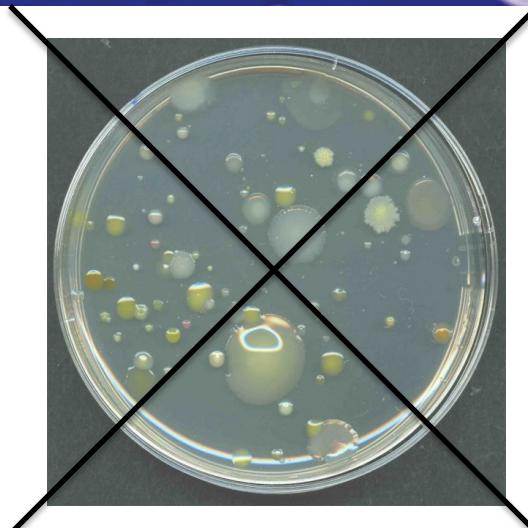
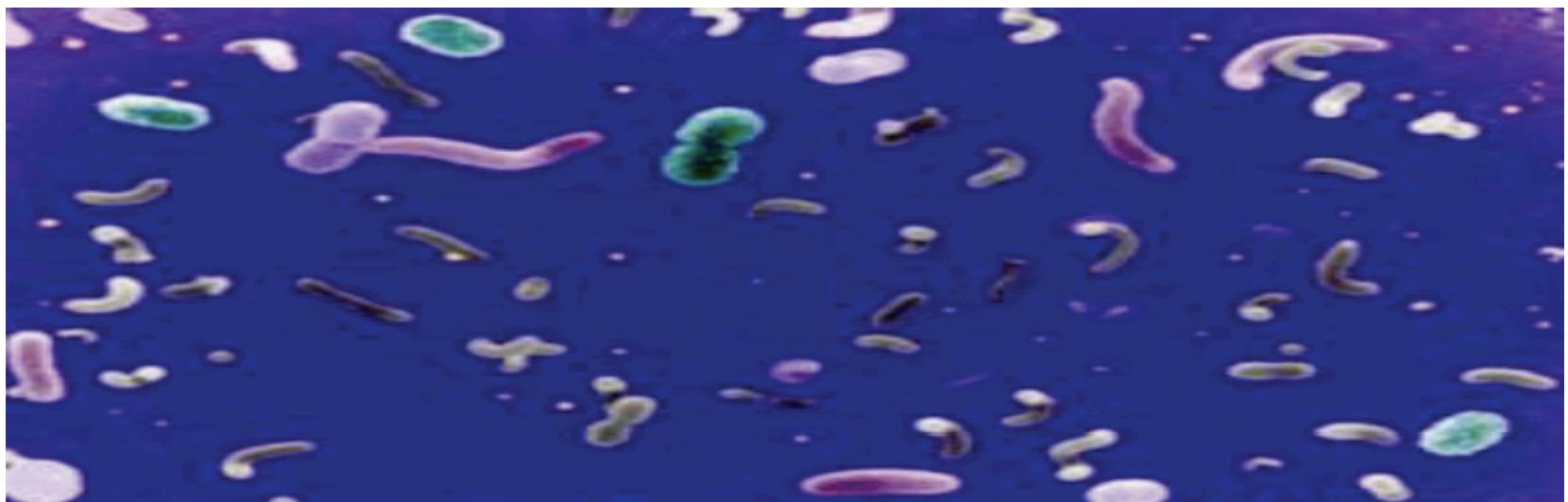
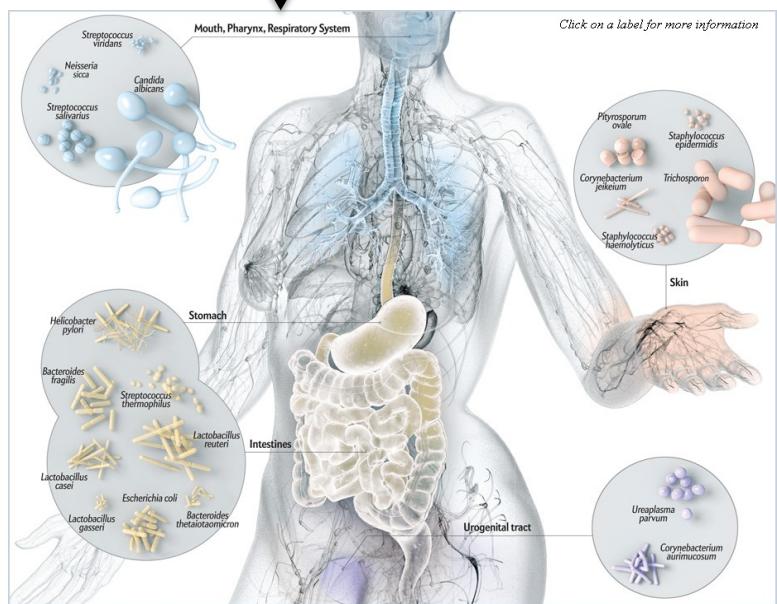
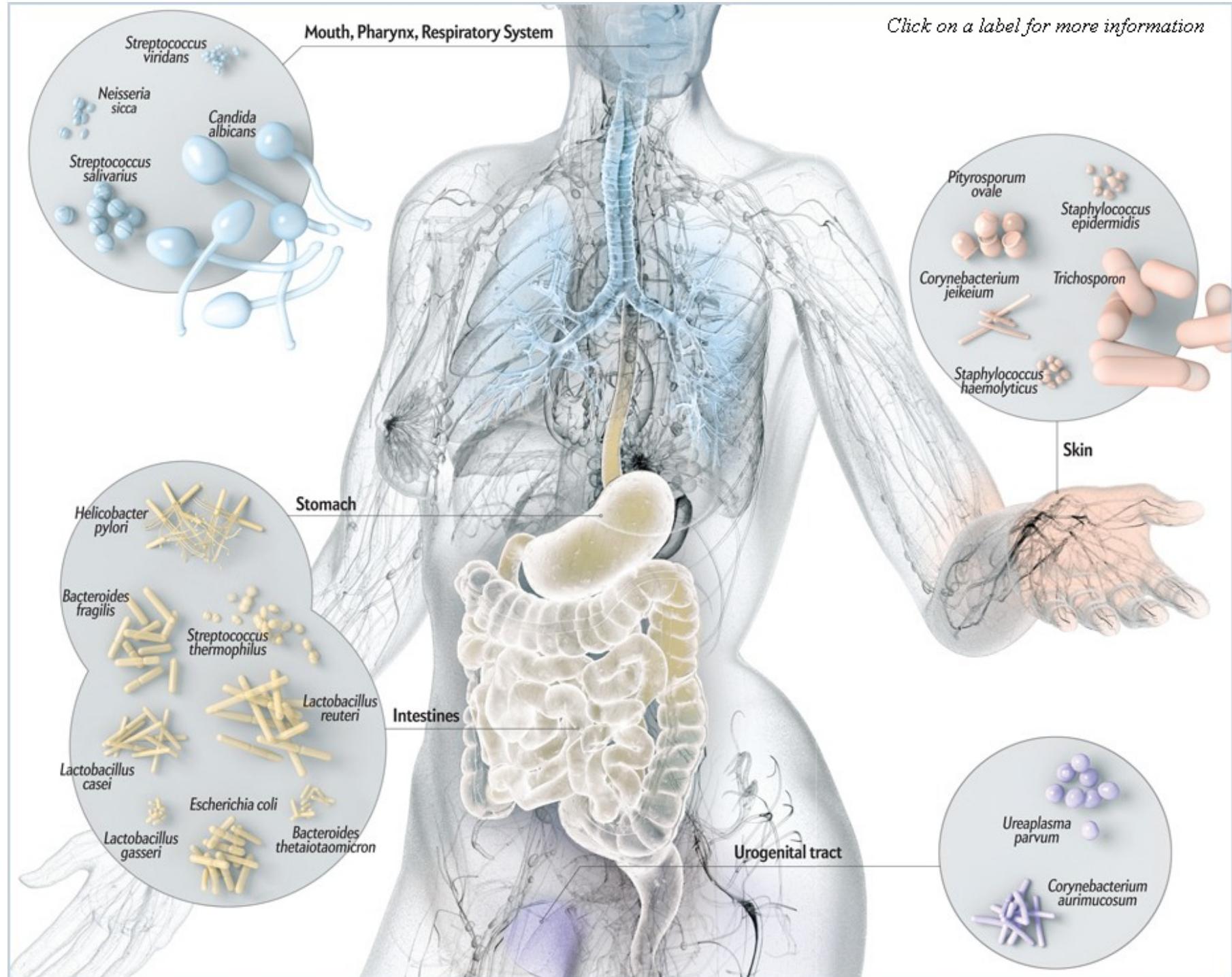


