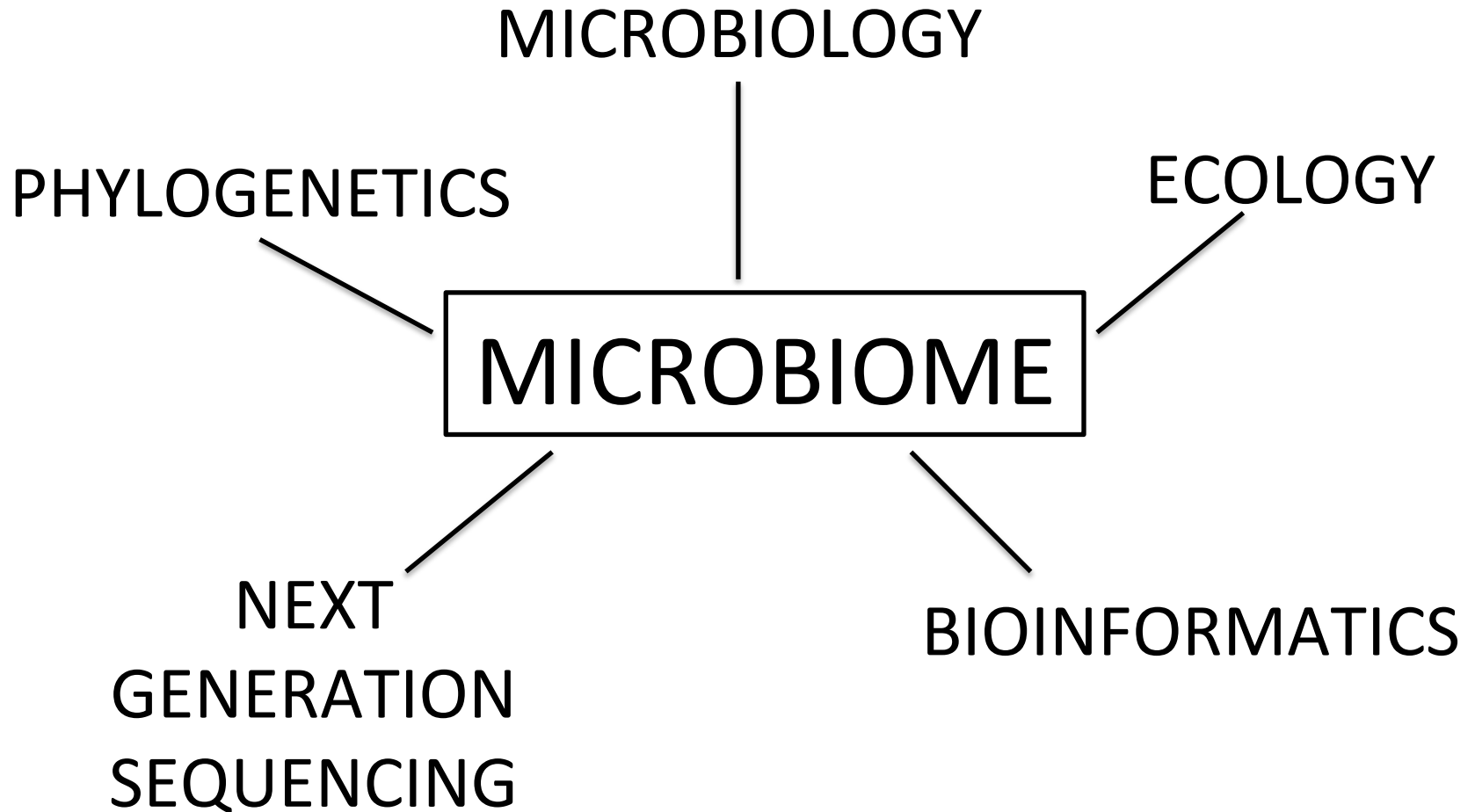


Microbiome Analysis

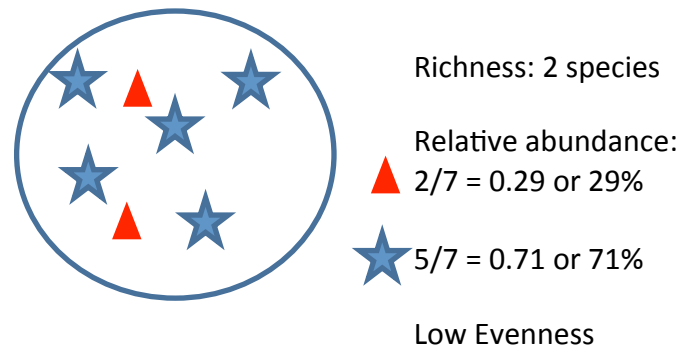
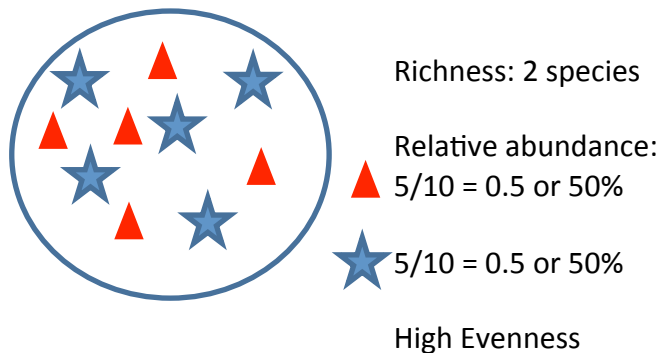


- How do we describe microbiomes?
- How do we compare microbiomes?

Important ecological concepts

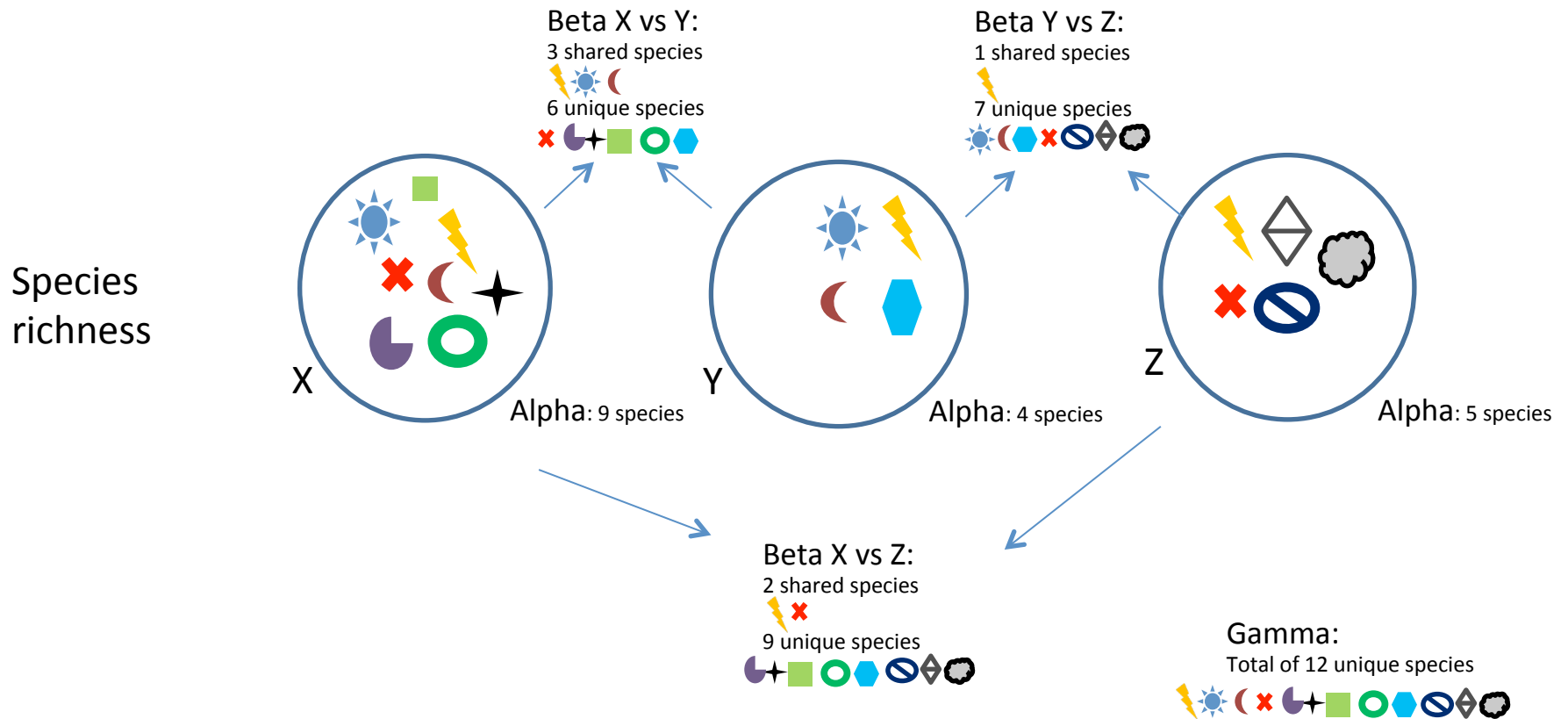
- How is biodiversity defined and measured?
- Component of biodiversity:
 - RICHNESS
 - EVENNESS
- Species richness: number of different species in a habitat/sample
- Species relative abundance: number of each species relative to total number of all species in a sample (number of reads per OTU in a sample relative to total number of reads in that sample)
- Species evenness: how close in numbers each species in an environment are; distribution

