Module 13: The phyloseq package

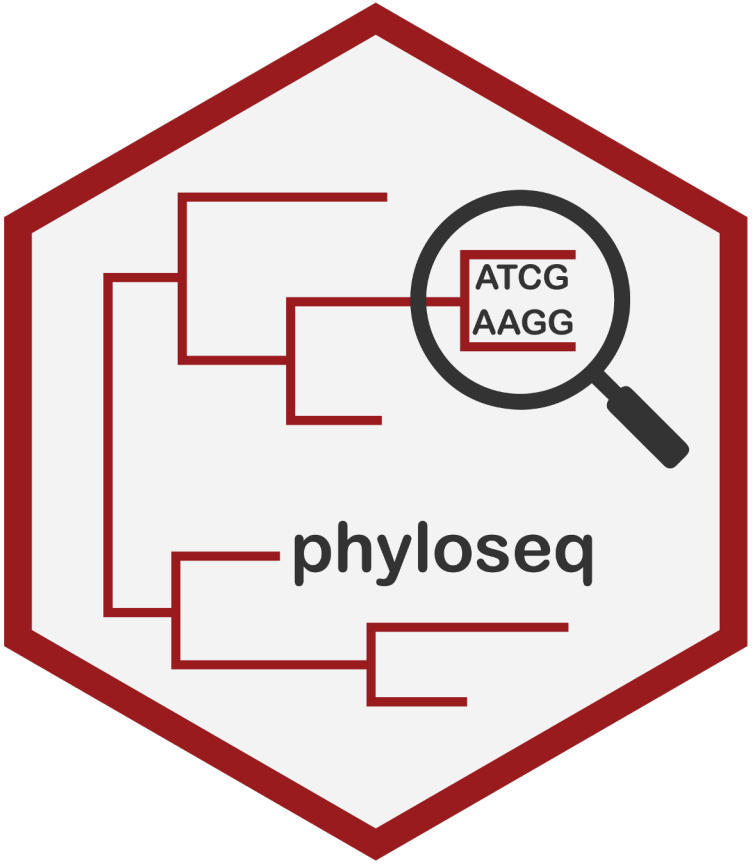
For microbial dataset analysis

Learning Outcomes

- Be able to create phyloseq objects
- Filter taxa and samples from phyloseq objects
- Rarefy ASV tables in a phyloseq object
- Be able to re-create taxa summary plots, alpha diversity boxplots, and beta diversity PCoA plots with phyloseq

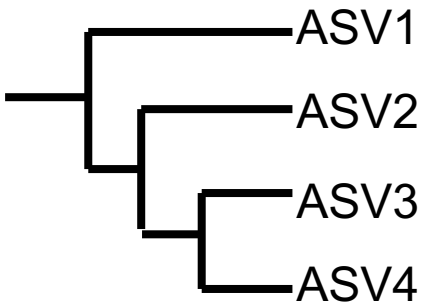
The ‘phyloseq’ package

- Enables analysis of microbial community data
- Integrates analysis, and creating graphics using a ‘phyloseq’ object



Phyloseq: reconciling all the outputs

Phylogeny (tree)



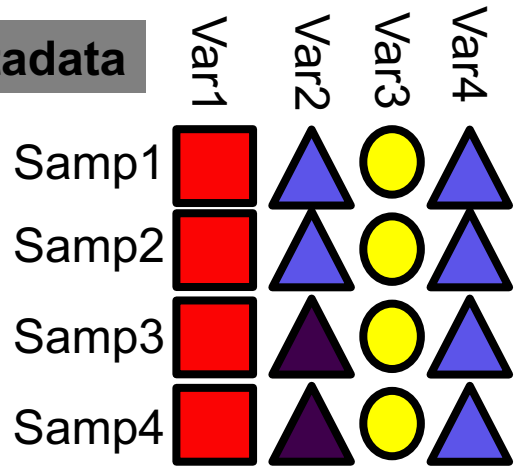
Taxonomy

ASV1	Bacteria	Proteo	...
ASV2	Bacteria	Cyano	...
ASV3	Bacteria	Bacter	...
ASV4	Bacteria	Planct	...

