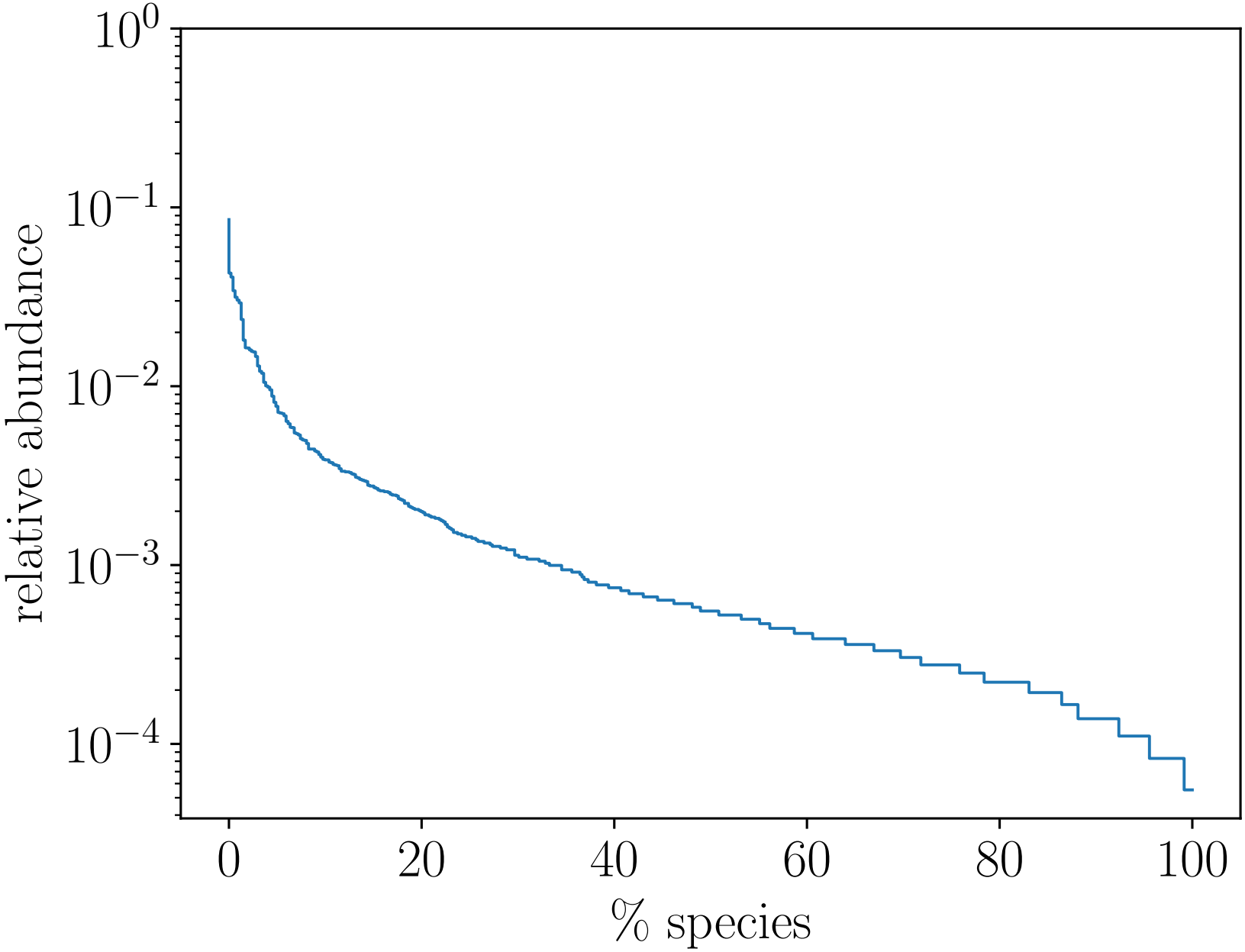
