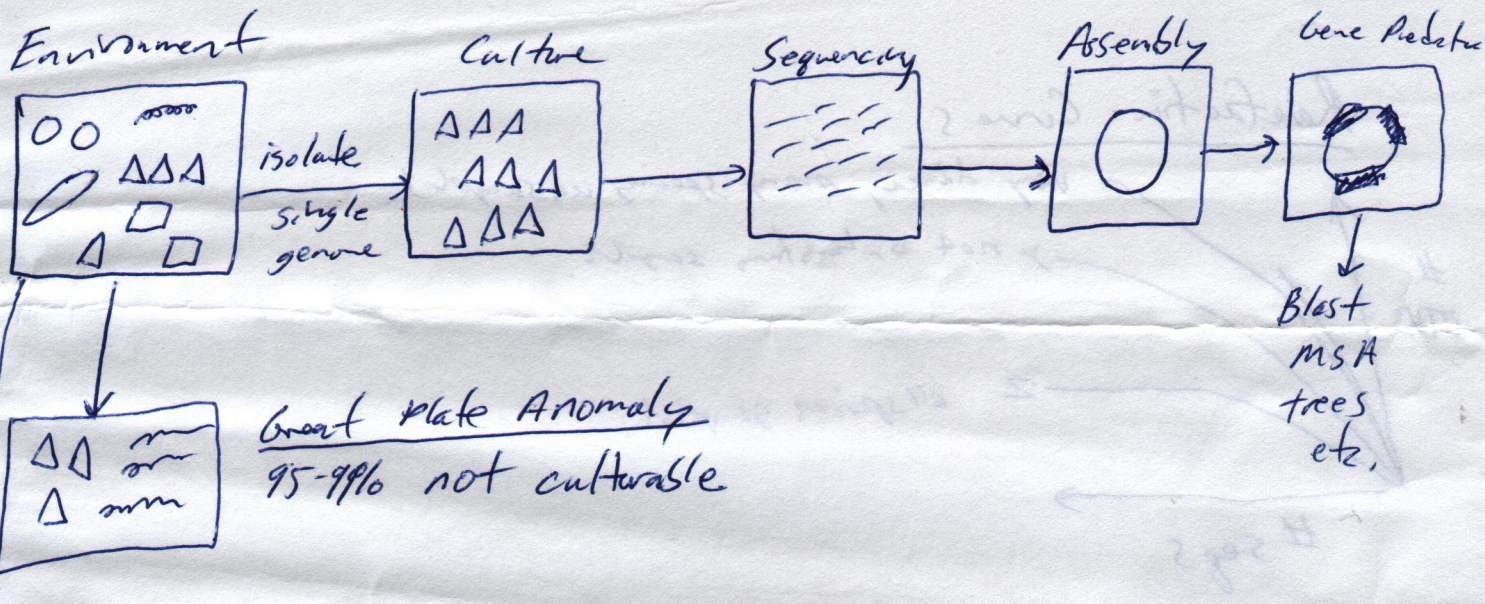
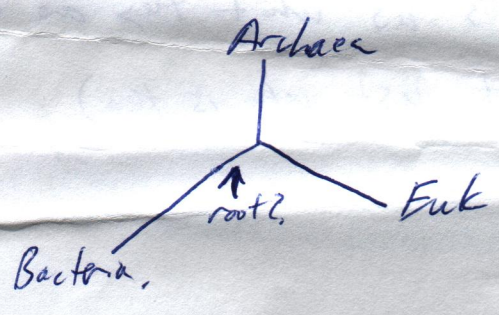
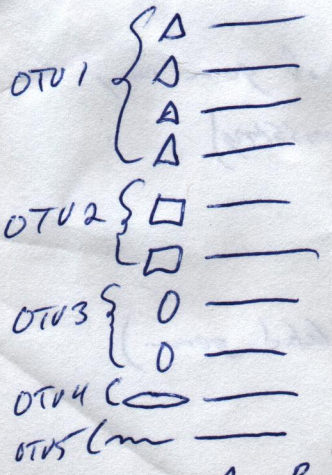
Environmental Sequencing + Metagenomics



① 16S ribosomal RNA (part of 30S subunit) in prokaryotes
18S " " 40S in eukaryote.