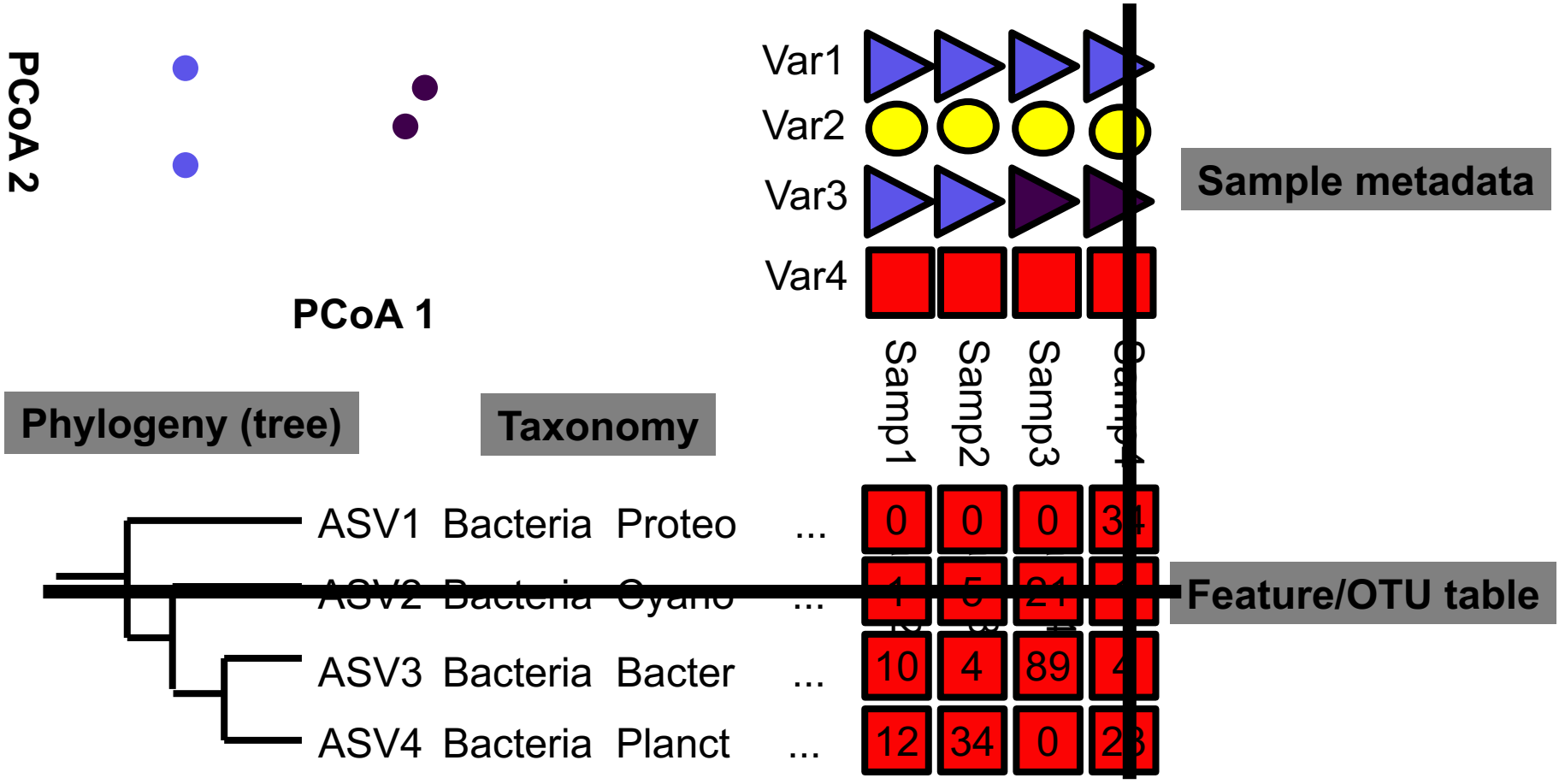
Feature/OTU table

	Samp1	Samp2	Samp3	Samp4
ASV1	0	0	0	34
ASV2	1	5	21	1
ASV3	10	4	89	4
ASV4	12	34	0	23

Sample metadata



Phyloseq objects allow you to filter, analyze, and plot data from multiple data frames/matrices together

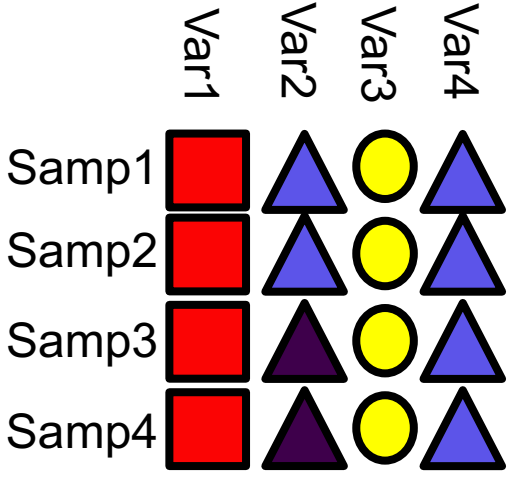


Step one: re-format tables

Feature/OTU table

	Samp1	Samp2	Samp3	Samp4
ASV1	0	0	0	34
ASV2	1	5	21	1
ASV3	10	4	89	4
ASV4	12	34	0	23

Sample metadata



Taxonomy

	Rank1	Rank2	...
ASV1	Bacteria	Proteo	...
ASV2	Bacteria	Cyano	...
ASV3	Bacteria	Bacter	...
ASV4	Bacteria	Planct	...