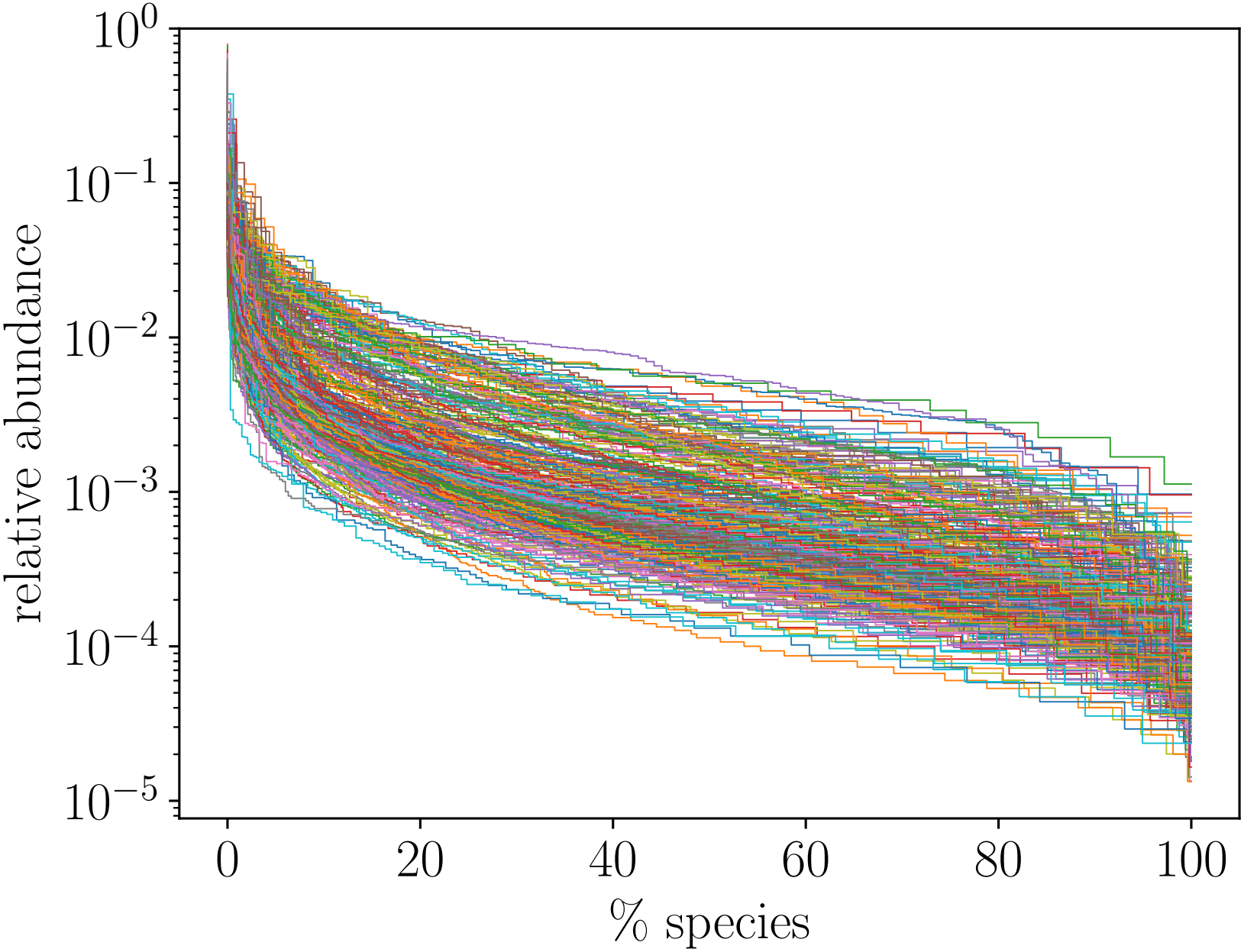
Initial results