Simple example:



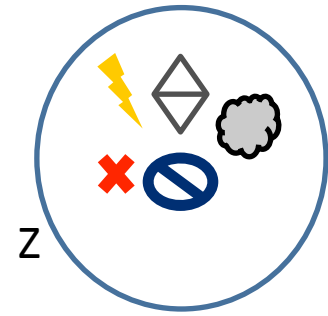
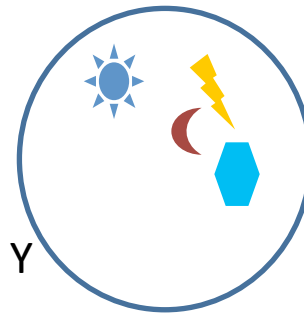
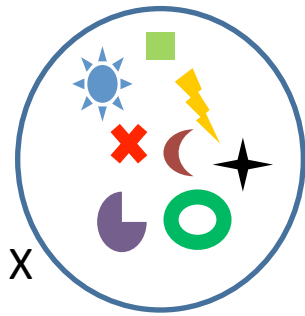
- Alpha diversity: Diversity within a single sample (Alone)
- Beta diversity: Diversity Between samples
- Gamma diversity: total diversity in a landscape

SIMPLE EXAMPLE

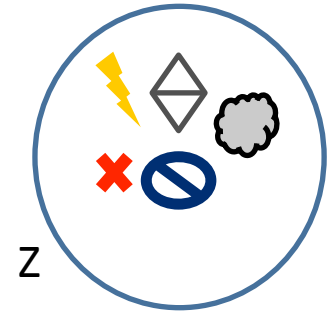
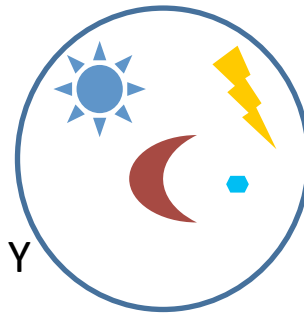
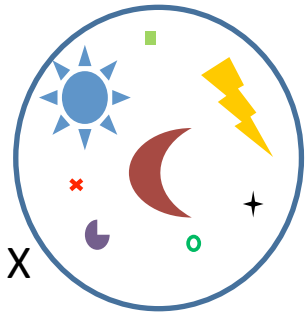


PROBLEM: Doesn't take abundance of each species OR relatedness of each species into account

