MMB-117

Amplicon sequence data analysis

Normalization / Transformation

- Sequence counts vary between samples
- No biological meaning (technical)
- Limits of sequencing instrument

  → normalisation/transformation of data

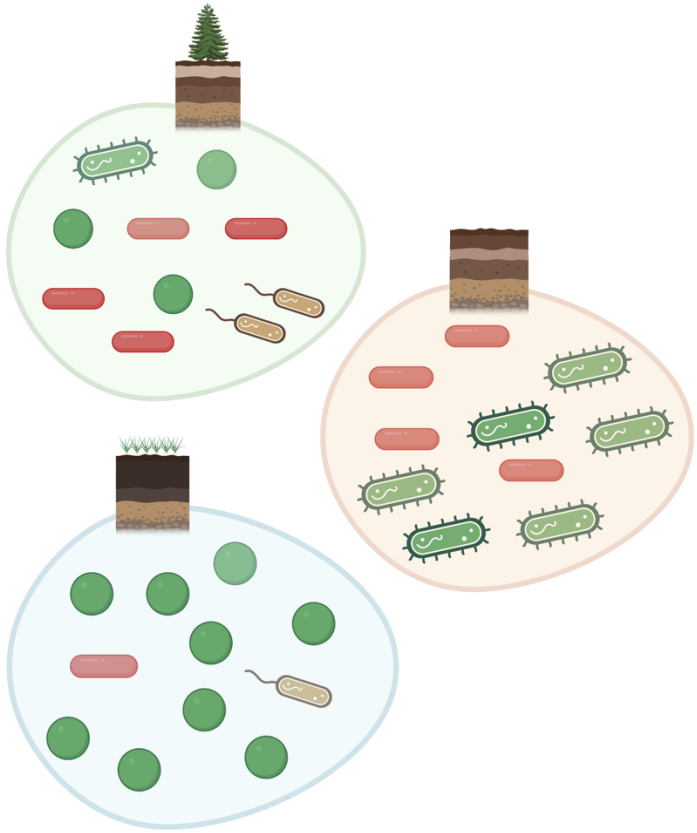
- Different ways to normalise:
  - Relative abundances / total sum scaling (TSS) / Proportions
  - Rarefaction: Subsample each library to same size
- Compositional data analysis:
  - Centered logratio transformation (CLR): $\mathrm{clr}(\mathrm{x}) = \log\frac{X}{g(x)}$

Alpha-diversity – within sample

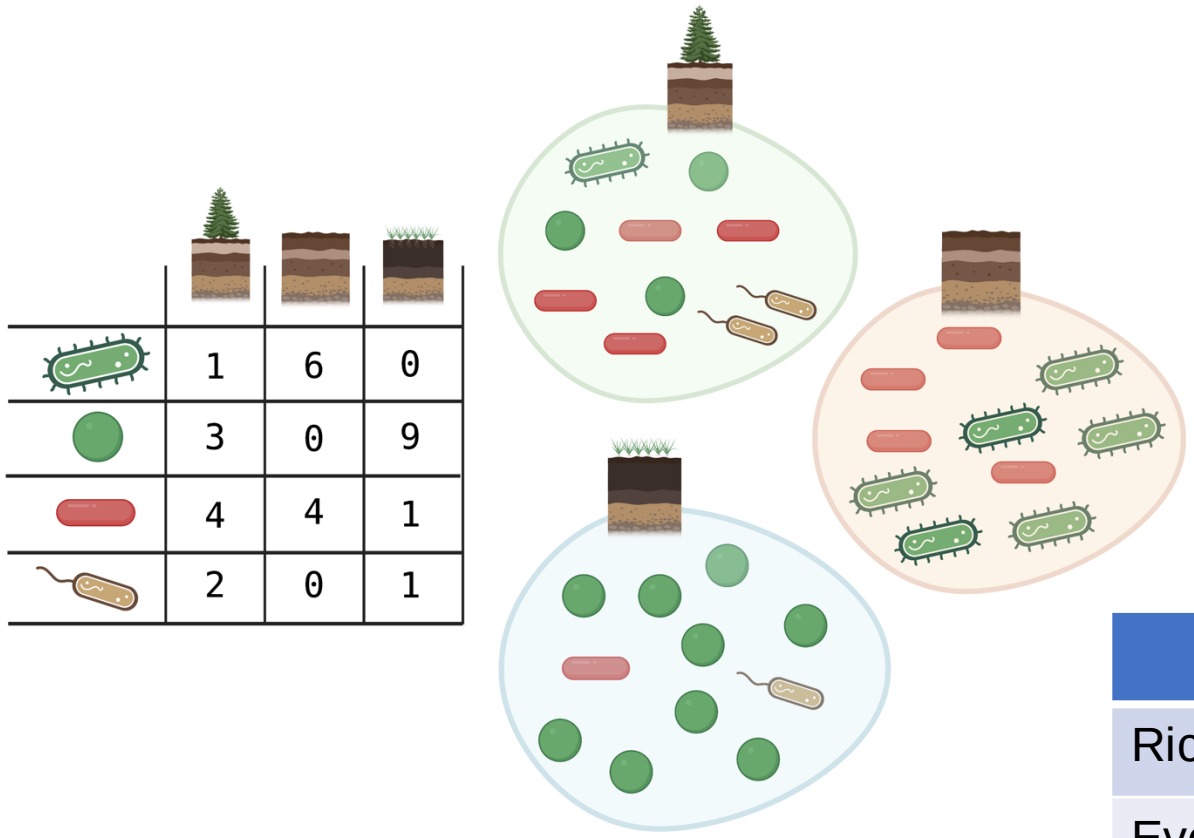
- Species richness
  - How many species we detected
  - Sequencing depth affects (rarefaction needed)
- Simpsons index
  - Evenness – odds that two randomly picked microbes belong to the same species
- Shannon’s diversity index
  - Diversity – combines richness and evenness

