https://eval.ucdavis.edu

Please take 10 minutes to fill out the online course evaluation. Professor Maloof will be back shortly.

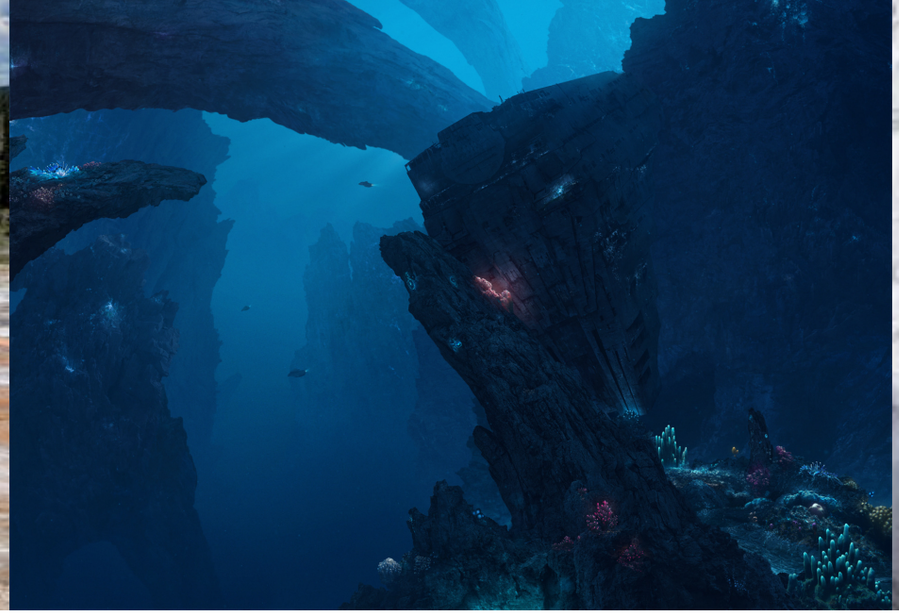
Getting Messy with Microbes:

An Introduction to Metagenomics

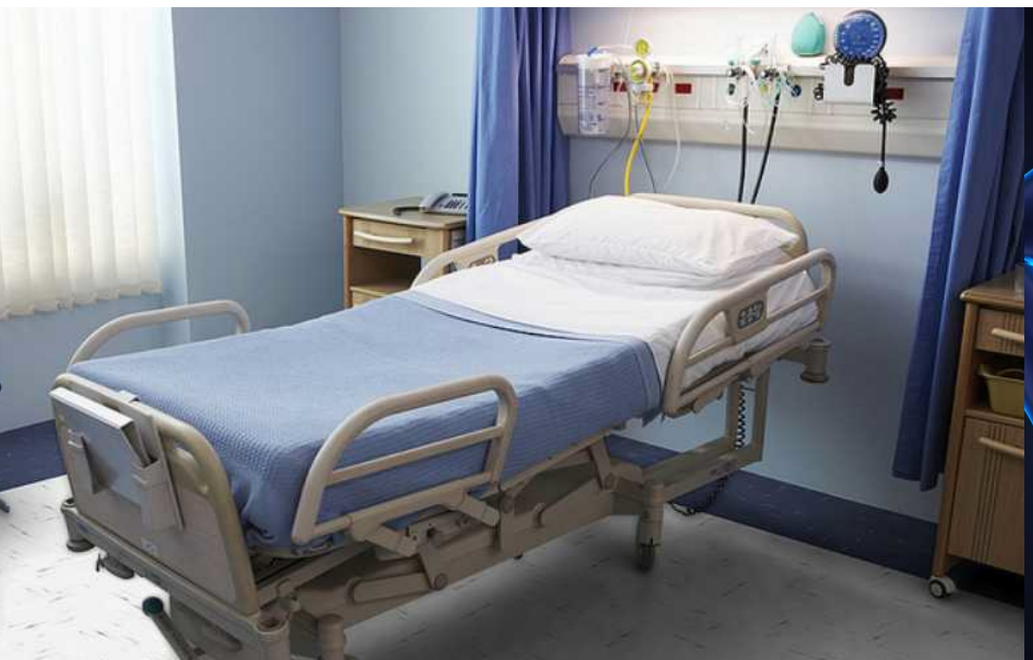
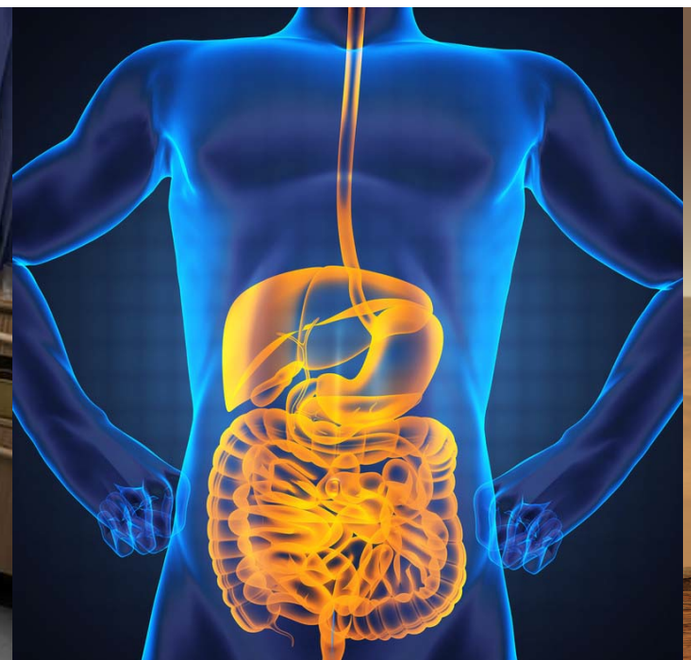
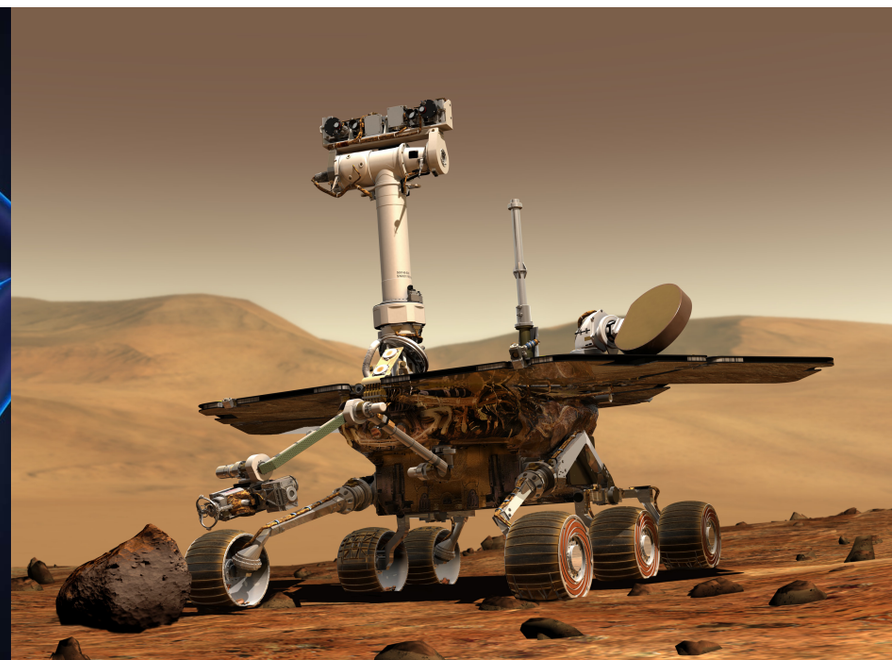
Slides from Kristen Beck, modified by Julin Maloof

Metagenomics.

The study of a collection of genetic material from a mixed community of organisms, typically microbial.


Metagenomics is a lens to understand life in countless environments.

Today's lab focuses on the rice root microbiome

- Plant health and nutrient acquisition
- Methane emission
  - Rice paddies produce 15-20% of global methane emission due to root-associated Archaebacteria
  - There are also methane-utilizing bacteria that could mitigate this problem
- Drought resistance
- Prescriptive innocula