Step two: Convert to phyloseq objects and load into phyloseq()

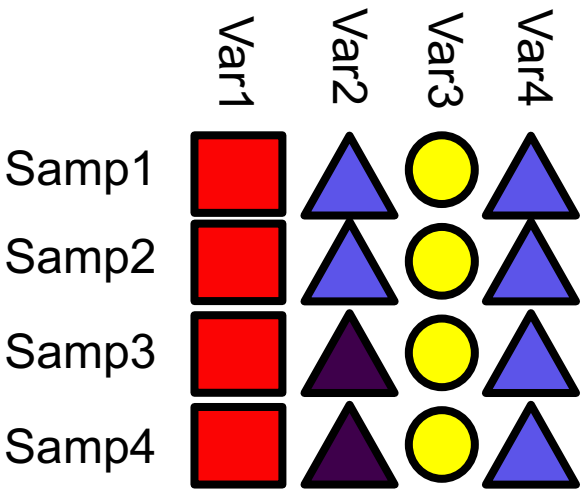
- sample_data() → metadata
- otu_table() → features table
- tax_table() → taxonomy table
- phyloseq() → combine all elements

Step three: Filter and rarefy as needed

- Filter out any mitochondria, chloroplasts, eukaryotes, archaea
  - subset_taxa()
- Filter low-abundance taxa
  - filter_taxa() or prune_taxa()
- Remove any “bad samples”
  - subset_samples() or prune_samples()
- Rarefy at appropriate depth
  - rarefy_even_depth()

subset vs prune

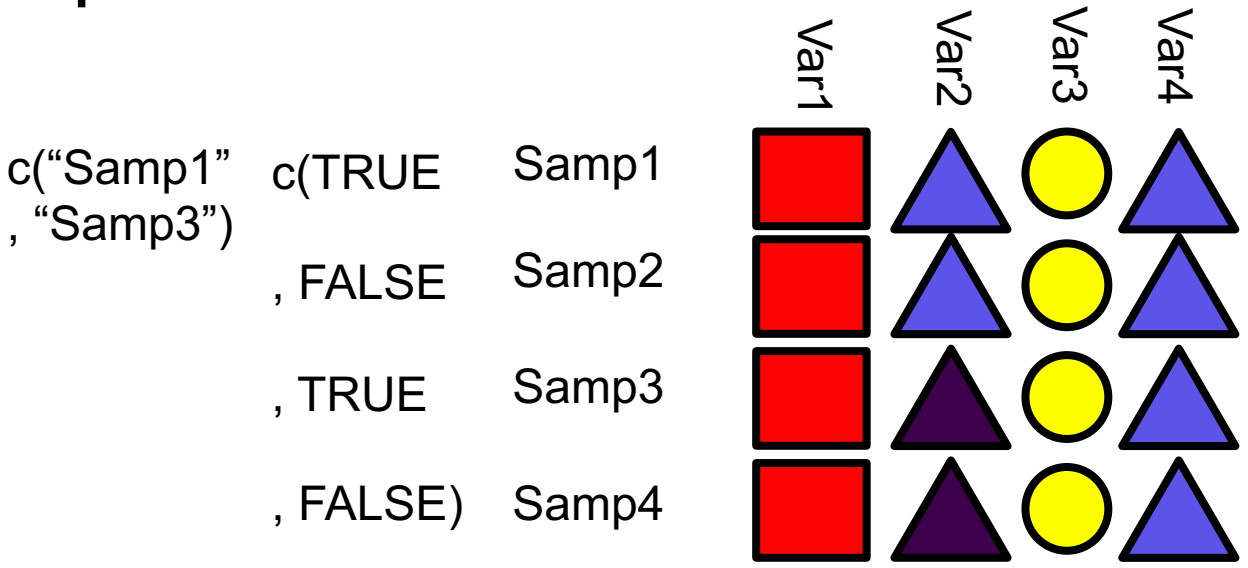
- Subset functions use the same table to filter
  - e.g. to subset samples, we use filtering criteria from the sample data table



	Rank1	Rank2	...
ASV1	Bacteria	Proteo	...
ASV2	Bacteria	Cyano	...
ASV3	Bacteria	Bacter	...
ASV4	Bacteria	Planct	...

subset vs prune

- Prune functions are more general; they use a vector of logicals or characters to choose what to keep



Recommend saving at this point

- Having a “final, clean, master dataset” is useful
- Can come back to this point to do many different analyses
- Ensure your data for all analyses is the same version

How to produce figures for microbial community analysis

- Alpha diversity plots
  - plot_richness()
- Beta diversity ordinations
  - dist(), ordinate() then plot_ordination()
- Taxonomic summaries
  - plot_bar(), taxa_glom() to combine taxa

Using phyloseq objects for other applications

- Can export data into regular data.frames
- Can export data into DESeq using the command phyloseq_to_deseq2()
- Can use phyloseq objects in “microbiome” package