Computation of Environmental Samples (a.k.a. Metagenomic Data)



Microbes are everywhere

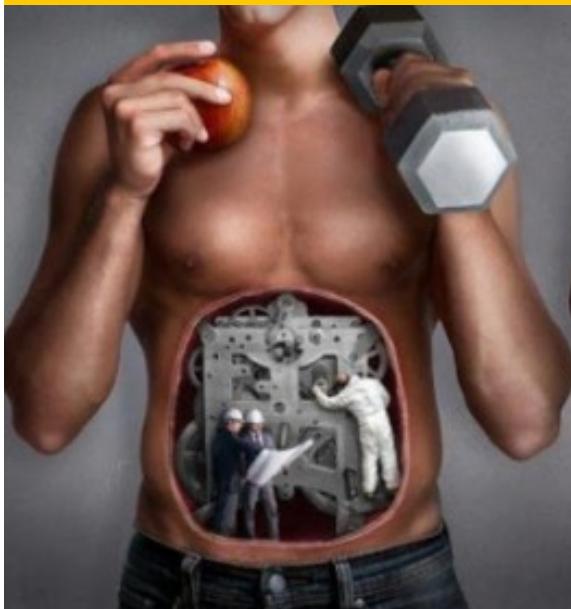
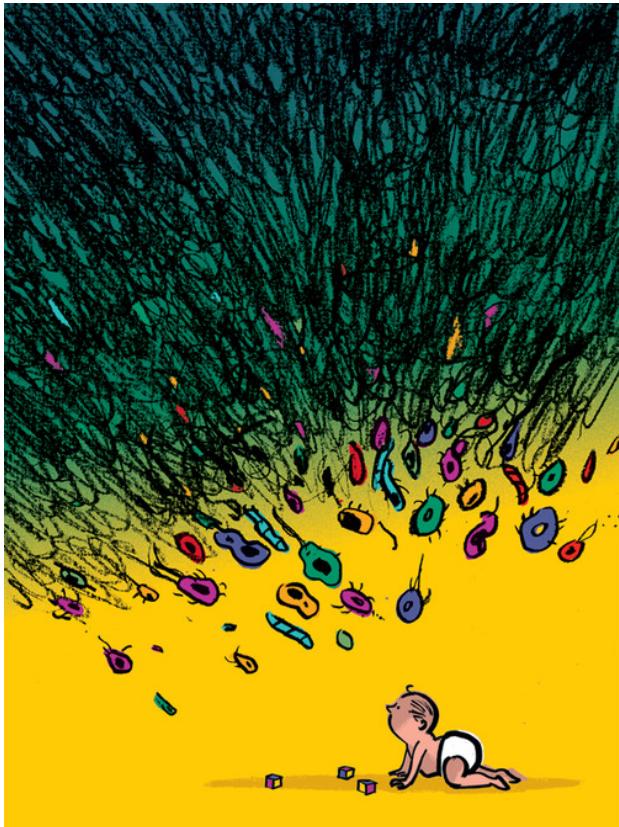




Health

Diversity protects

Are we losing it?





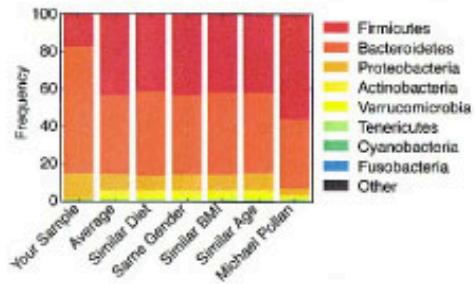
Crowdfunding Science



YOUR AMERICAN GUT SAMPLE

ME

What's in your American Gut sample?



Your most abundant microbes:

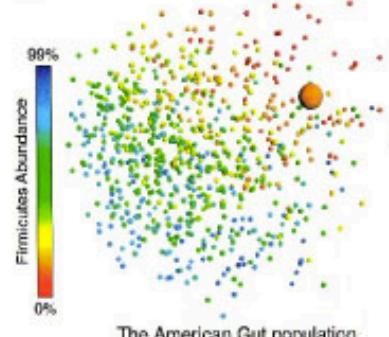
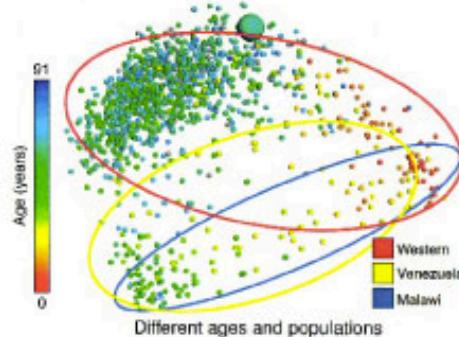
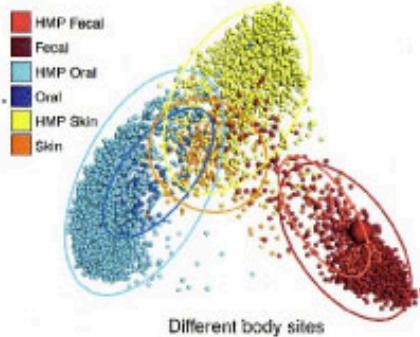
Taxonomy	Sample
Genus <i>Bacteroides</i>	64.9%
Family Enterobacteriaceae	7.8%
Family Lachnospiraceae	4.4%
Family Ruminococcaceae	4.4%

Your most enriched microbes:

Taxonomy	Sample	Population	Fold
Genus <i>Bacteroides</i>	64.9%	32.13%	2x
Family Enterobacteriaceae	7.8%	2.88%	3x
Genus <i>Holdemania</i>	0.1%	0.03%	3x
Family Barnesiellaceae	1.7%	0.64%	3x

Your sample contained the following rare taxa: Genus *Fusobacterium*, Genus *Citrobacter*, Genus *Plesiomonas*, Genus *Acinetobacter*.