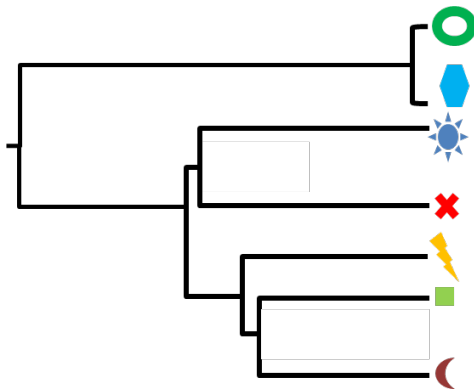
Species
richness



Size adjusted
according to
abundance



Phylogenetic
relationship



Metrics used to describe diversity measure
different aspects of the community

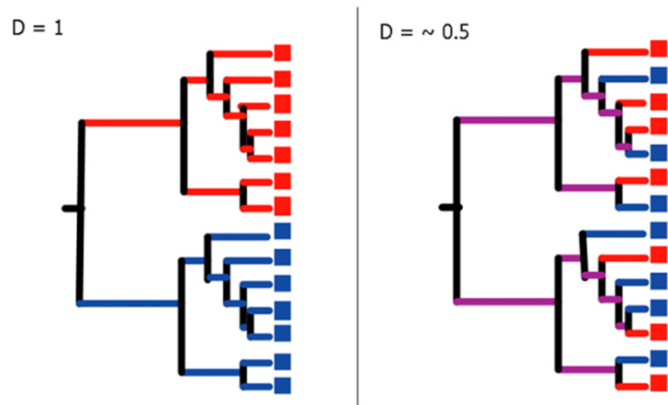
Alpha metrics

- Richness: observed species, chao1
- Diversity (Richness and Abundance / Evenness): Shannon, Simpson
- Phylogenetic: PD, Faith's PD or PD_whole tree

Beta diversity

- Diversity between samples
- Single metric to describe difference or similarity between samples
- Non-phylogenetic metrics
- Phylogenetic metrics

- Raw unweighted Unifrac: sum of branch length that is unique to one environment or the other

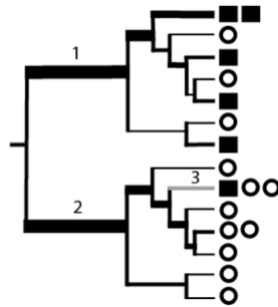


$$u = \frac{\sum_{i=1}^N l_i |A_i - B_i|}{\sum_{i=1}^N l_i \max(A_i, B_i)}$$

l_i is the branch length between node i and its parent, and A_i and B_i are indicators equal to 0 or 1 as descendants of node i are absent or present in communities A and B respectively

A = red, B = blue, branches in common are purple, branches unique to A are red and unique to B are blue. Presence/absence metric.

- Raw weighted Unifrac: Branch lengths are weighted by the relative abundance of sequences

