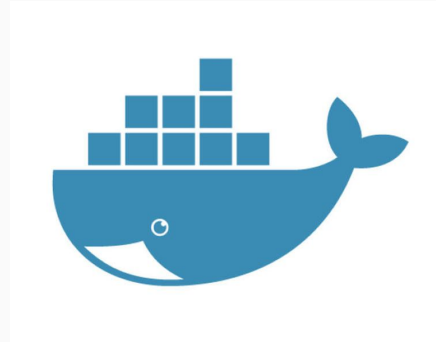
a
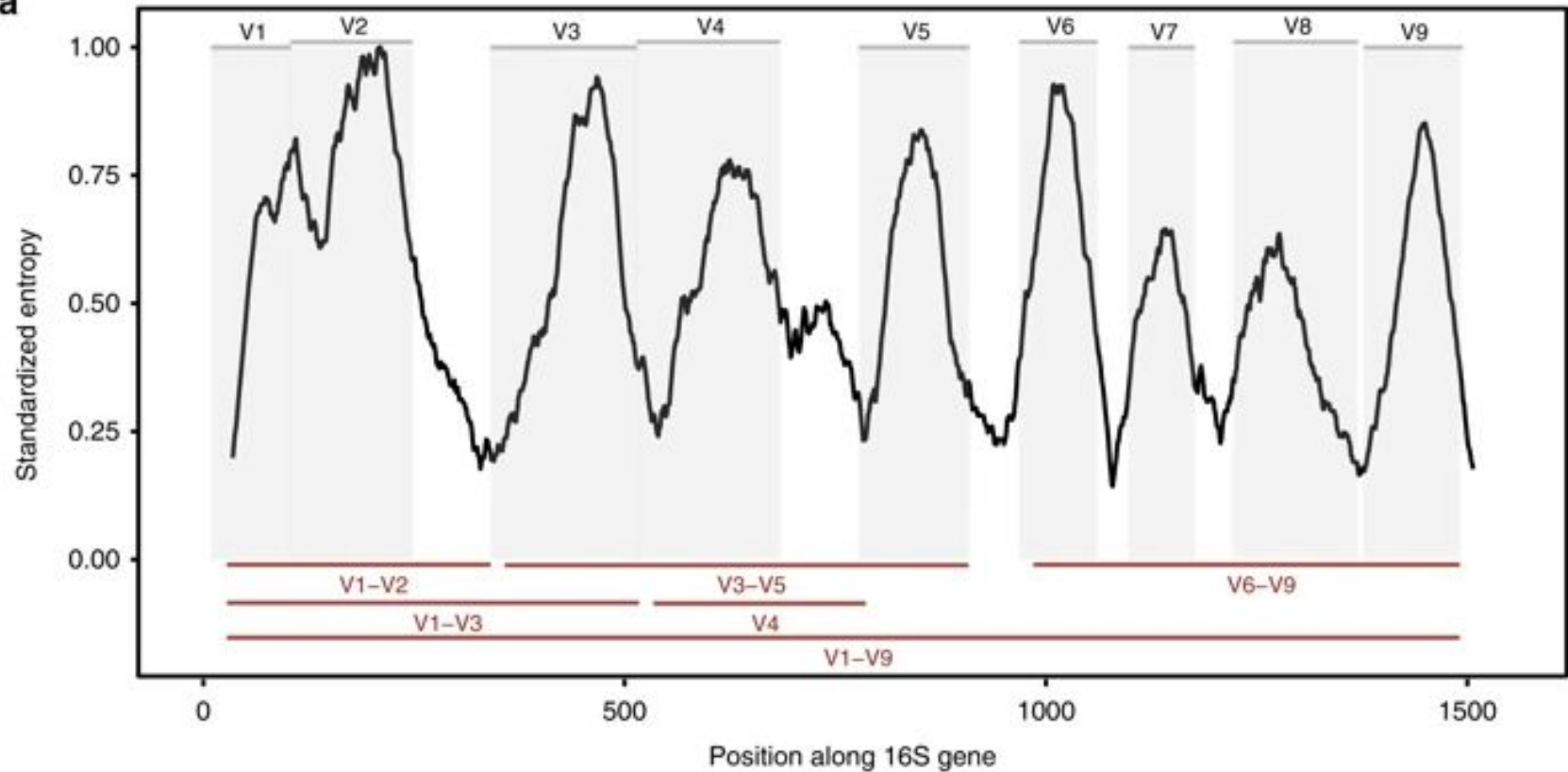
What is qiime2?

- Chiime
- Quantitative Insights Into Microbial Ecology
- Open-source pipeline
- Python
- Wraps Popular algorithms
- Comparison and analysis of microbial communities
- Next iteration of QIIME
  - Cited 24,228 times (07 APR 2021)
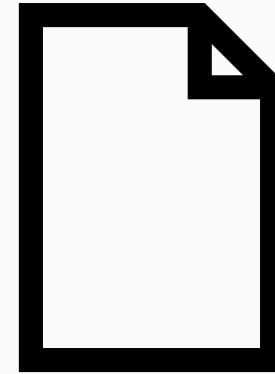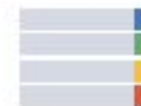 - Publication: Caporaso, J. Gregory, et al. (2010)

Qiime2 Workflow

- Most critical analysis choices:
  - Whether to do OTU picking or error correction
  - What α and β metrics to pick
    - Some are phylogenetically aware, some aren't

- Sequence files

  – Fastq

- Artifact files

  – .qza: Contains all data QIIME2 requires

 – .qzv: Files containing visualisation information

DADA2

- Divisive Amplicon Denoising Algorithm
- Models and corrects Illumina-sequenced amplicon errors
- Pipeline
  - Filtering
  - Dereplication
  - Chimera Identification
  - Merging paired-reads
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4927377/