How do your gut microbes compare to others?

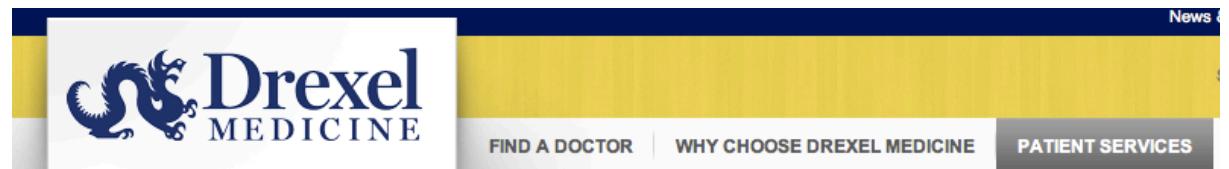


What are two systems at our “interfaces”?

- Microbiome
- Immune System

Causality?

- Fecal Transplants



Home > Patient Services > Gastroenterology > Services > Fecal Microbiota Transplant

Gastroenterology

Our Doctors

Meet a Doctor

Services

Anorectal Manometry

Balloon Enteroscopy

Capsule Endoscopy

Cholangioscopy

Colonoscopy

Endoscopic Ultrasound

Fecal Microbiota Transplant

Impedance/pH Monitoring

Conditions Treated

Preparing for Your Visit

Fecal Microbiota Transplant

Fecal microbiota transplant, or FMT, takes stool from a healthy donor and transplants it into the digestive tract of a patient suffering from recurrent *Clostridium difficile* infection, or RCDI. Patients who have had three or more relapses of mild to moderate RCDI that have not responded to antibiotics are eligible to receive FMT at Drexel Gastroenterology.

How Does FMT Work?

Patients with RCDI have too much of a bad bacteria (*C. diff*) in their digestive tracts, in particular in the large intestine. This often results from taking antibiotics that are strong enough to kill the good bacteria but not the *C. diff*. With no competition, the *C. diff* takes over, causing RCDI, which is characterized by:

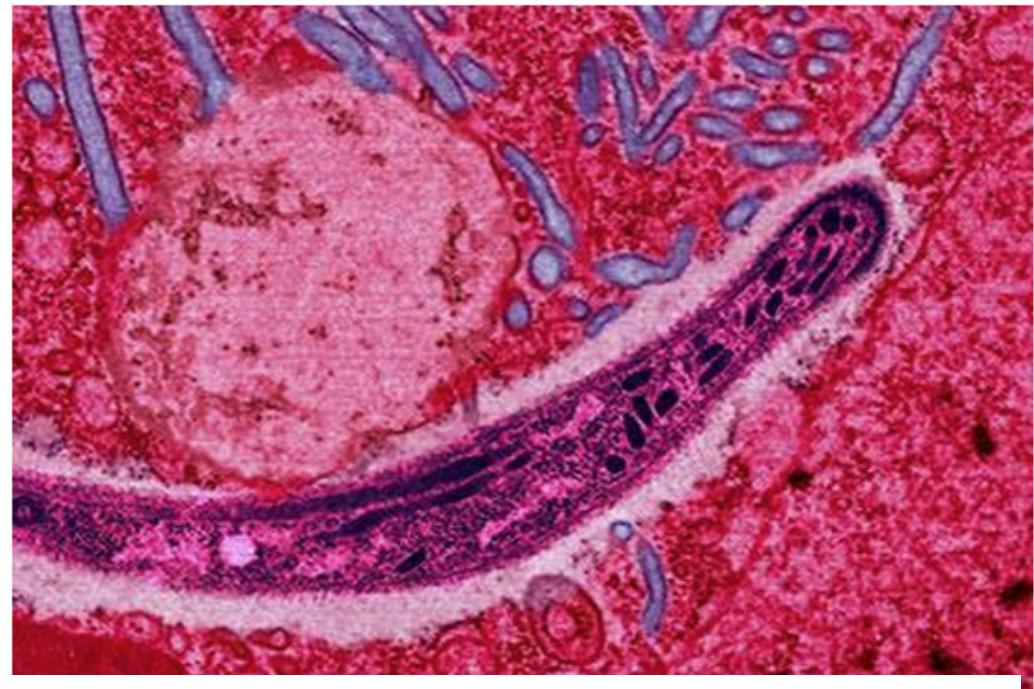


Engineering Therapies

- Rewiring 100 B. thetaiotaomicron circuits to detect “seeing” different bacteria/molecules
- Secrete Anti-inflammatory molecules when see Inflammation
- Optimize Processes by getting the “Right mix” of bacteria -- Biofuels

Engineering mosquito gut bacteria to fight malaria

By Ed Yong | July 16, 2012 3:00 pm



Engineering the Human Microbiome Shows Promise for Treating Disease

Synthetic biology may lead to the creation of smart microbes that can detect and treat disease