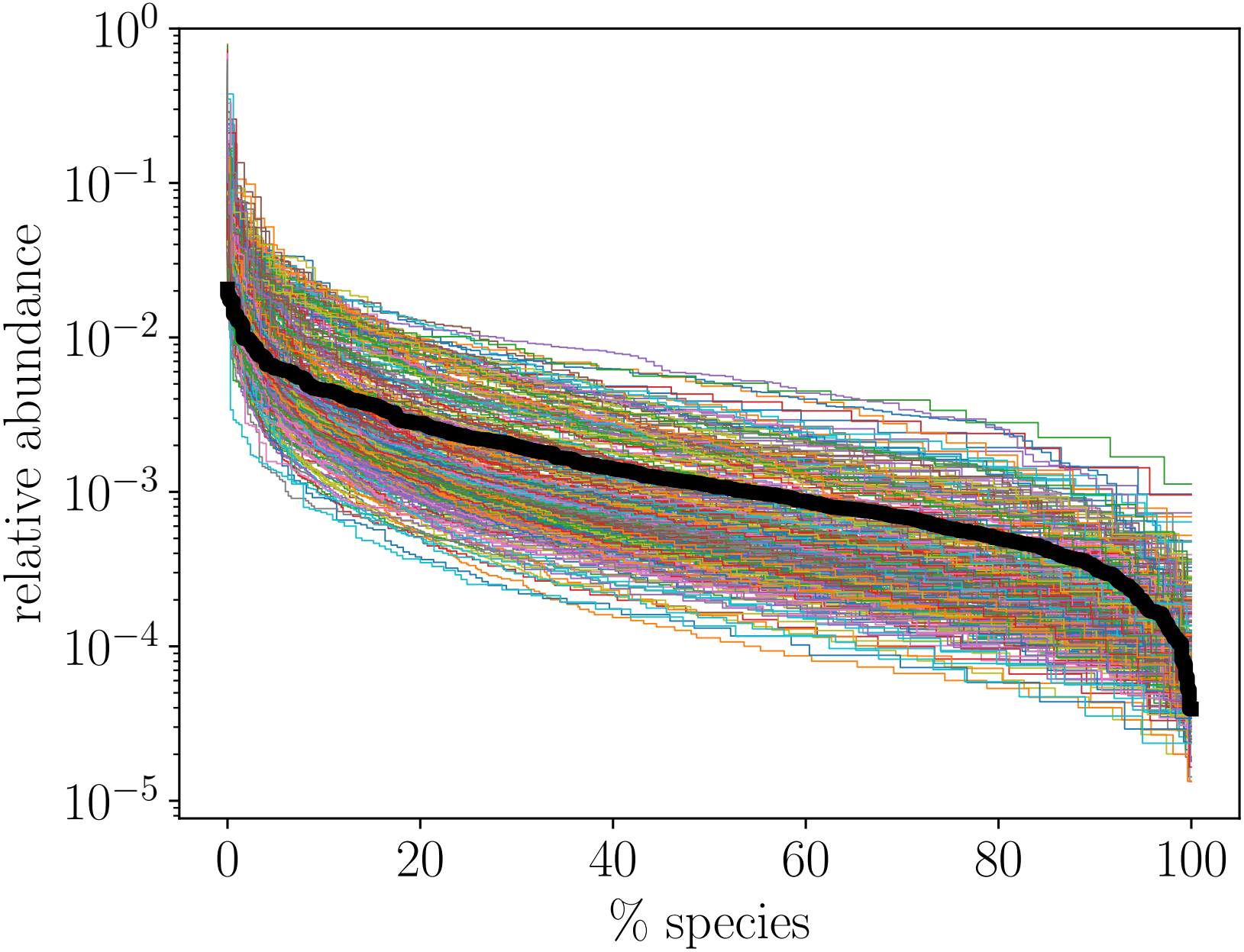
Species abundance distributions



Species abundance distributions



Species abundance distributions with lognormal curve