PhiX removal

- Remove sequences similar to PhiX
  - **BLAST**
- PhiX Quality control
  - cluster generation
  - Sequencing
  - alignment.
- PhiX virus

https://en.wikipedia.org/wiki/Phi_X_174
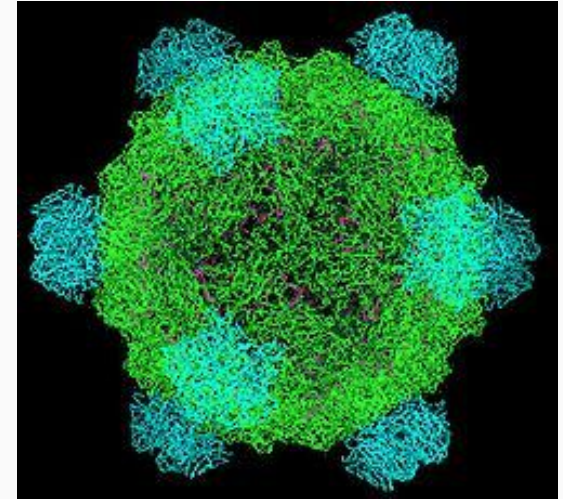
Chimera removal

DNA from two or more parent molecules
- PCR artifact
- Erroneous "novel" sequence

aborted amplification

next cycle's "primer"

chimeric sequence

for.   rev.

Merging reads

- Align read pairs R1/R2 to each other
- Improves quality of reads
- Longer reads

|	
R1 (30bp)	ACCGTACGTATGCGTAGCTGACGTAGCATG----------
R2 (30bp)	----------TGCGTAGCTGACGTAGCATGCGCGATTCGA
Overlap (20bp)	----------TGCGTAGCTGACGTAGCATG----------
Stitched read (40bp)	ACCGTACGTATGCGTAGCTGACGTAGCATGCGCGATTCGA

# OTUs vs Amplicon Sequence Variants (ASV)

- Previous methods
  - OTUs (Operational taxonomic units)
  - Cluster sequences by identity e.g. 97% similarity for species
  - Can cause over clustering
  - 97% is chosen due to errors within Illumina data
- DADA2
  - Denoises and cleans reads so they represent real sequences
  - Much finer resolution on sequences
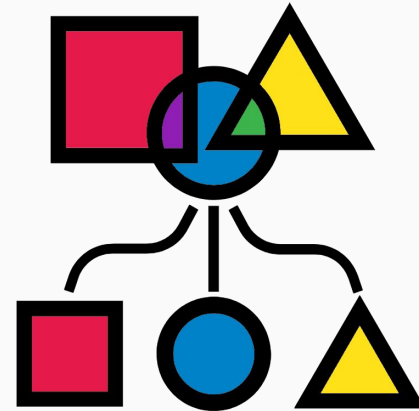 - Can differentiate sequences that have only 1bp difference

