Diversity metrics

There are many different ways that ecologists (microbial or otherwise) measure diversity. Diversity metrics are not created equal. When interpreting diversity metrics, make sure to take into account which factors of diversity are used, how they are weighted, and how they are scaled (e.g. linear, logarithmic).

Here, we provide descriptions of the most common diversity metrics, but this is by no means an exhaustive list.

Alpha diversity

Measures the diversity of a single sample (local diversity). Alpha diversity metrics are generally calculated as some combination of species richness and evenness.

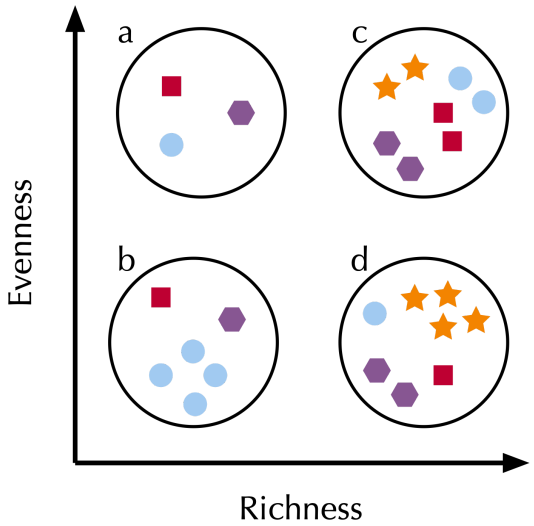
Richness:

- The number of species observed in a sample.

Evenness:

- A measure of how close the species in a sample are to one another in abundance.

Since richness and evenness are involved in calculating most alpha diversity metrics, here's a diagram to make them a little clearer:



Communities **a** and **b** are equally rich (3 species), as are communities **c** and **d** (4 species). Communities **a** and **c** are perfectly even, with all species being equally abundant. Communities **b** and **d** are uneven, as some species occur in higher abundance than others.

Observed OTUs / ASVs:

- The number of OTUs observed in a sample (essentially equivalent to richness).

Shannon's H:

- A measure of entropy. Takes into account both richness and evenness. Expressed on a logarithmic scale. Probably the most common alpha diversity metric.
- Puts more weight on rare species than Simpson's D.

Simpson's D:

- The probability that two randomly selected individuals will belong to the same species. Takes into account both richness and evenness.
- Gives abundant species more weight than Shannon's H.
- This is also called Simpson's Index. Simpson's Index of Diversity is 1 - D

Chao1:

- Abundance-based estimated richness. Assumes that rare species are more informative about missing species. Uses the number of singletons and doubletons to estimate richness.

Faith's Phylogenetic Diversity:

- The sum of the length of all branches of a phylogenetic tree containing all species within a sample (or a group of samples).

Effective number of species (Hill numbers):

- Sometimes referred to as the true diversity of a sample. Equal to the number of species should all species be perfectly even. Can be calculated from many other diversity indices, including Shannon's H and Simpson's D.

Beta diversity

The spatial variation of diversity. Cannot be defined outside of the concept of diversity across space. True beta diversity is calculated as the ratio between gamma diversity and alpha diversity, where gamma diversity is the mean number of species per

sample. Beta diversity is also sometimes calculated as the total variance in a species composition table.

Distance metrics may be used to help describe beta diversity, but are not measures of beta diversity themselves. They are used in conjunction with sample clustering and ordination for analysis. Not all distance metrics are considered valid descriptors of beta diversity.

Euclidean

- The distance between points in Euclidean (geometric) space. This is distance as you normally think of it.

Aitchison

- The distance between points in simplex space.

Bray-Curtis (dissimilarity)

- Quantifies the differences in species populations between two sites. Bounded from 0 to 1, with 0 indicating that two sites are identical and 1 indicating that two sites share no species.

Jaccard index

- Also called the Jaccard similarity coefficient. Calculated as the intersection over the union of two samples (shared species over total species). Jaccard distance can be calculated by subtracting the Jaccard index from one.

Sorensen

- Similar to Jaccard. Gives greater weight to more common species.

*UniFrac (***uni**que **frac**tion metric)

- Incorporates phylogenetic distances to account for relative relatedness of communities.
- The unweighted version is calculated as the sum of unshared branch lengths between samples over the sum of all branch lengths. The weighted version also takes into account species abundance.

Gamma diversity

Regional diversity. The species diversity across all samples in a sequencing study. Calculated as the mean number of species per sample.