Alpha-diversity – within sample

	1	6	0
	3	0	9
	4	4	1
	2	0	1



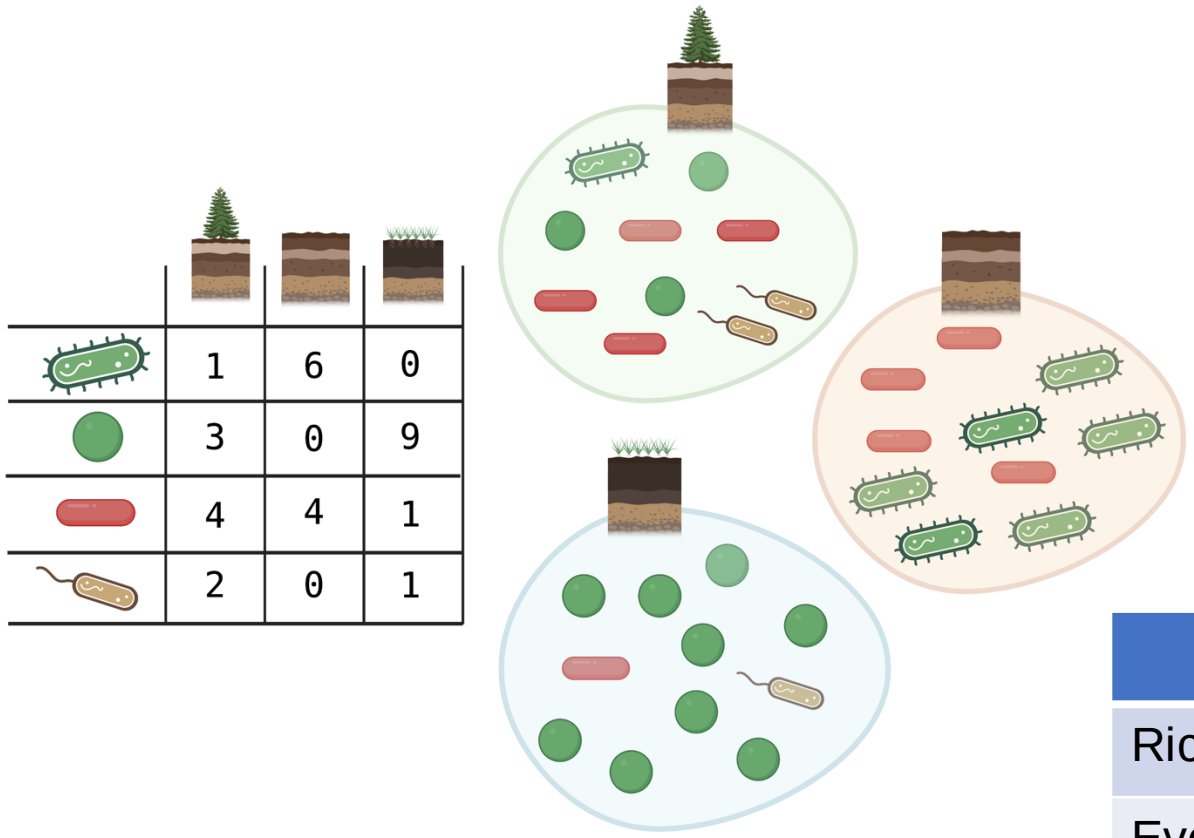
Alpha-diversity – within sample



	1	6	0
	3	0	9
	4	4	1
	2	0	1

	Forest	Agri	Park
Richness			
Evenness			
Diversity			

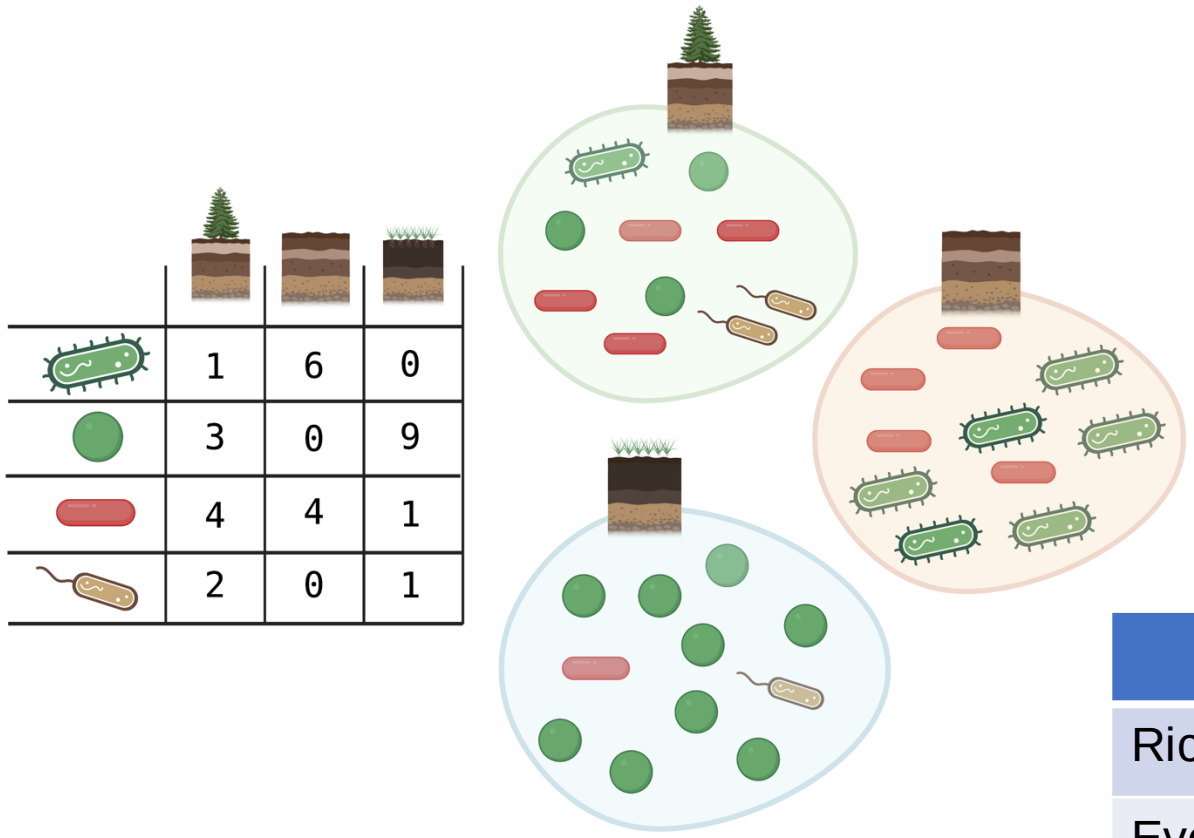
Alpha-diversity – within sample



	1	6	0
	3	0	9