Innovations®

Feb 17, 2015 | By Justin L. Sonnenburg

Earth Microbiome Project



Microbiome and Plant Yield

Soil microbiomes can set plant flowering time

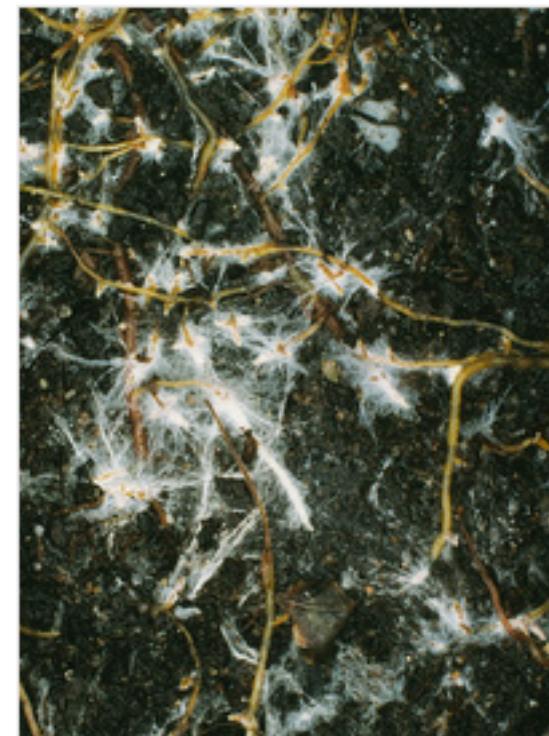
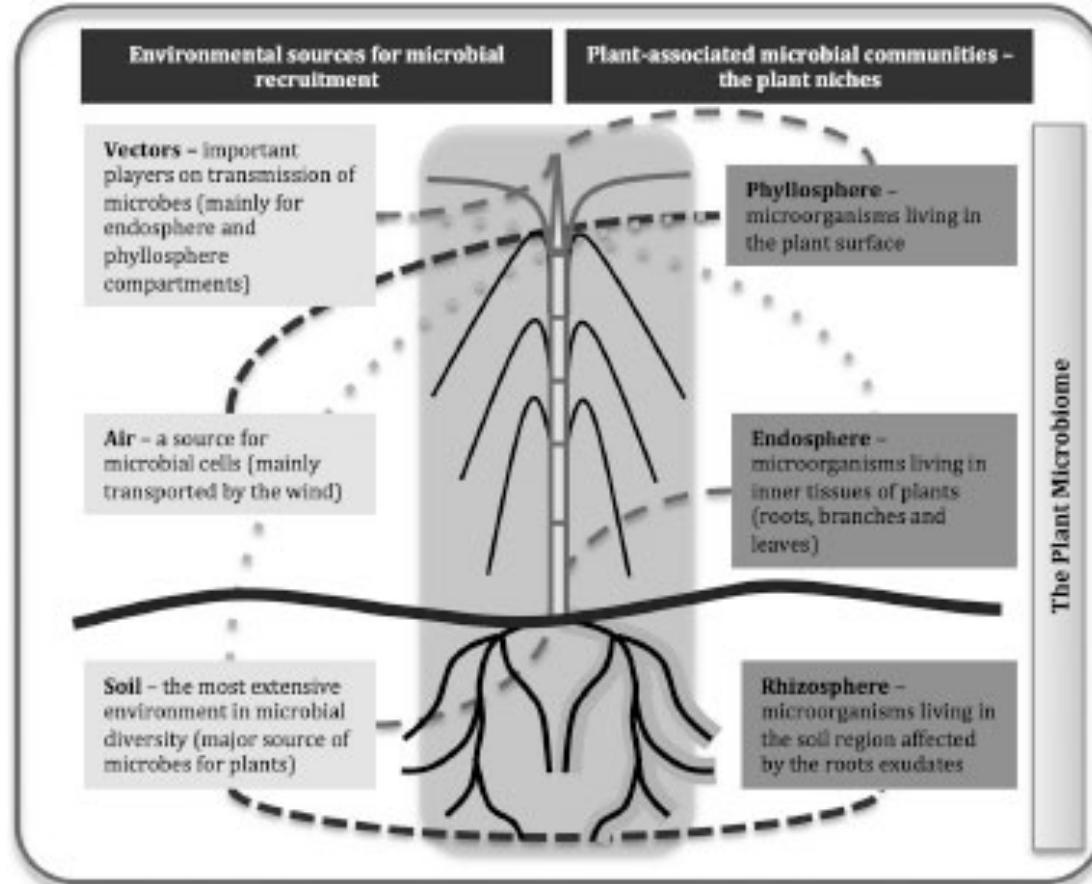
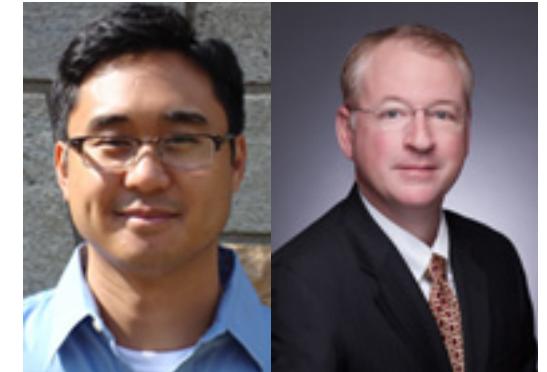


Figure 1 – Schematic representation of the major sources for microbes that compose the plant-associated communities: the rhizosphere, endosphere and phyllosphere. Thickness and fill of connections indicate the contribution of environmental sources for the composition of microbial communities in plant-harboring niches.

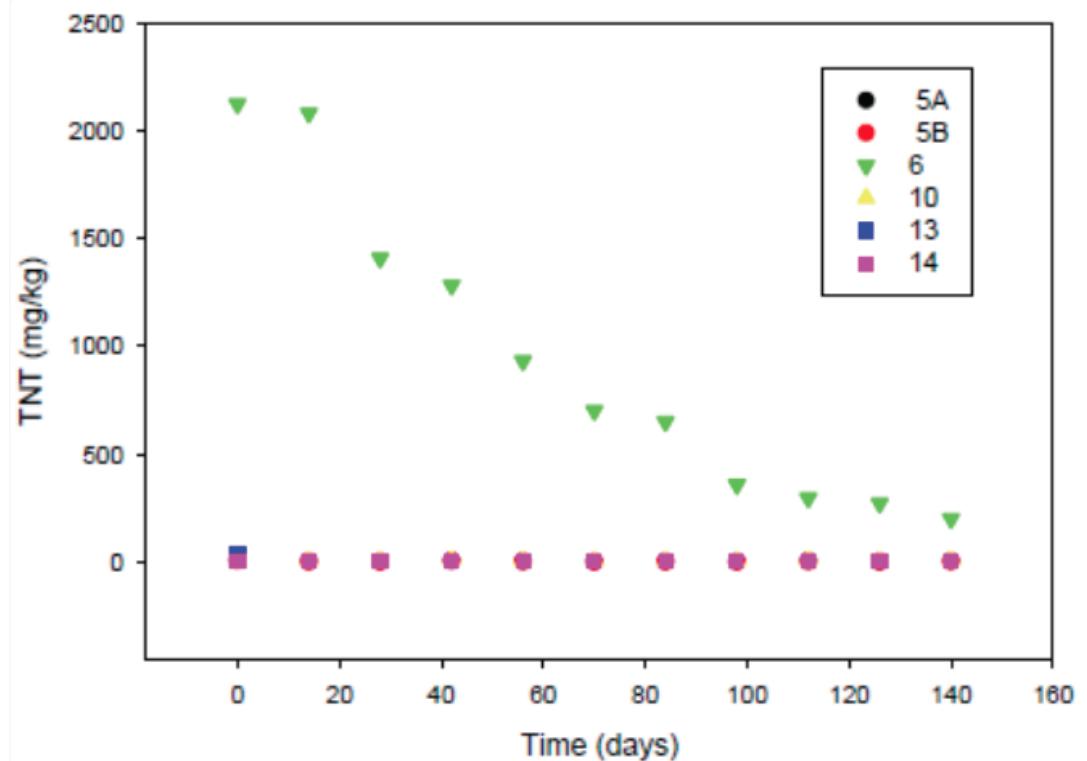
JEREMY BURGESS/SCIENCE
PHOTO LIBRARY

Mycorrhizal fungi (white) help plant roots (brown) absorb water and nutrients from the soil.