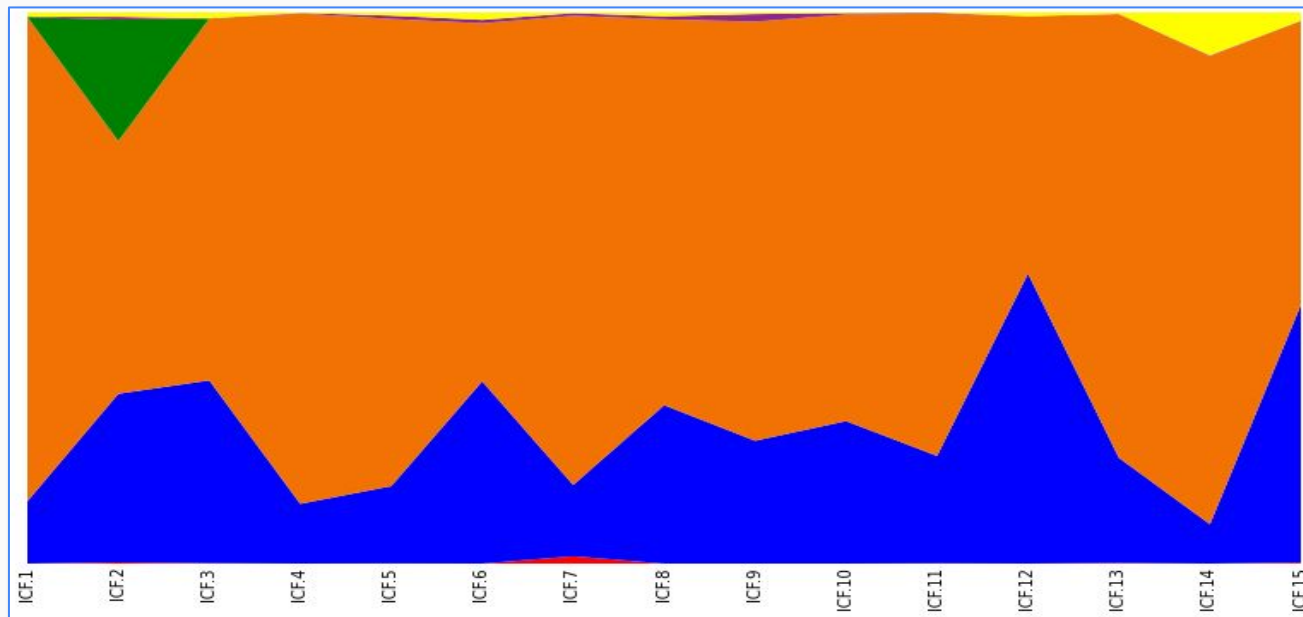https://www.youtube.com/watch?v=PmtqSa4Z1TQ
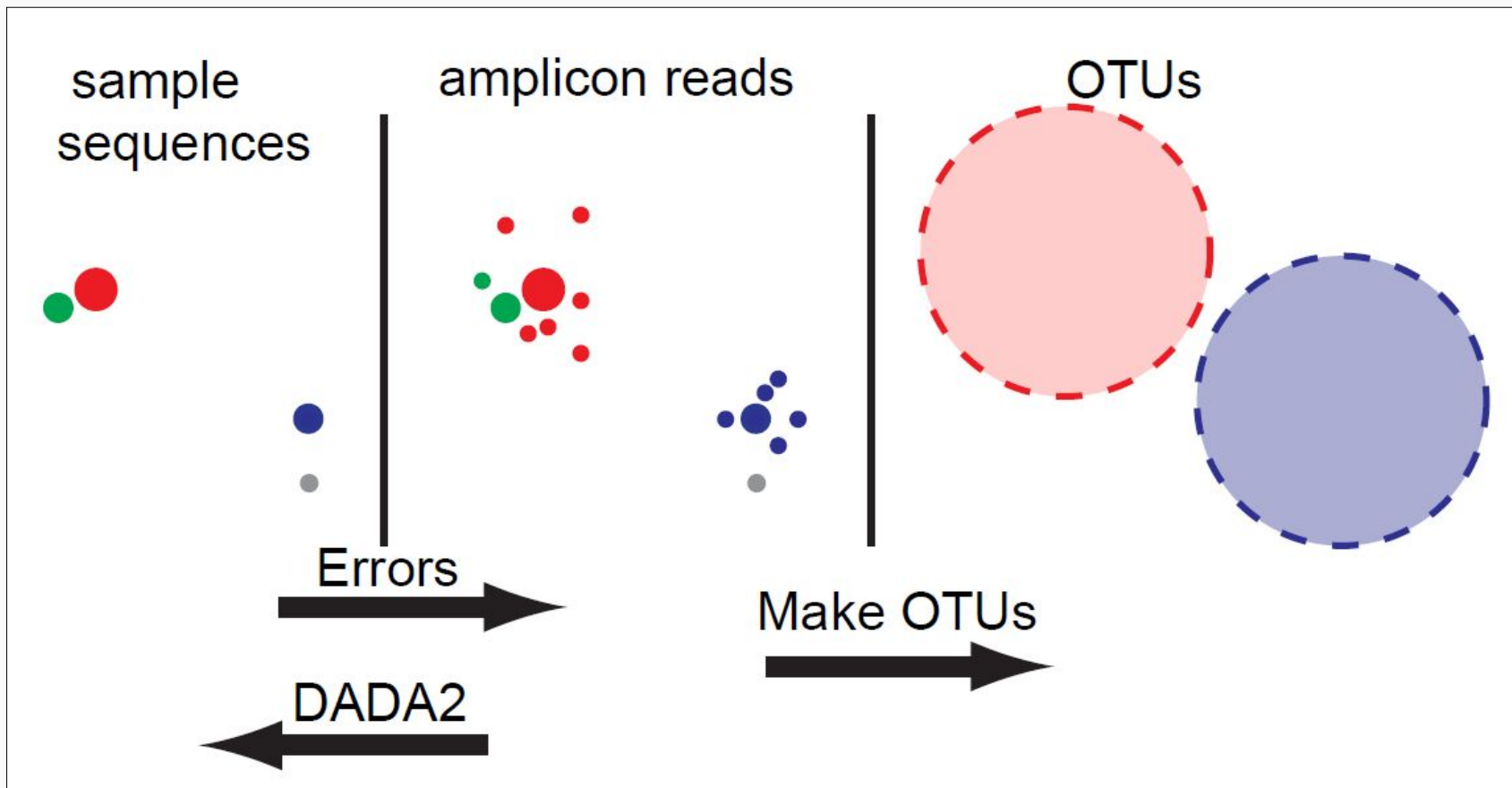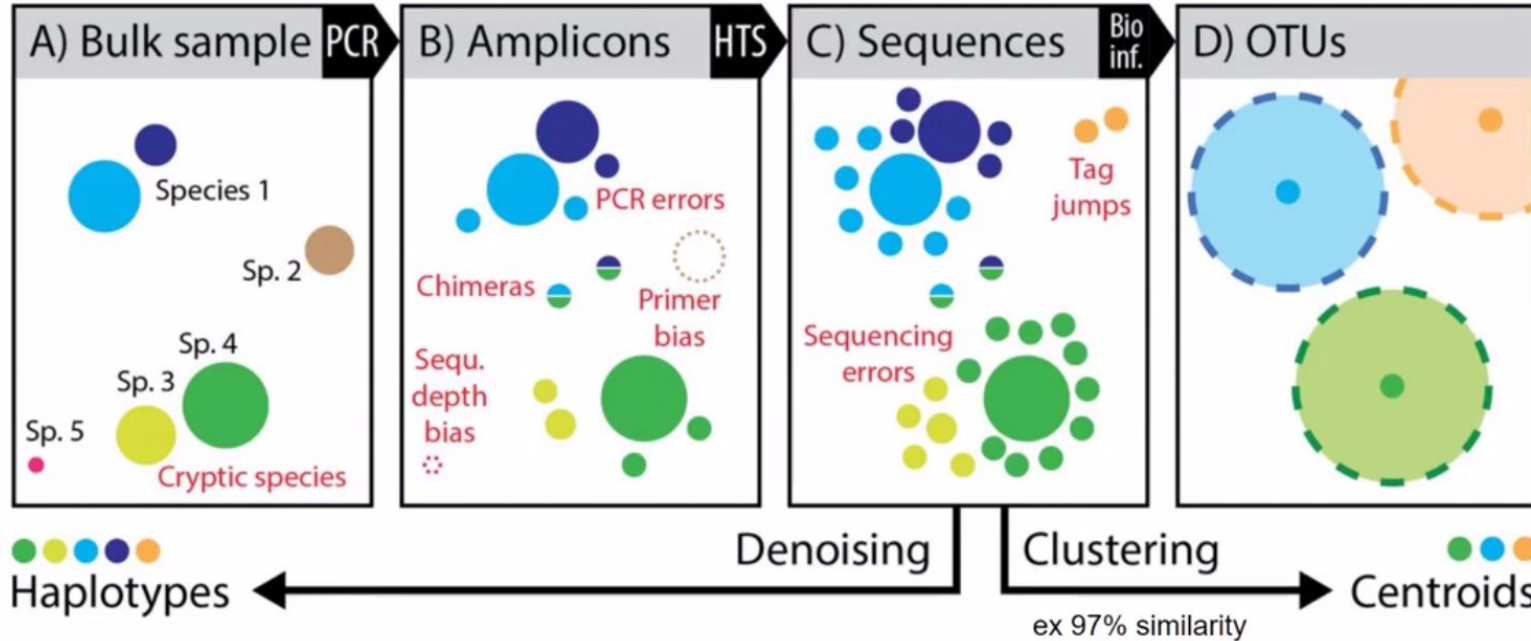
Taxa classification