Shannon diversity

The Shannon diversity S is defined

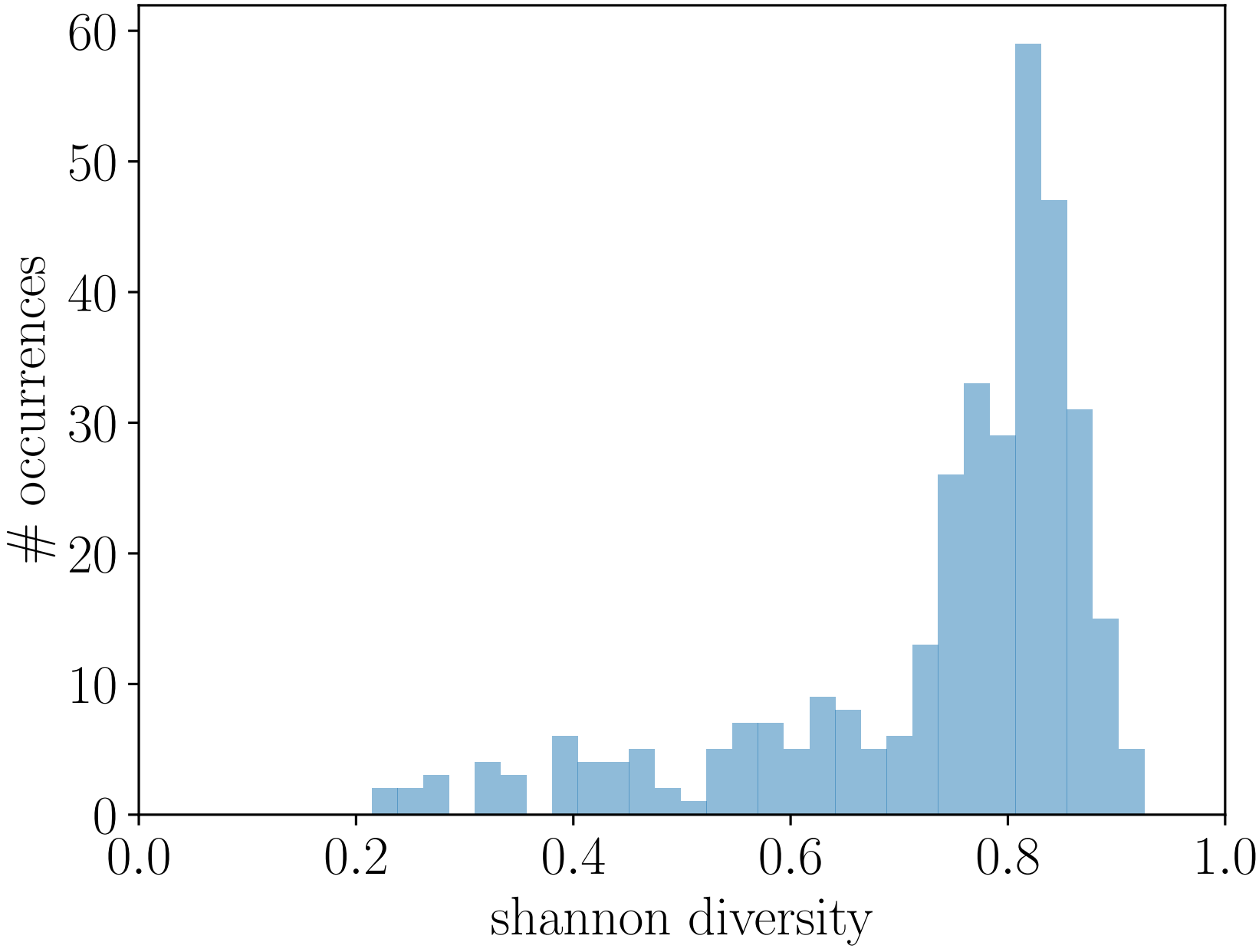
$$S = -\sum_{i=1}^{N} p_i \ln(p_i).$$

$\Longrightarrow$ maximum for $y_i = 1/N$ for all i