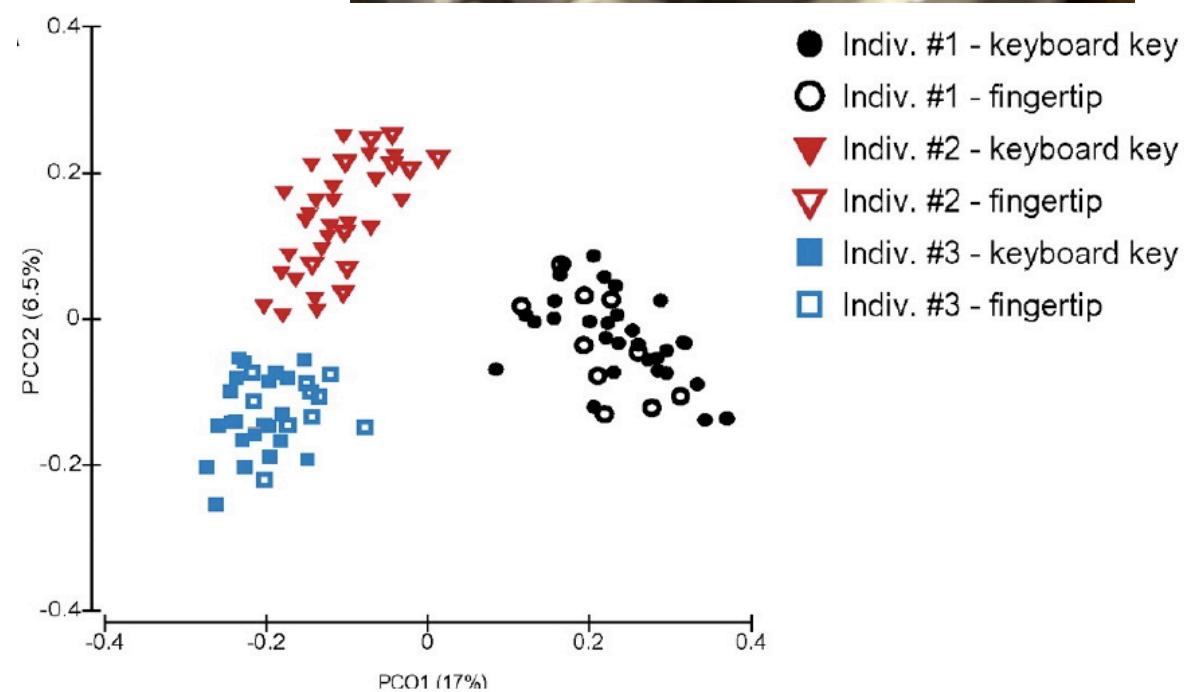
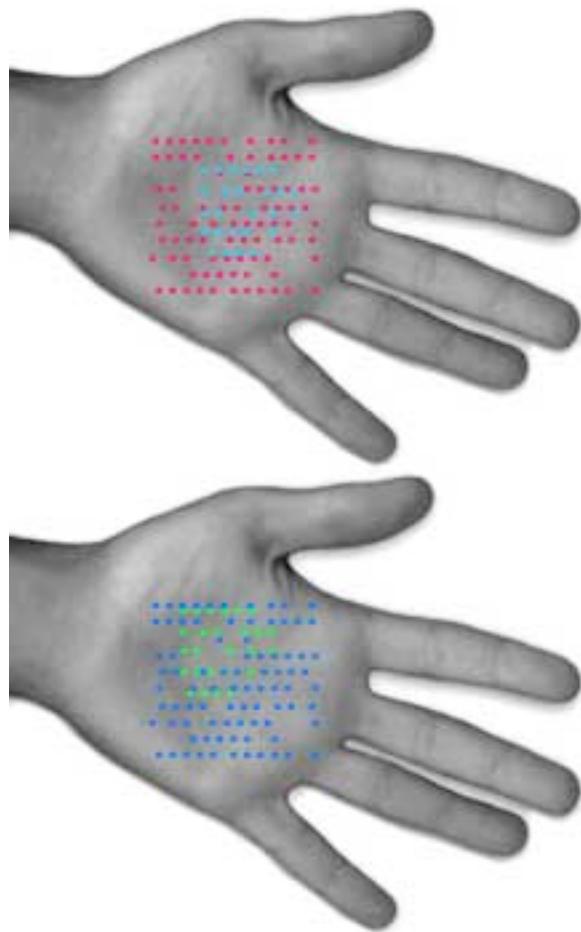
TNT Bioremediation