- Assigns each ASV taxonomy based on a DB
- Reliant on:
  - Quality and completeness of DB
  - Tool used to search DB
- Two highly used DBs
 - Greengenes
  - Silva

Phylum level Taxonomy

| Legend	Taxonomy	Total count	%	ICF.1 %	ICF.2 %	ICF.3 %	ICF.4 %	ICF.5 %	ICF.6 %	ICF.7 %	ICF.8 %	ICF.9 %	ICF.10 %	ICF.11 %	ICF.12 %	ICF.13 %	ICF.14 %	ICF.15 %
	k__Bacteria;p__Actinobacteria	0	0.1%	0.0%	0.3%	0.1%	0.0%	0.0%	0.0%	1.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.2%
	k__Bacteria;p__Bacteroidetes	4	24.5%	11.3%	30.6%	33.2%	10.8%	14.0%	33.1%	12.9%	28.7%	22.3%	25.9%	19.5%	52.6%	19.1%	7.1%	46.7%
	k__Bacteria;p__Firmicutes	11	72.5%	87.8%	45.8%	65.6%	89.0%	84.8%	64.9%	85.1%	70.0%	76.0%	73.8%	80.2%	46.7%	80.5%	84.9%	51.5%
	k__Bacteria;p__Fusobacteria	0	1.5%	0.0%	22.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k__Bacteria;p__Proteobacteria	0	0.3%	0.0%	0.5%	0.0%	0.1%	0.6%	0.6%	0.5%	0.5%	1.4%	0.3%	0.2%	0.0%	0.0%	0.1%	0.0%
	k__Bacteria;p__Tenericutes	0	1.1%	0.8%	0.8%	1.1%	0.1%	0.6%	1.4%	0.1%	0.7%	0.3%	0.1%	0.0%	0.7%	0.3%	7.8%	1.5%

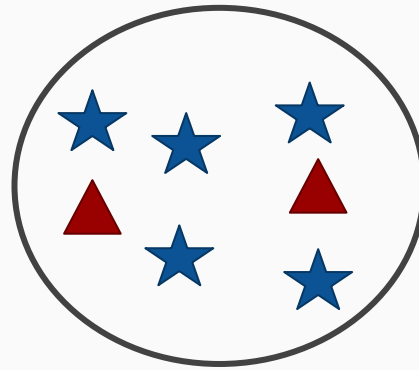# Genus level taxonomy
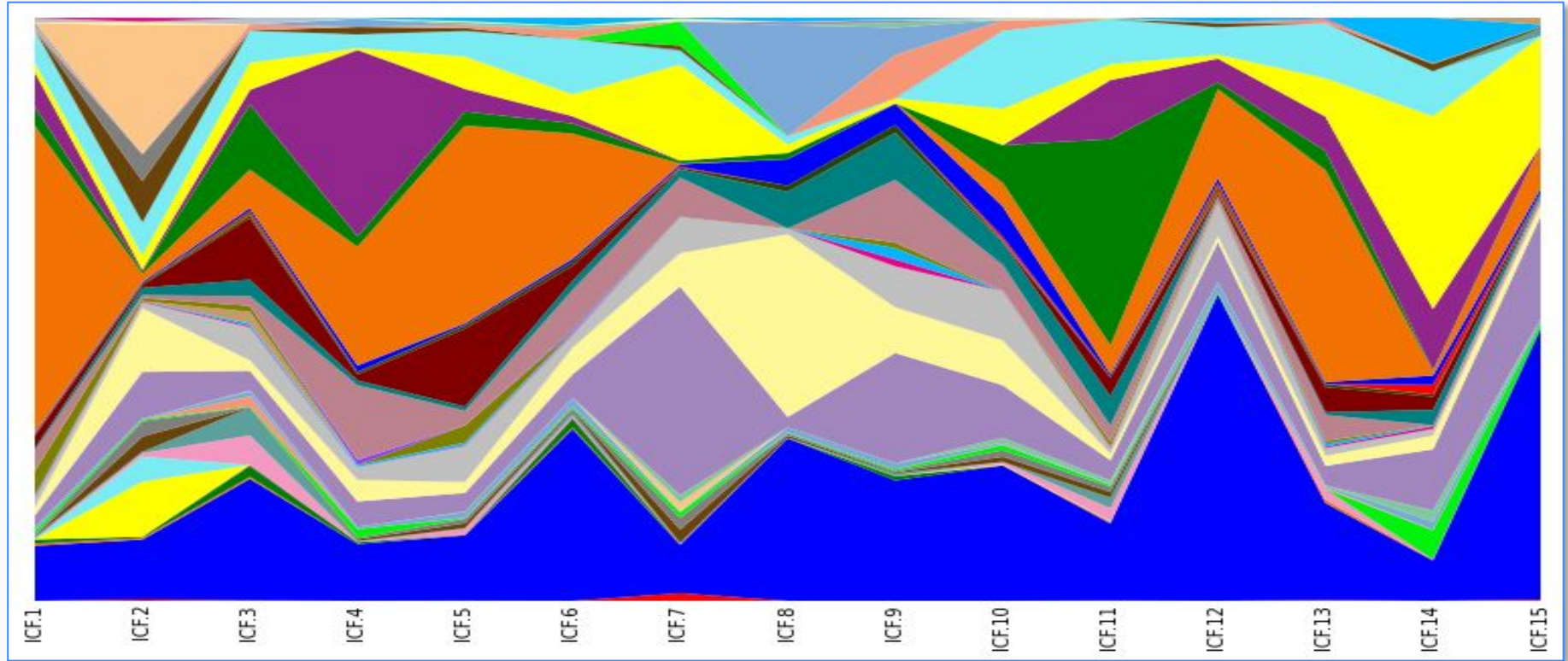
Biodiversity

- Richness
- Relative abundance
- Evenness
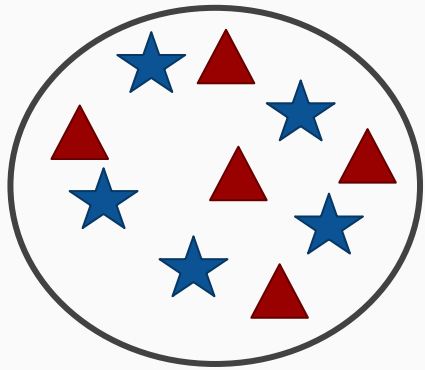
Richness: 2