	4	4	1
	2	0	1

	Forest	Agri	Park
Richness	4	2	3
Evenness			
Diversity			

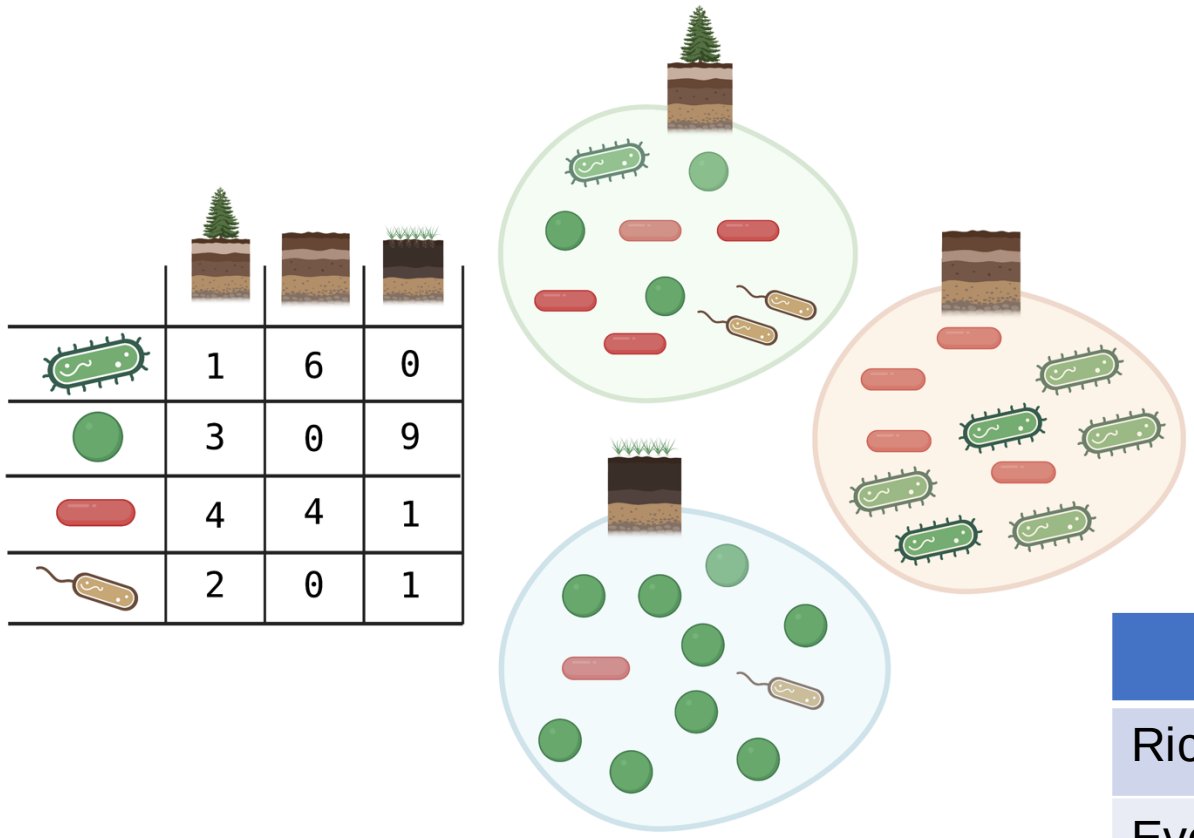
Alpha-diversity – within sample



	1	6	0
	3	0	9
	4	4	1
	2	0	1

	Forest	Agri	Park
Richness	4	2	3
Evenness	0.2	0.5	0.7
Diversity			

Alpha-diversity – within sample



	1	6	0
	3	0	9
	4	4	1
	2	0	1

	Forest	Agri	Park
Richness	4	2	3
Evenness	0.2	0.5	0.7
Diversity	1.3	0.7	0.6

Beta-diversity – between samples

- Different distance (or dissimilarity) indices
  - Euclidean (with CLR data)
  - Bray-Curtis (with relative abundances)

- Different ordination methods
  - Dimension reduction techniques
  - PCA, PCoA, nMDS