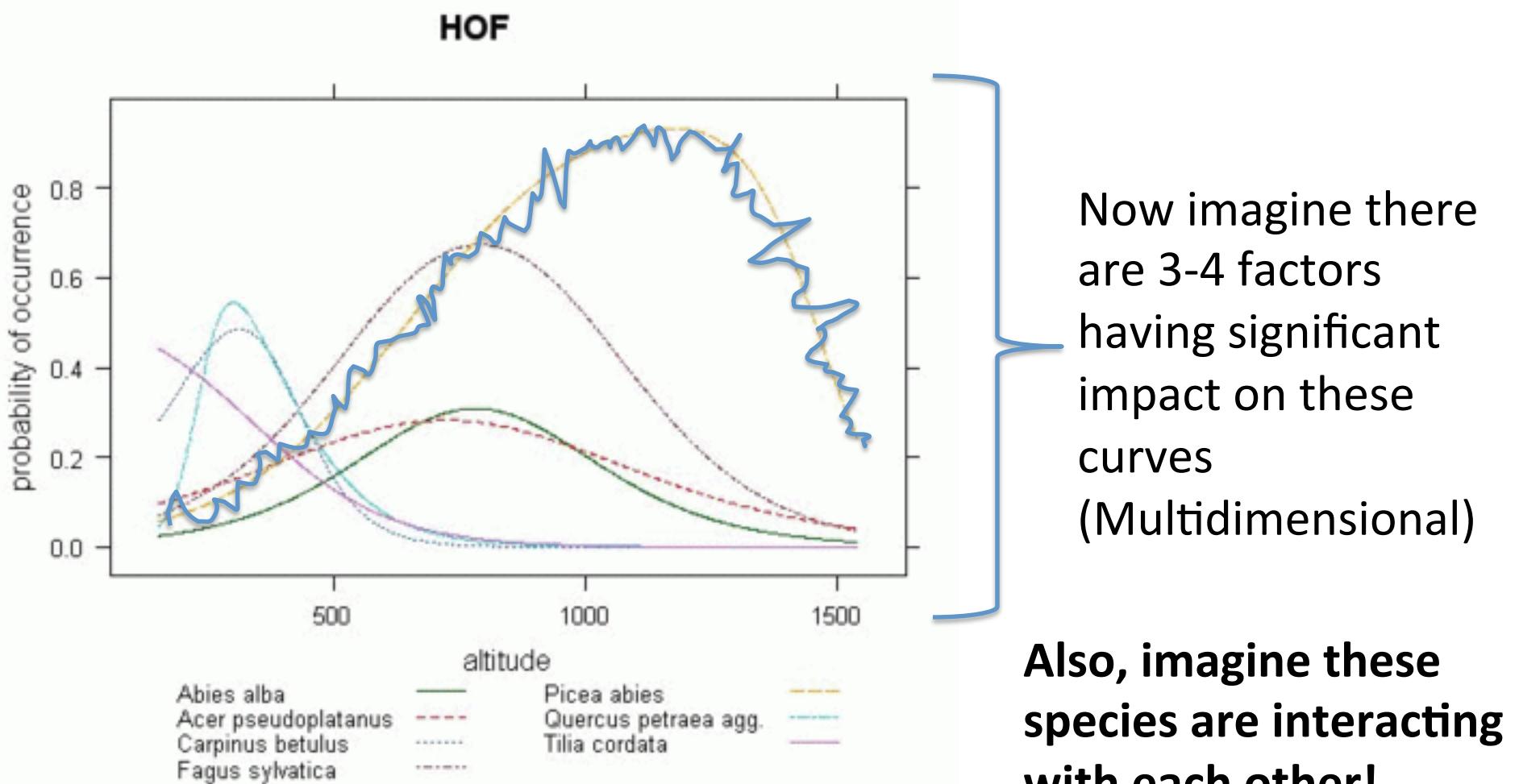
Tilling accelerates
TNT degradation



Example Application: Forensics



Underlying Assumptions of Data: Ecological Theory



Different Types of Factors

Soil environment and physical factors controlling microbial activity

From MicrobeWiki, the student-edited microbiology resource

This is a curated page. Report corrections to [Microbewiki](#).

Contents [hide]

- 1 Introduction
- 2 Chemical Factors
 - 2.1 pH
 - 2.2 Oxygen
 - 2.3 Cation Exchange Capacity (CEC)
- 3 Physical Factors
 - 3.1 Soil Texture
 - 3.2 Soil pores
 - 3.3 Soil Structure
 - 3.4 Soil water
 - 3.5 Temperature
 - 3.6 Soil aggregates
- 4 Biological Factors
 - 4.1 Soil Fauna
 - 4.2 Organism Interactions
 - 4.3 Bioavailability
 - 4.4 Plant Growth-Promoting Rhizobacteria (PGPR)
- 5 Relevant Organisms
- 6 Current Research
- 7 References



[15]

These Factors are in the body too...
but probably way more susceptible to
biotic or host-influenced environment

Our Observations

Samples

Ecological data matrix.								
<i>Objects</i>	<i>Descriptors</i>							
	\mathbf{y}_1	\mathbf{y}_2	\mathbf{y}_3	...	\mathbf{y}_j	...	\mathbf{y}_p	
\mathbf{x}_1	y_{11}	y_{12}	y_{13}	...	y_{1j}	...	y_{1p}	
\mathbf{x}_2	y_{21}	y_{22}	y_{23}	...	y_{2j}	...	y_{2p}	
\mathbf{x}_3	y_{31}	y_{32}	y_{33}	...	y_{3j}	...	y_{3p}	
.	
.	
.	
\mathbf{x}_i	y_{i1}	y_{i2}	y_{i3}	...	y_{ij}	...	y_{ip}	
.	
.	
.	
\mathbf{x}_n	y_{n1}	y_{n2}	y_{n3}	...	y_{nj}	...	y_{np}	

Calculate Distances between samples

Sites	Species		
	y_1	y_2	y_3
x_1	0	1	1
x_2	1	0	0
x_3	0	4	4

From these data, the following distances are calculated between sites:

Sites	Sites		
	x_1	x_2	x_3
x_1	0	1.732	4.243
x_2	1.732	0	5.745
x_3	4.243	5.745	0

	1	2	...	<i>j</i>	...	<i>p</i>
1						
2						
↓						
O						
b						
j						
e				y_{ij}		y_{ip}
c						
t						
s						
↓						
<i>i</i>						
↓						
n				y_{nj}		y_{np}
↓						

← Descriptors →

Dispersion (covariance) matrix

	1	2	...	<i>j</i>	...	<i>p</i>
1						
2						
↓						
D		s_{22}		\dots	\dots	s_{2p}
e						
s						
c						
r						
i						
p		s_{j2}		\dots	\dots	s_{jp}
t						
o						
r						
s						
↓		s_{p2}		\dots	\dots	s_{pp}

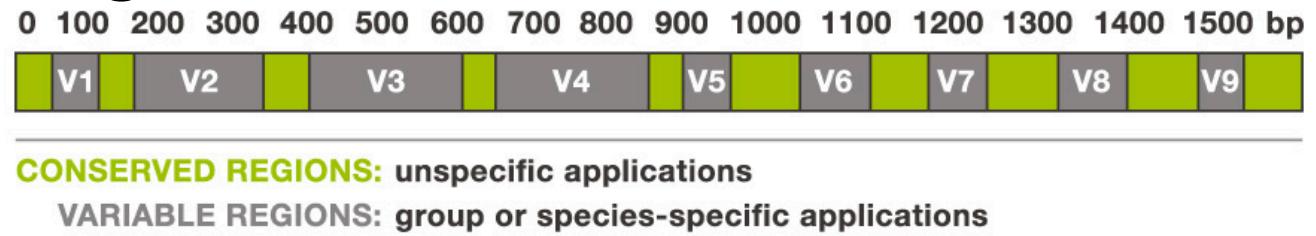
← Descriptors →

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (y_{ij} - \bar{y}_j) (y_{ik} - \bar{y}_k)$$

Figure 4.2 Structure of ecological data. Given their nature, ecological descriptors are *dependent* of one another. In statistics, the objects are often assumed to be *independent* observations, but this is generally not the case in ecology (Section 1.1)

How to get these species profiles?