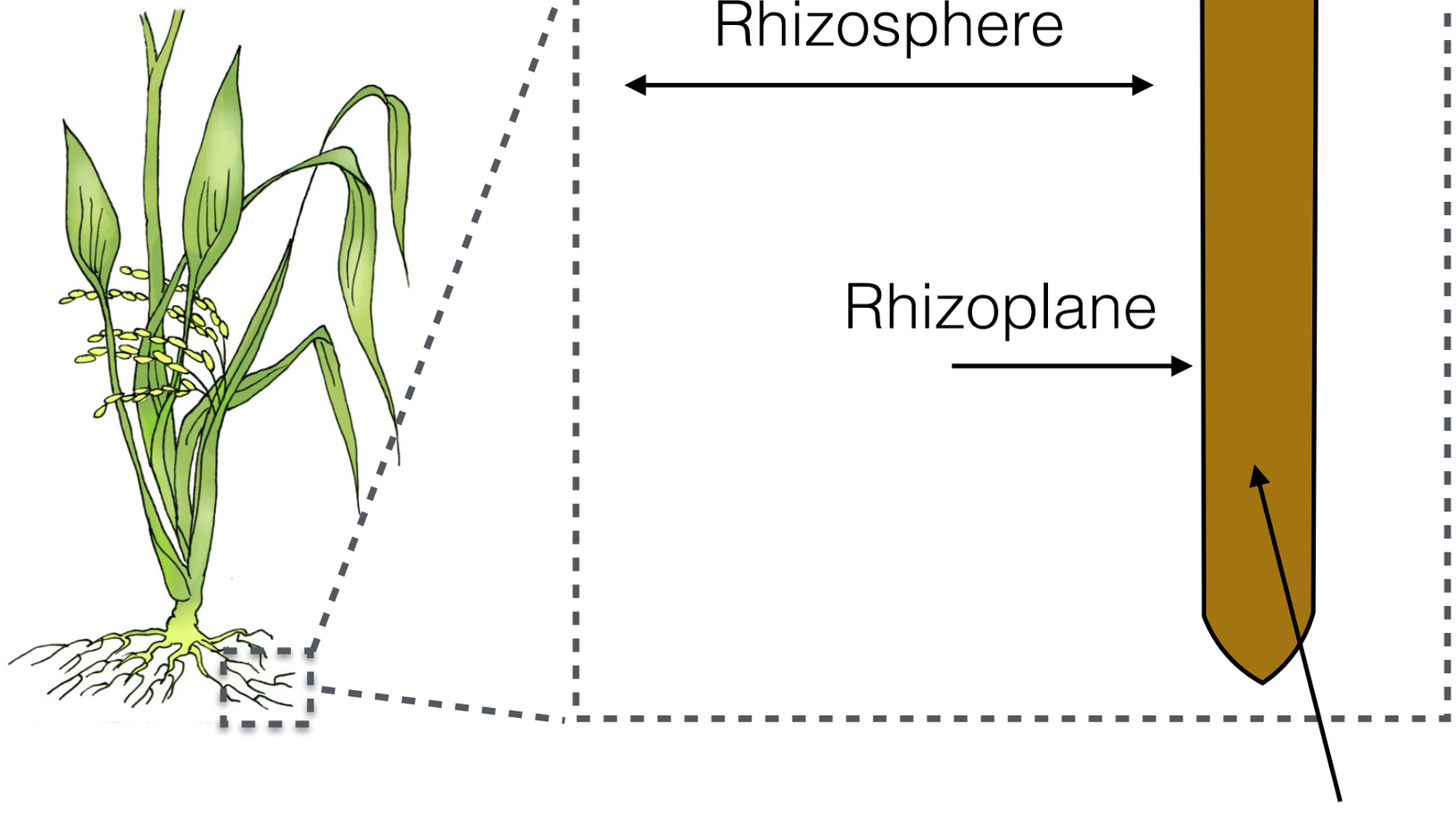
Photo credit: Sundaresan Lab, UC Davis

Today: how does the microbe community differ between cultivars and locations?

Nipponbare

M104

IR50



How do we assess microbial community composition?