Relative
abundance:
▲ 5/10 = 0.5
★ 5/10 = 0.5

Even
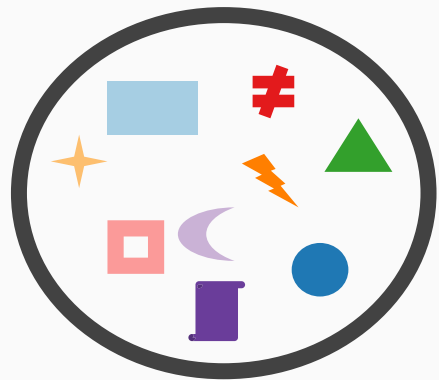
Richness: 2

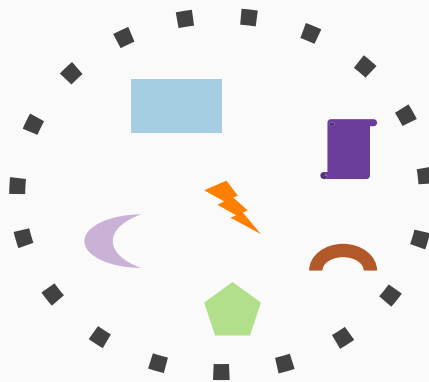Relative
abundance:
▲ 2/7 = 0.29
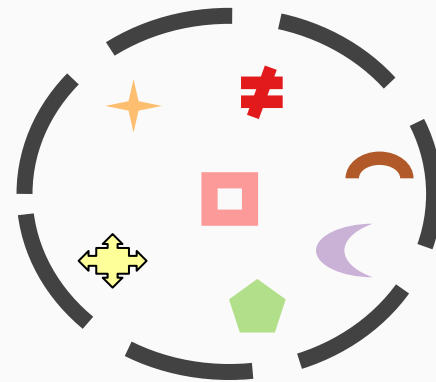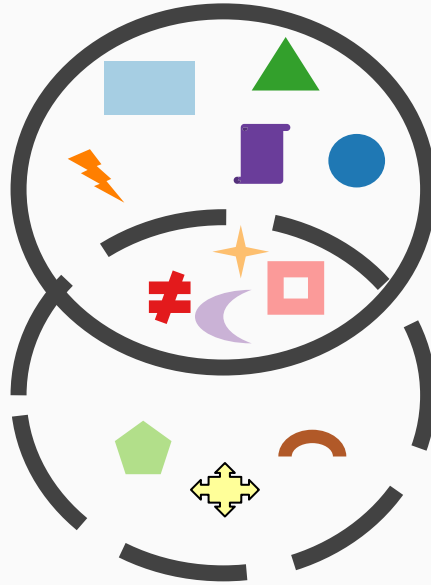★ 5/7 = 0.71

Uneven

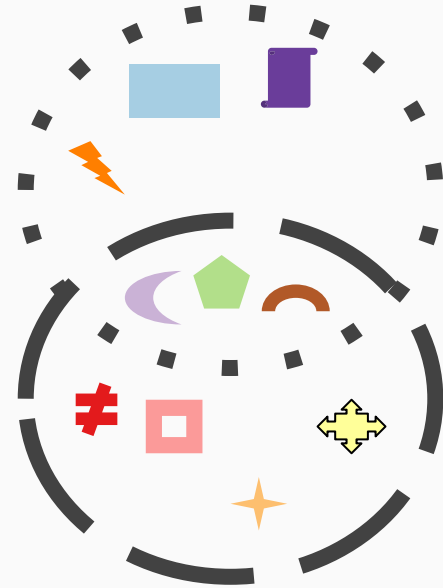Richness example

A
Richness: 9

B
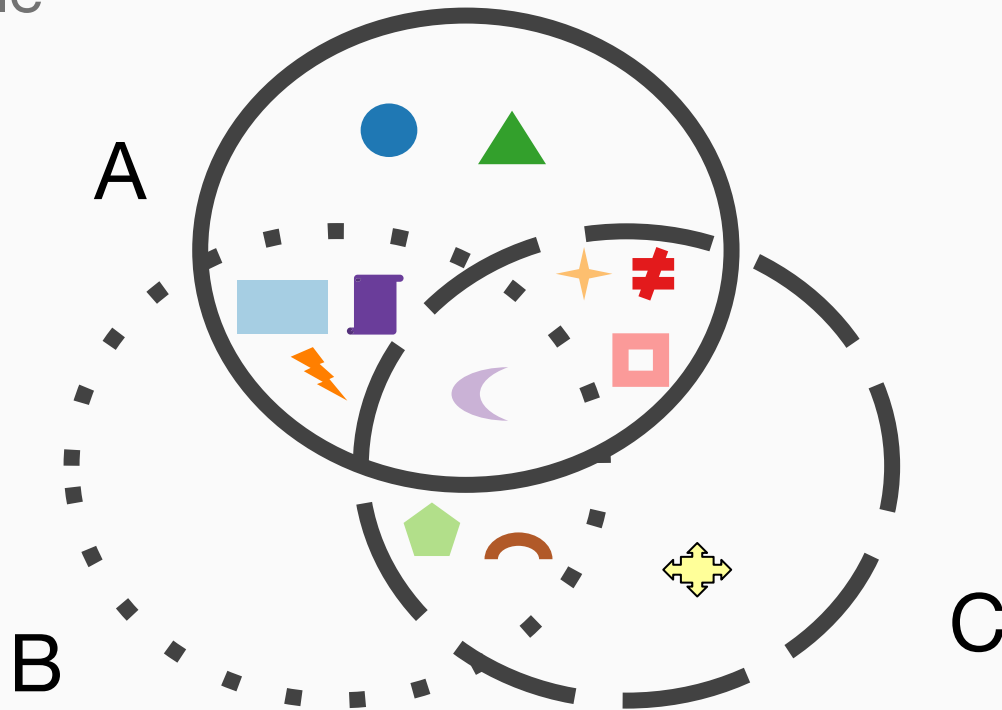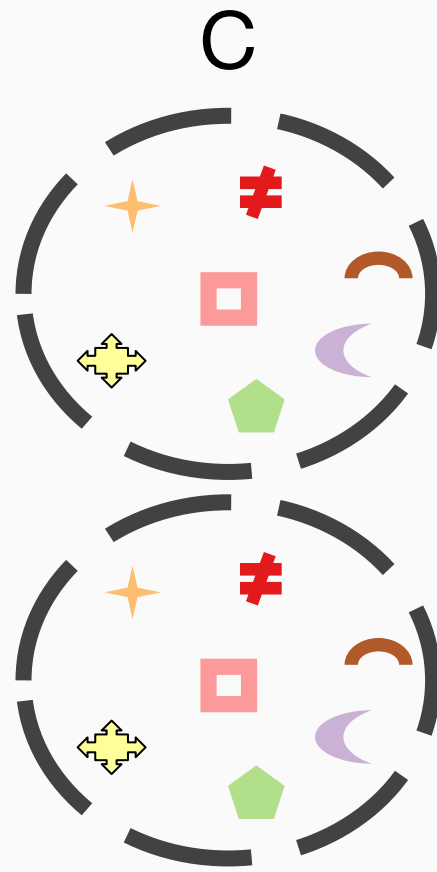Richness: 6