- 16S rRNA Sequencing
- DNA Sequencing



Gene profiles are probably more informative

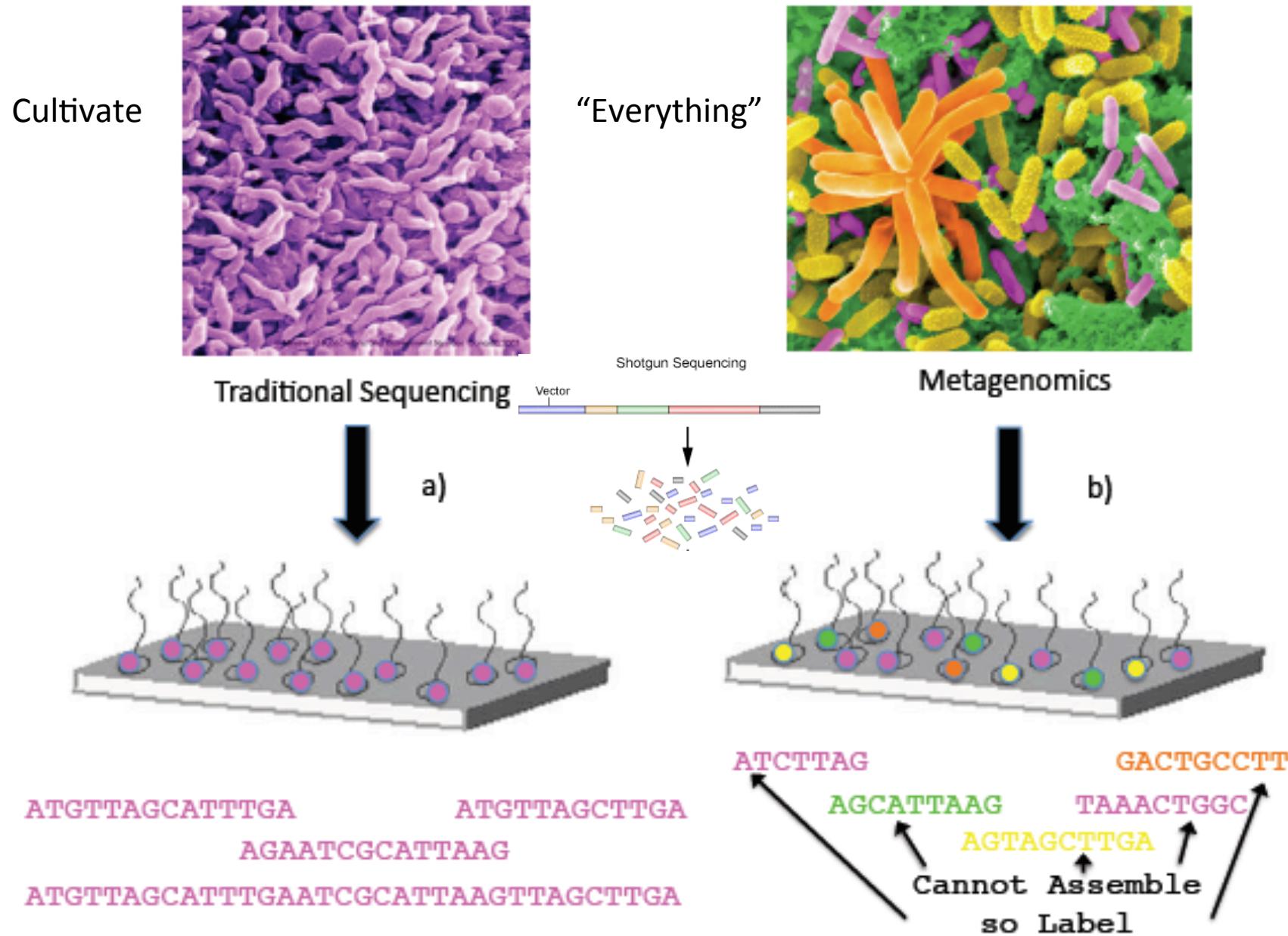
- DNA Sequencing
- RNA Sequencing

Genome Assembly: Culturing – not an option

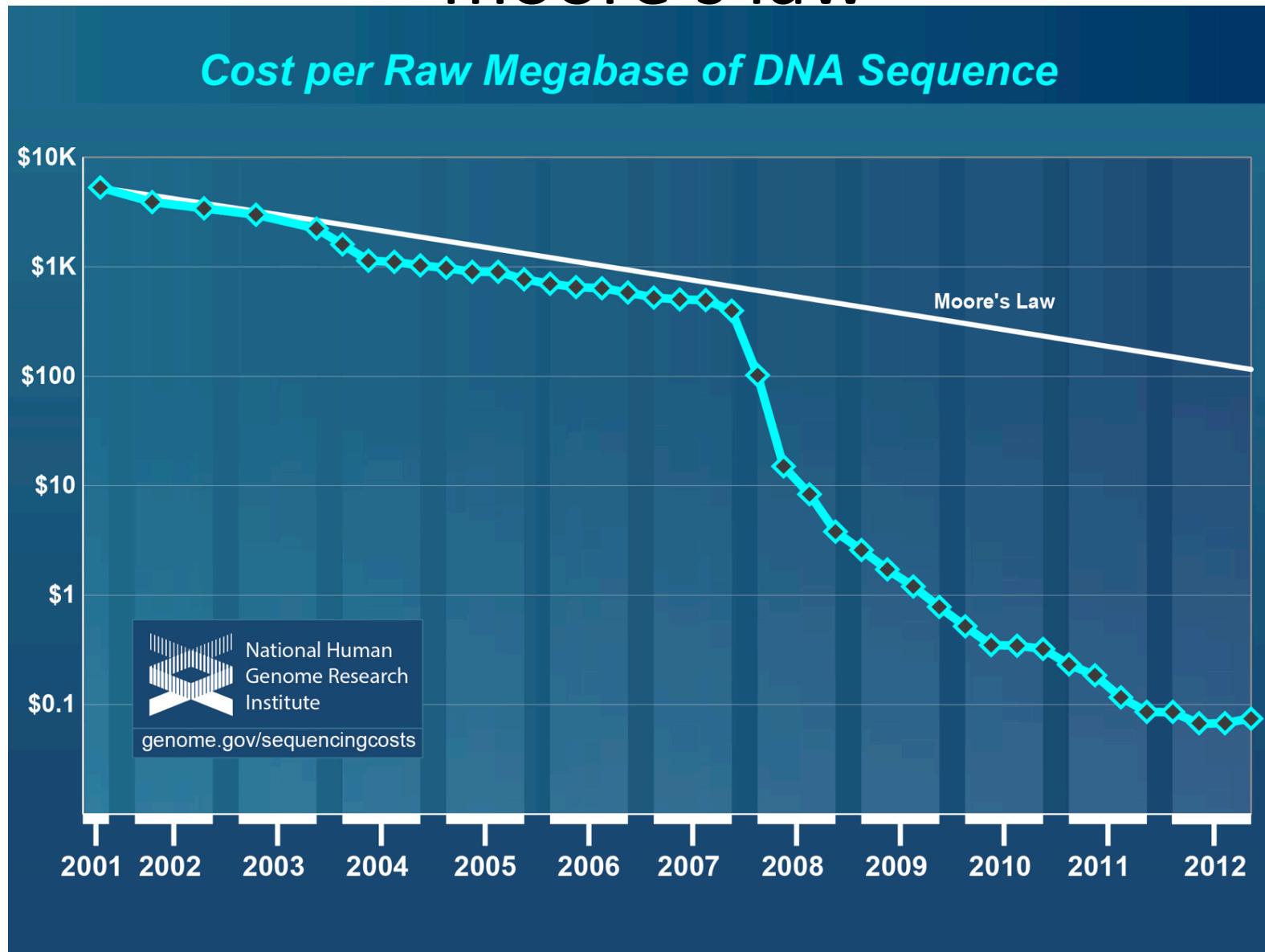
- Isolate ONE new species at a time
- Culture
- Sequence DNA
- Applies to <10% of microbes