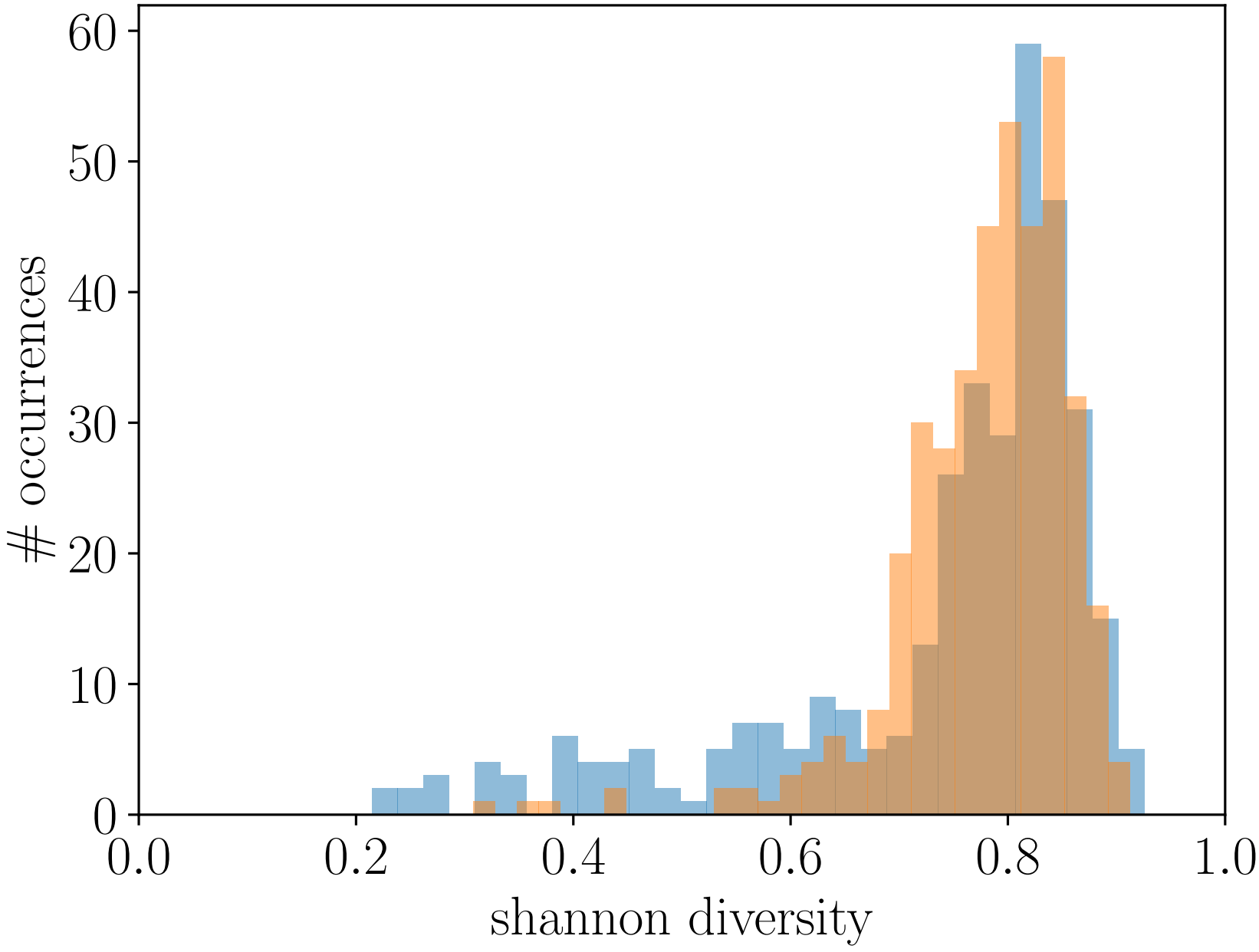
$\Longrightarrow$ minimum for $y_i = 1$, $y_j = 0$ for $j \neq i$

(one of many metrics for "diversity" of a population)