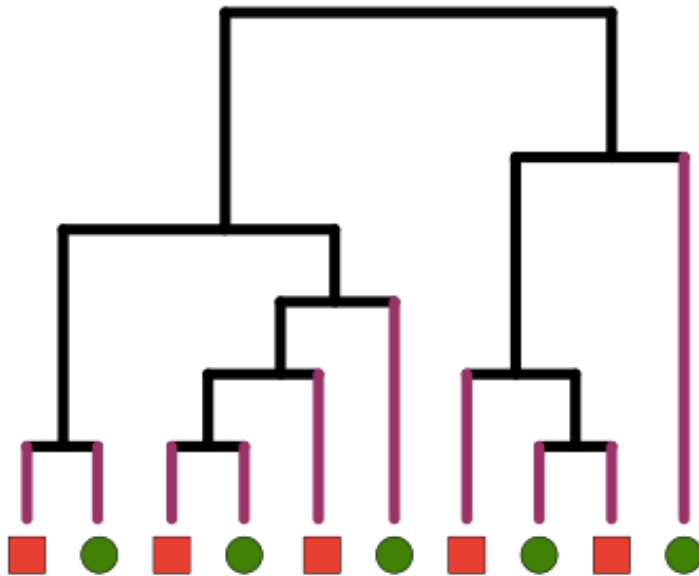


$$u = \sum_i^n b_i \times \left| \frac{A_i}{A_T} - \frac{B_i}{B_T} \right|$$

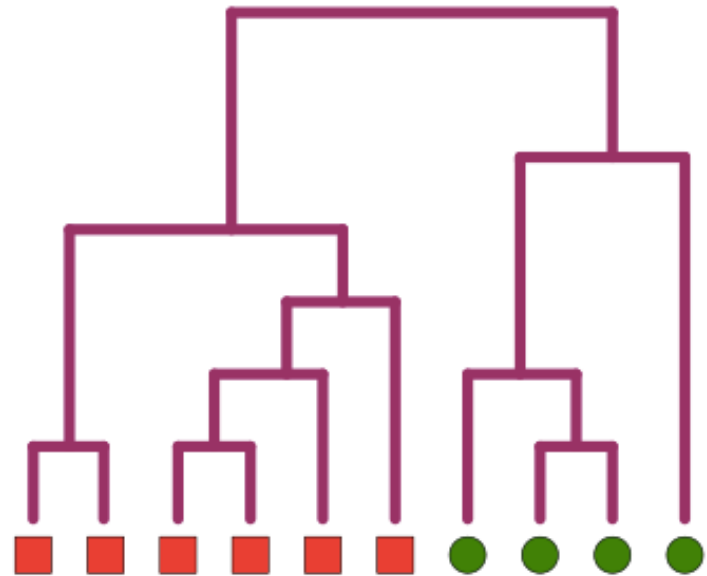
Here, n is the total number of branches in the tree, b_i is the length of branch i , A_i and B_i are the number of descendants of branch i from communities A and B respectively, and A_T and B_T are the total number of sequences from communities A and B respectively. In order to control for unequal sampling effort, A_i and B_i are divided by A_T and B_T .

- Normalised weighted Unifrac: takes abundance and normalises branch length
 - Rapidly evolving lineages (with long branch length can skew unifrac)

Similar Communities




Maximally Different Communities



$$\text{UniFrac Distance Measure} = \frac{\text{—}}{\text{—} + \text{—}}$$

 Branch length of unique branches

 Branch length of common branches

Unweighted Unifrac does not take abundance into account

Weighted Unifrac takes abundance into account – branch lengths weighted by relative abundance

Completely genetically different communities $D=1$

Exactly same communities $D=0$

PCoA is a method to explore and to visualize similarities or dissimilarities of data. It starts with a similarity matrix or dissimilarity matrix (= distance matrix) and assigns for each item a location in a low-dimensional space (2D or 3D graphics).

	Sample1	Sample2	Sample3	Sample4	Sample5	Sample6	Sample7	Sample8	Sample9	Sample10
sample1	-0.45272	0.11954	-0.081764	0.037794	-0.14426	-0.054638	0.03717	-0.045374	-0.0079748	-0.12573
sample2	-0.47655	0.089467	-0.067195	0.056111	-0.12911	0.057788	-0.025686	0.066917	-0.12545	-0.020286
sample3	-0.38372	0.045654	-0.098504	0.019517	0.2348	-0.16791	-0.019422	0.00015095	0.306	0.059276
sample4	0.40561	0.1961	-0.26087	0.35721	0.064719	0.049181	-0.058984	0.036286	-0.030997	-0.024002
sample5	-0.49884	0.076615	-0.10997	0.0848	-0.065639	-0.033154	-0.05038	0.02358	0.0064334	-0.069332
sample6	-0.46592	0.13353	-0.088553	0.026915	-0.092687	0.013809	0.07571	0.056925	-0.052395	0.028462
sample7	0.06255	0.12978	-0.049563	-0.43325	0.16133	0.3547	0.043509	0.064613	0.019788	0.085759
sample8	0.084894	0.017048	0.21576	-0.088903	-0.022797	-0.050414	0.17675	0.5677	0.063088	-0.15751
sample9	0.41945	0.28248	-0.12415	0.014038	-0.085404	-0.052874	-0.05431	0.0076042	0.039428	-0.01937
sample10	-0.44193	0.073919	-0.11333	0.070209	-0.095491	-0.048107	0.00069784	-0.0089602	-0.029425	0.082447