Sequence Classification Problem



DNA Sequencing has outpaced moore's law



Problems

- What species and functions are in a sample?
- How and why does this sample differ from another sample?
- Are these differences causing disease or a particular sample attribute/phenotype?

Two Ways about finding out “Who is Here”?

16S

Advantage: Highly Conserved

Disadvantage: May not be able to distinguish some species and most strains

Everything

Advantage: Access to full genomes (use other housekeeping genes or can observe HGT among strains)

Disadvantage: So much info – that it can easily make mistakes (especially with HGT, highly similar COGs)

16S rRNA

- Who?



What is happening? (Instruction? Test?)

Am I teaching the
students?

Are the students
presenting?



Whole-genome Sequencing

- Who and what are their “tools”?



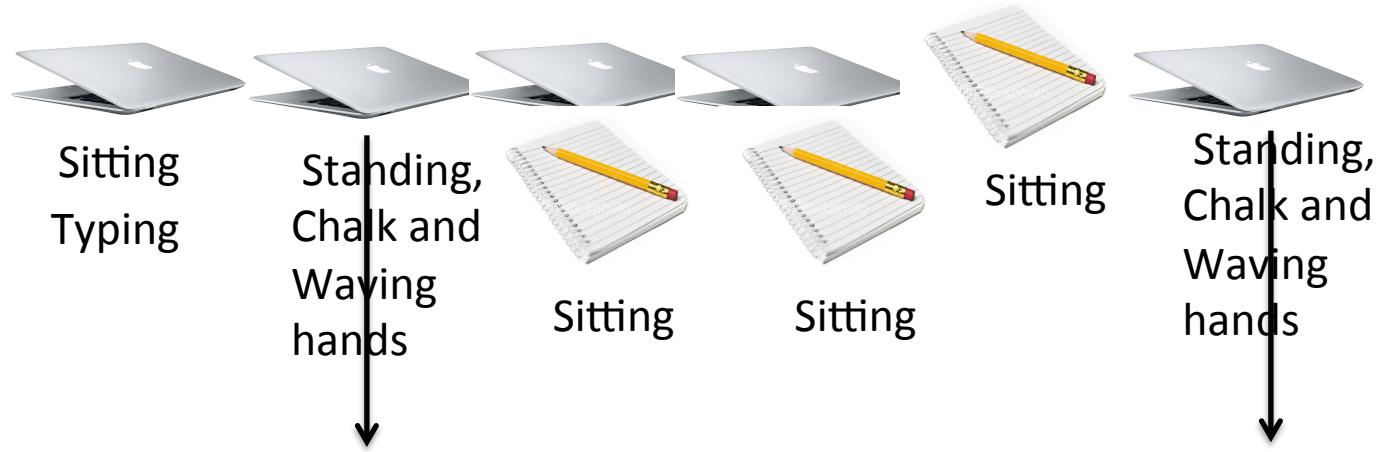
Most likely class



Most likely test

RNA-seq

- Who, what are their “tools”, and how much are they doing it?

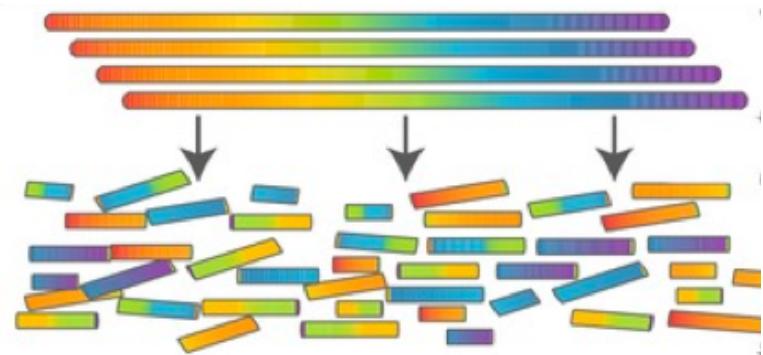


- Presenting to class

- Presenting to class

Types of Sequencing Summary

- 16S rRNA -- Barcode gene
- WGS -- DNA
- RNA-seq – Sequencing transcripts
(expressed genes)
- Limitations – Sequencing “coverage” or sampling depth

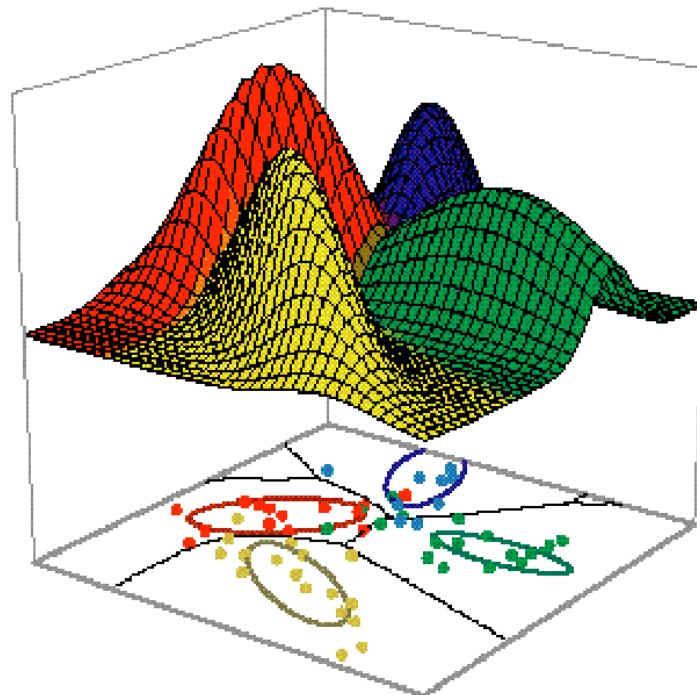


Other limitations

- Perhaps “Factor of Interest” is a **signal buried in noise**
- Statistical significance tests
- Variance studies will be pivotal

Signal Processing View of Metagenomics

- Mixture of “Stuff”
 - Pattern Recognition heaven
 - Machine Learning Techniques



Analogue of speech recognition to Nmer-Based feature Metagenomics