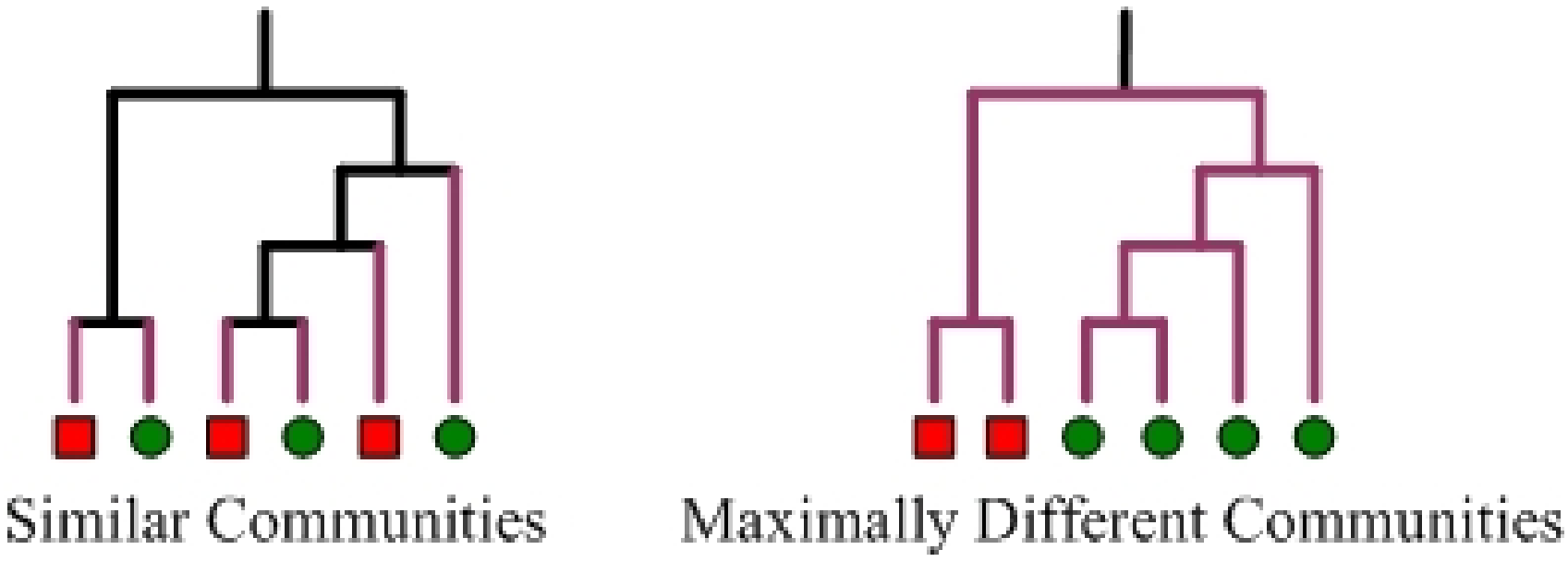
Shannon diversity



Shannon diversity with lognormal distribution

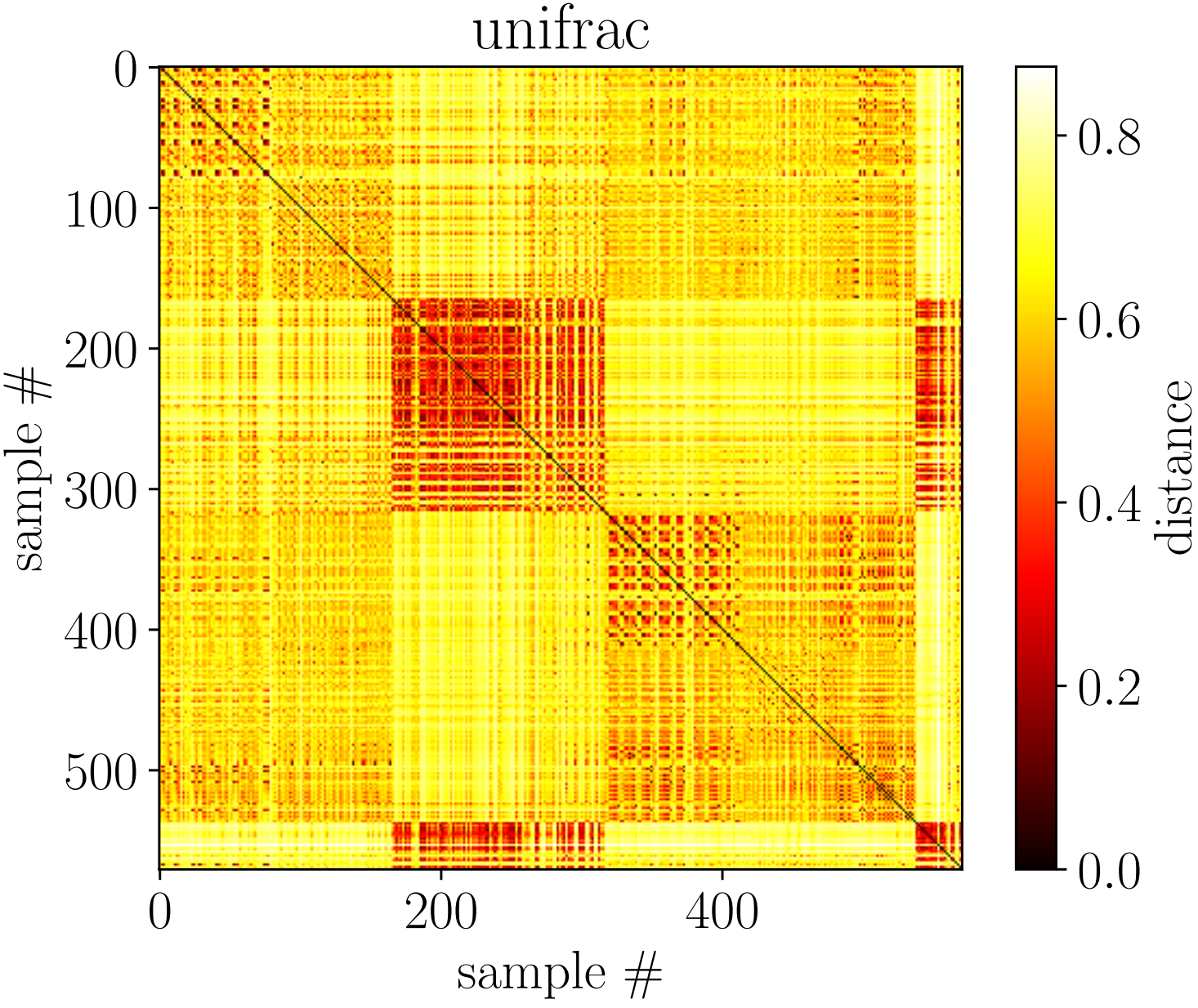


Unifrac[1]