Sequencing 16S



	A	B	C
OTU1	4	1	2
OTU2	2	1	2
OTU3	2	2	2
OTU4	1	4	2
OTU5	1	2	2

Who is there?

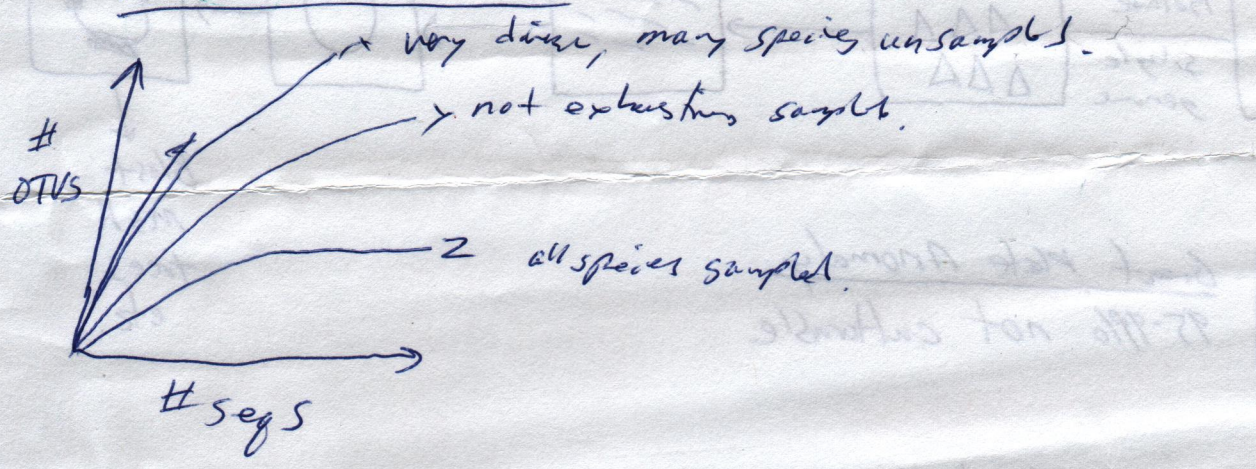
↳ assign taxonomy names using Blast (homology)
↳ names are problematic so cluster into
OTUs (Operational Taxonomic Units)
↳ simply cluster sequences at 97% ≈ species.

α-Diversity ⇒ diversity within single sample
↳ measures number of species and their relative abundance.

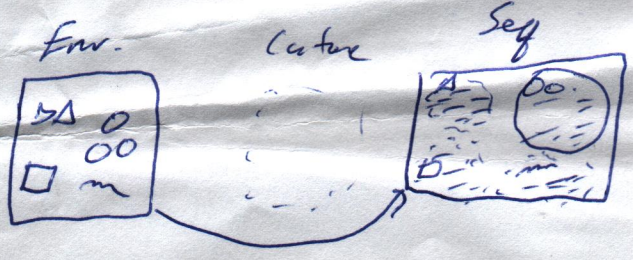
Shannon's Index $= -\sum_{i=1}^{S}\left[\left(\frac{n_i}{N}\right)\ln\left(\frac{n_i}{N}\right)\right] = -(0.4\cdot\ln(0.4)+0.2\ln(0.2)\ldots$
$= 1.47 = B \quad C = 1.6$

β-diversity = Bray-Curtis (similar to correlation)

Rarefaction Curves



Metagenomics



Tells us what they are doing (not just who is there)

Binning ⇒ trying to place reads into the genome they came from.

⇒ hard if ① ↑ diversity
② low sequence coverage (parts of genome may be missing)
③ LGT

Methods ① Assembly (requires deep seq $$$)
② Reform genome alignment (requires closely related genome)
③ Composition (k-mers; codon usage)
④ Phylogenetics (Lowest Common Ancestor)