Sequence the 16S ribosomal RNA

Sequencing of the 16S ribosomal RNA

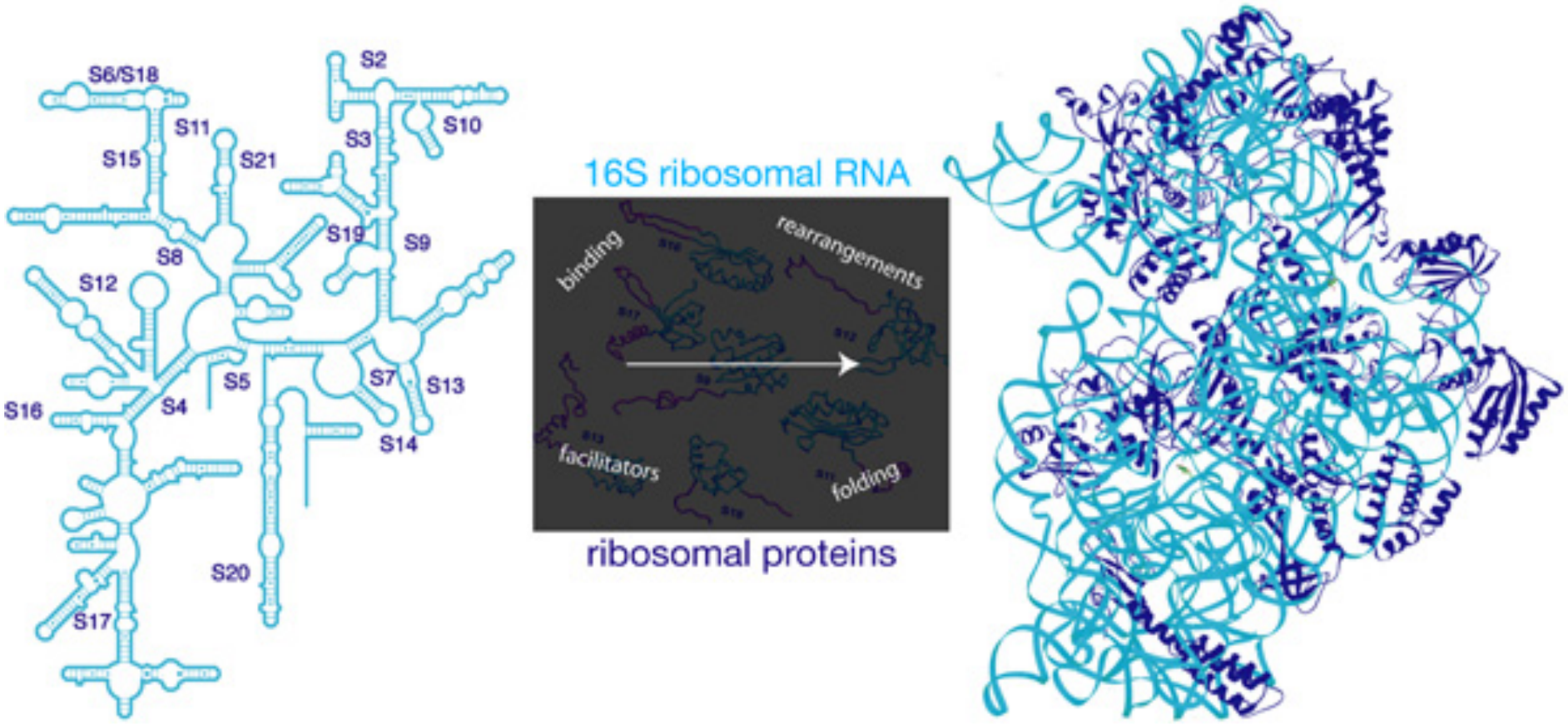


Photo credit: Culver Lab, Univ. of Rochester

What are the components of a metagenomics analysis?

Operational Taxonomic Unit (OTU)

A cluster of 16S rRNA reads with >97-99% similarity to define “species”

Questions that can be answered by 16S sequencing:

- Which microbes are in my sample? What might they be doing?
- Is microbe X present at different levels between samples (differential abundance)?
- Why might one phylum be more abundant in this environment?

Alpha Diversity

- Alpha diversity: how diverse is the microbial community?
- Can compare diversity between samples
    - e.g. is the microbial community more diverse at the root surface or in bulk soil?

Alpha Diversity Method 1

- **Count observed OTUs**
- Simple, but…any issues?
- (Illustrate on board)

Alpha Diversity Method 2

- **Shannon's Diversity**

- s = number of species

$$H = -\sum_{i=1}^{s} \left(p_i \log_2 p_i \right)$$

- p_i = proportion of counts attributable to species i

- equal representation leads to high diversity index (play with it in R)

Alpha Diversity Method 3

- **Chao1**

$$chao1 = S_{obs} + \frac{F_1^2}{2F_2}$$

- Adjust for unobserved samples based on observed singletons

- *Sobs* = total observed species

- F_1 = number of singletons (species observed only once in a sample)

- F_2 = number of doubletons

Questions that can be addressed with alpha diversity:

How many taxa are in a sample? What is the richness of my sample?

Have I sequenced to a depth (coverage) that describes the diversity of my sample?

Does condition X have higher phylogenetic diversity than condition Y?

Beta Diversity

- Beta diversity measures the diversity *between* samples.

- The distance between each pair of samples (with respect to community composition) is calculated.

- There are many ways to calculate beta diversity; we will use one: Bray-Curtis

Bray-Curtis Dissimilarity

- For two samples, *i* and *j* :

$$\frac{\text{unique species counts}}{\text{total counts}} \qquad BC_{ij} = \frac{S_i + S_j - 2 * C_{ij}}{S_i + S_j}$$

- S_i and S_j = sum of counts in samples *i* and *j*
- C_{ij} = sum of min(counts) for taxa observed in both *i* and *j*
- What is Bray-Curtis if there are no taxa in common?
- What is Bray-Curtis if the two samples are exactly the same?

Bray-Curtis Dissimilarity

$$BC_{ij} = \frac{S_i + S_j - 2 * C_{ij}}{S_i + S_j}$$

	Sample 1	Sample 2
Species A	5	0
Species B	4	8
Species C	7	3
Species D	10	5

- $S_i = 26; S_j = 16$
- $C_{ij} = 0 + 4 + 3 + 5 = 12$
- $BC_{ij} = \frac{26 + 16 - 2 * 12}{26 + 16} = 0.43$

Beta Diversity Questions

- Do samples contain different microbial communities?
- Which microbial taxa have increased abundance compared to another sample?
- Are there broad trends that relate many samples? Can these trends be explained by an environmental or genetic factor?

Other questions that can be addressed (but not in this lab)

- Given the observed microbial community, what services, biochemical/metabolic/biological functions might they be providing?

- Also can sequence all DNA (or mRNA) instead of 16S to get a better idea of functions present in the community.