C
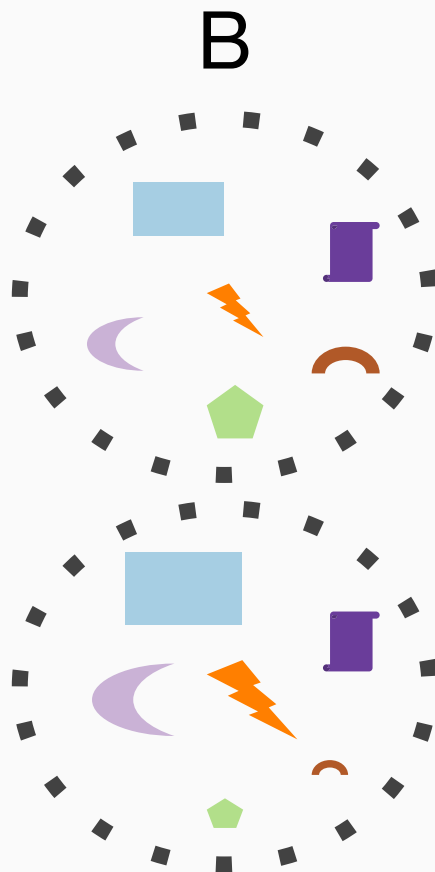Richness: 7
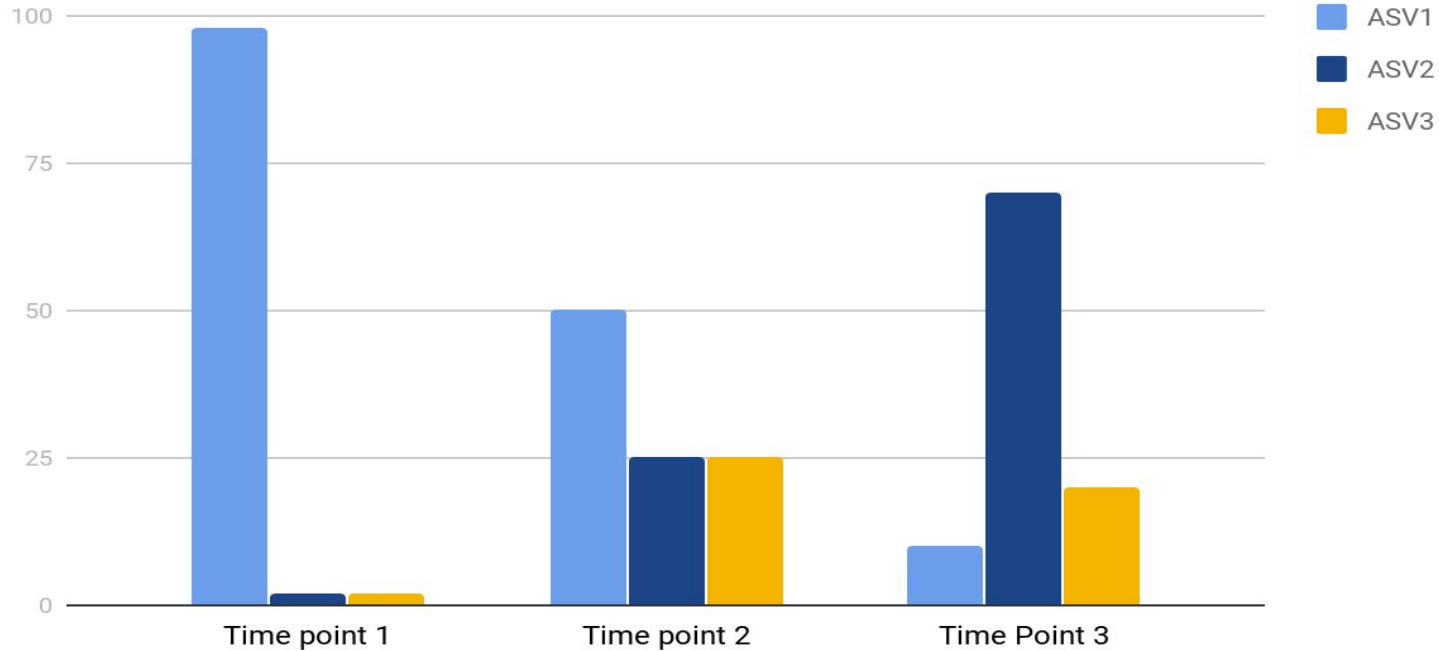
Richness example