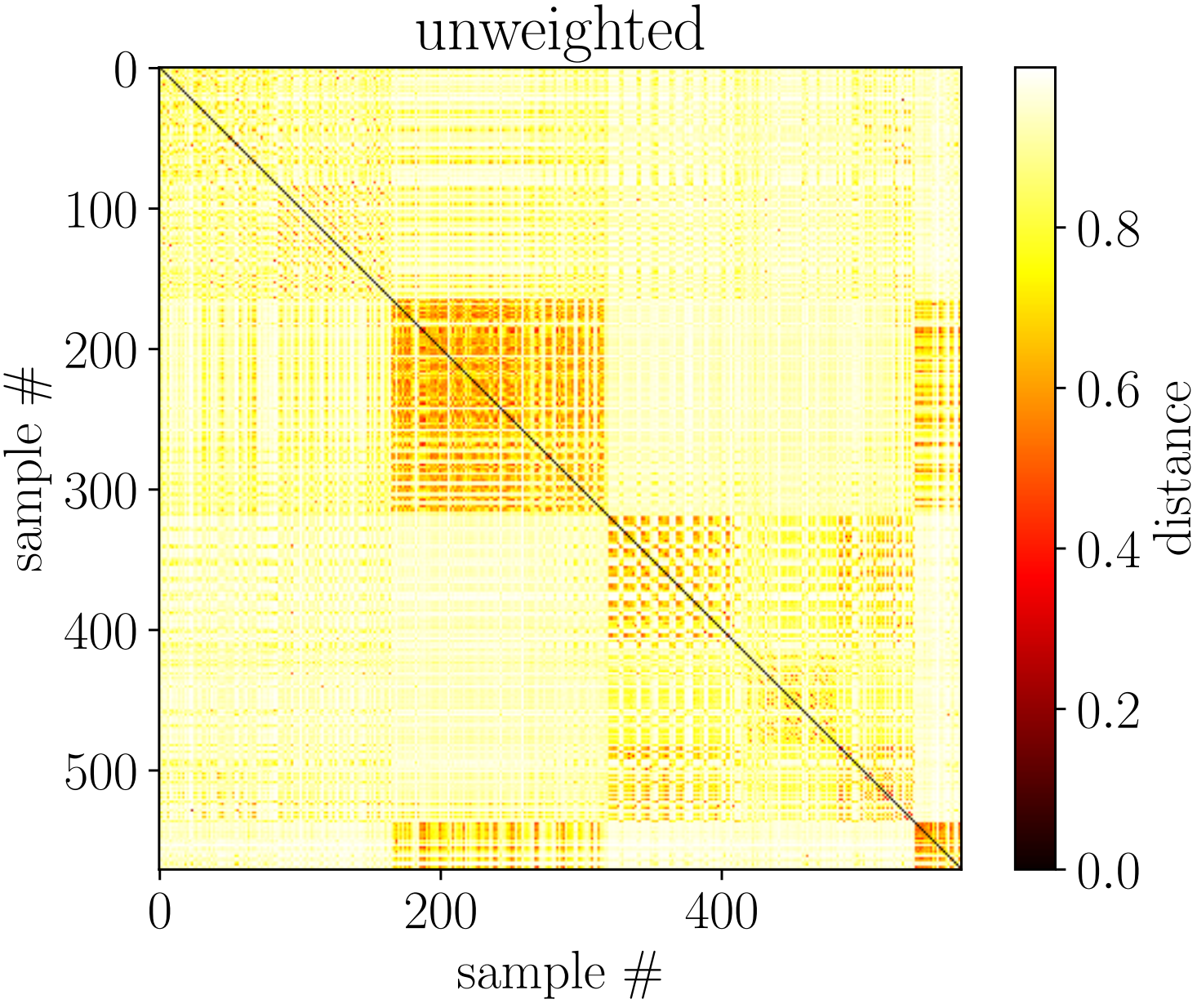


[1] image from mothur.org/w/images/5/5b/UnweightedUniFracMeasure.jpg

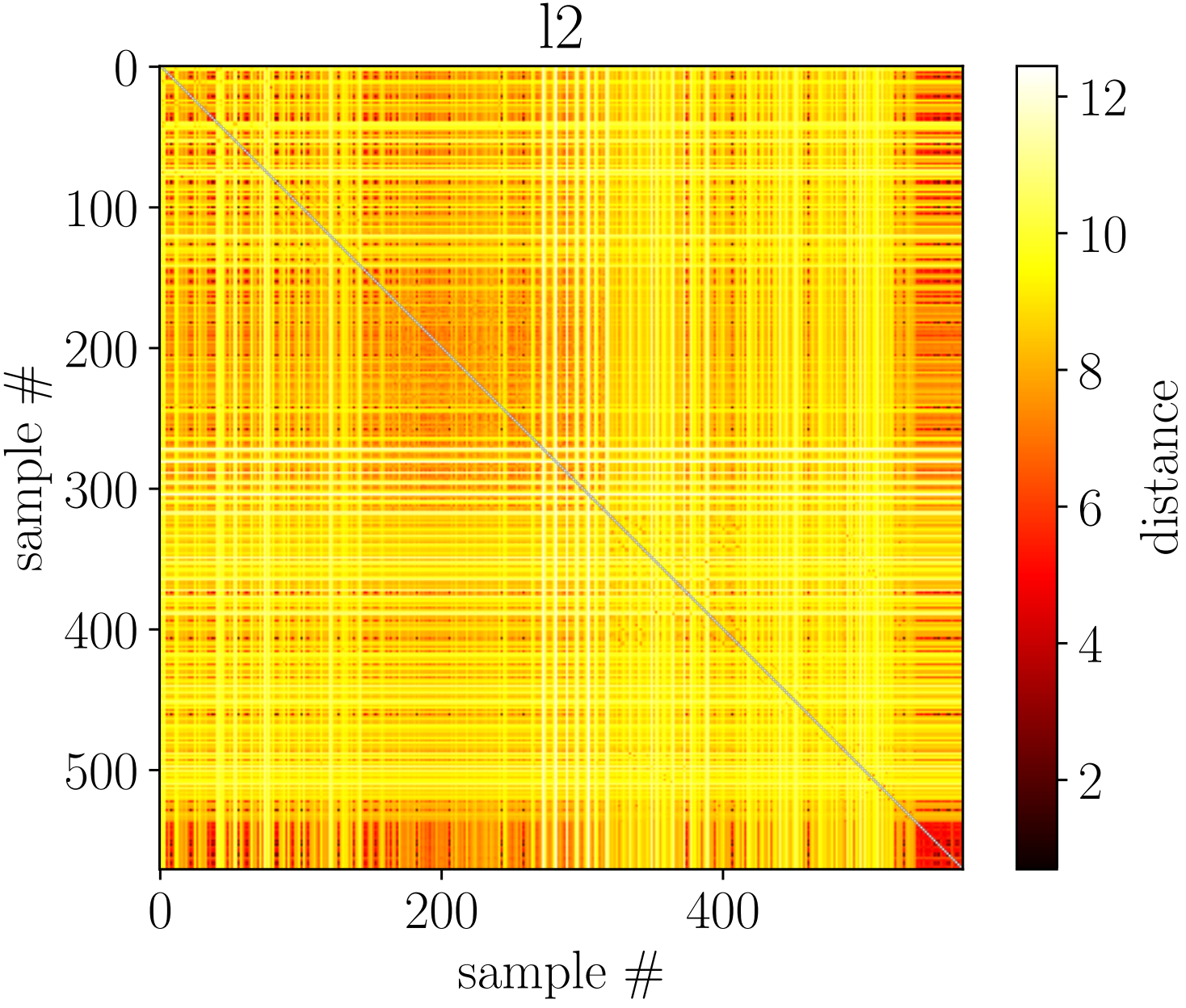
Unifrac distance matrix (weighted)



Unifrac distance matrix (unweighted)



L_2-norm distance matrix (weighted)



Future work?

- Estimate size of available microbiome pool (species abundance curve)
- Coarse-grain at different taxonomic resolutions
- Include metadata analyses
- How can we address nestedness?