A vs B   A vs C   B vs C

Richness example

Relative abundance

Richness

Relative abundance

A   B   C

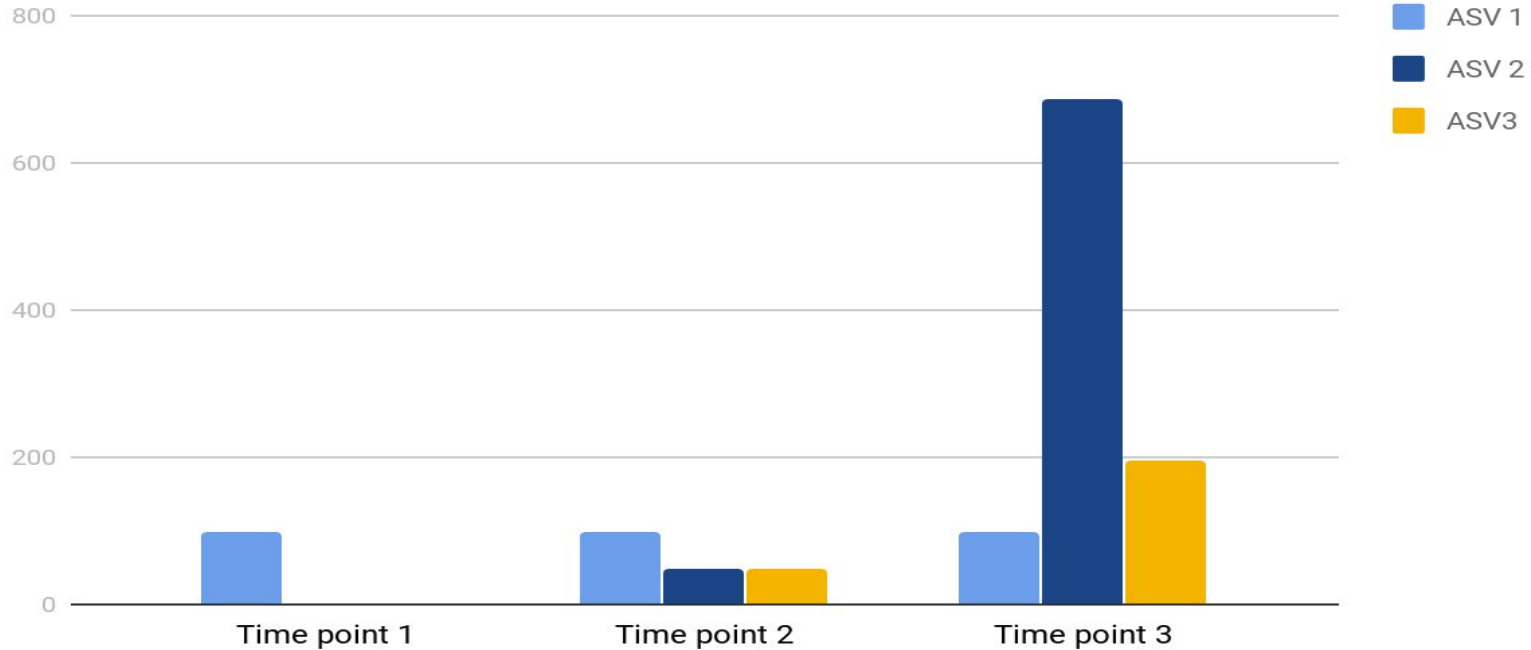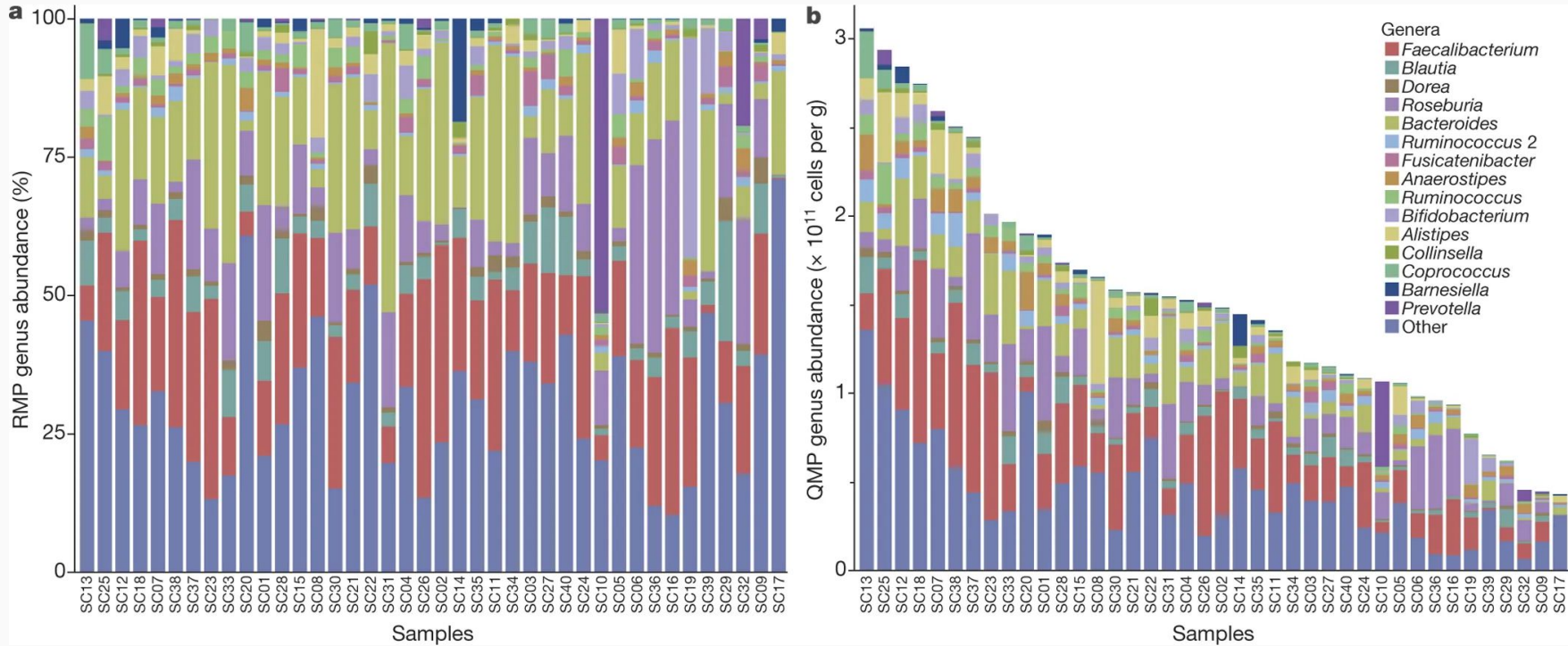# Relative abundance

Total abundance

Relative versus quantitative

Vandeputte, Doris, et al. "Quantitative microbiome profiling links gut community variation to microbial load." *Nature* 551.7681 (2017): 507-511.

- Biomarker detection
- Both statistically heavy methods
- ANCOM
  - Analysis of composition of microbiomes: a novel method for studying microbial composition
  - Differential abundance analysis
  - ANCOM assumes that less than ~25% of features are changing between groups.
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4450248/
- gneiss
  - Differential abundance analysis
  - Attempts to account for using relative abundance through balances
  - In essence it looks at log ratios try to determine real changes

Online Resources

- QIIME2 documents
    - https://docs.qiime2.org/2021.2/tutorials/

Useful papers

- What is new and relevant for sequencing-based microbiome research? A mini-review
  - Johnson, Jethro S., et al. "Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis." Nature communications 10.1 (2019): 1-11
- Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis
  - Vandeputte, Doris, et al. "Quantitative microbiome profiling links gut community variation to microbial load." Nature551.7681 (2017): 507-511.
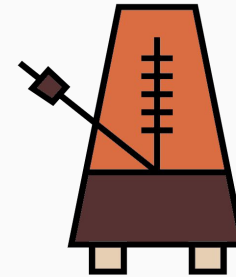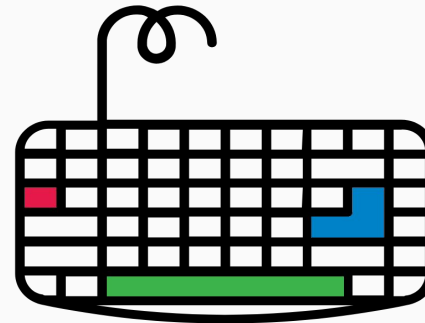
Reminders and Tips

Work at your own pace

Typos

Ask questions
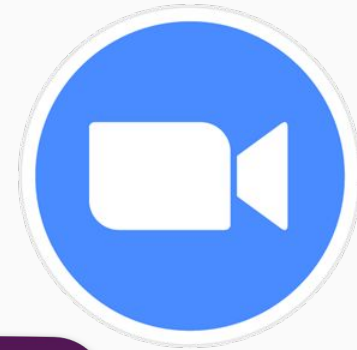
Breaks are important

Tab, space, and enter
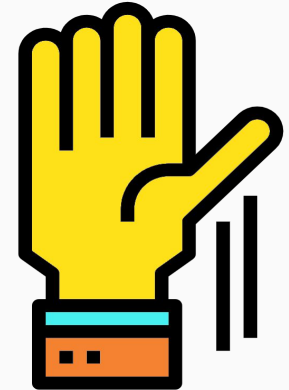
Online class info

Zoom - Ask via microphone if no question currently being asked/answered

Slack - Ask questions via the channel or ask to go into a zoom breakout room with one of us

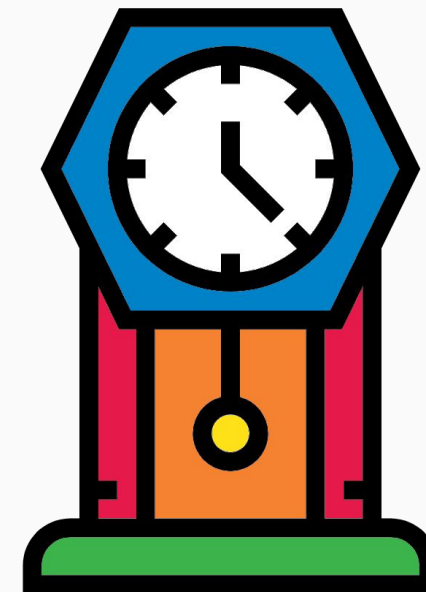WebVNC - We can connect to your webVNC to see and help with issues.

Format & Schedule

This intro
Bookdown
Theory
Practice
Exercises
MCQs
Optional materials

Work at your own pace on your own time

Book

Online HTML book
Read through and follow instructions
Link is in workshop agenda
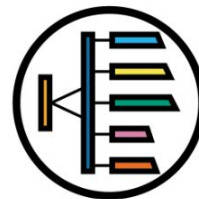Active learning

Bacterial 16S metabarcoding

Luca Lenzi and Matthew R. Gemmell

2022-08-22

Chapter 1   Introduction

This practical session aims to introduce you to the analysis of bacterial 16S metabarcoding with QIIME2. The topics covered are:
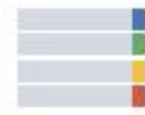
- Background on the biology
- Introduction to QIIME2
- Cluster and webVNC information
- Information on initial data
- QIIME2 analysis workflow

Recap

- Why 16S rRNA is used for microbial community analysis
- Qiime2 and its 16S rRNA workflow
- DADA2 quality control
- Alpha, Beta and Gamma Diversity analysis
- Relative abundance
- ANCOM & GNEISS
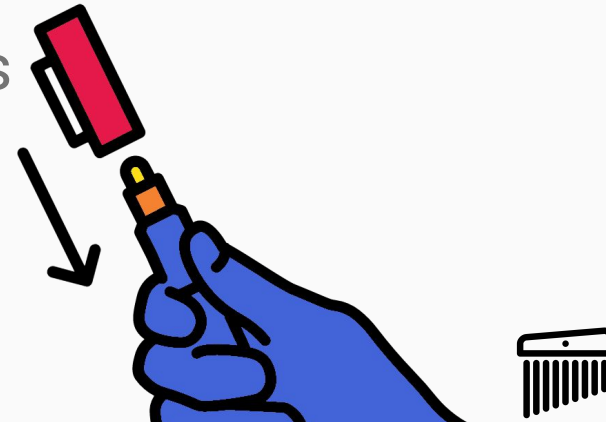
Segmentation fault

Thank you!

Questions?