Rob's various things

Contact details:

beiko@cs.dal.ca

@rob_beiko

Metagenomics* I: Who is There?

Rob Beiko

15 January 2013

*

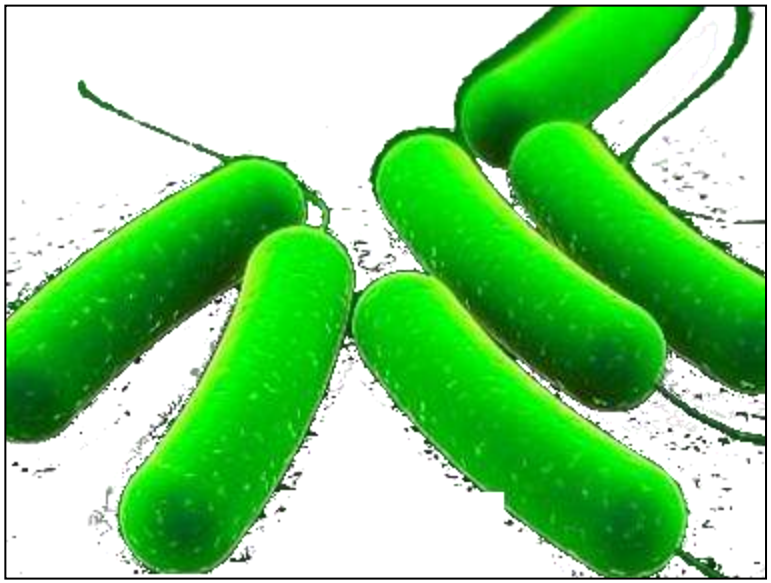
“Metagenomics” describes the functional and sequence-based analysis of the collective microbial genomes contained in an environmental sample."

"The definition applied here **excludes** studies that use PCR to amplify gene cassettes or random PCR primers to access genes of interest, since these methods do not provide genomic information beyond the genes that are amplified."

Riesenfeld et al., *Annu Rev Genet* 2004

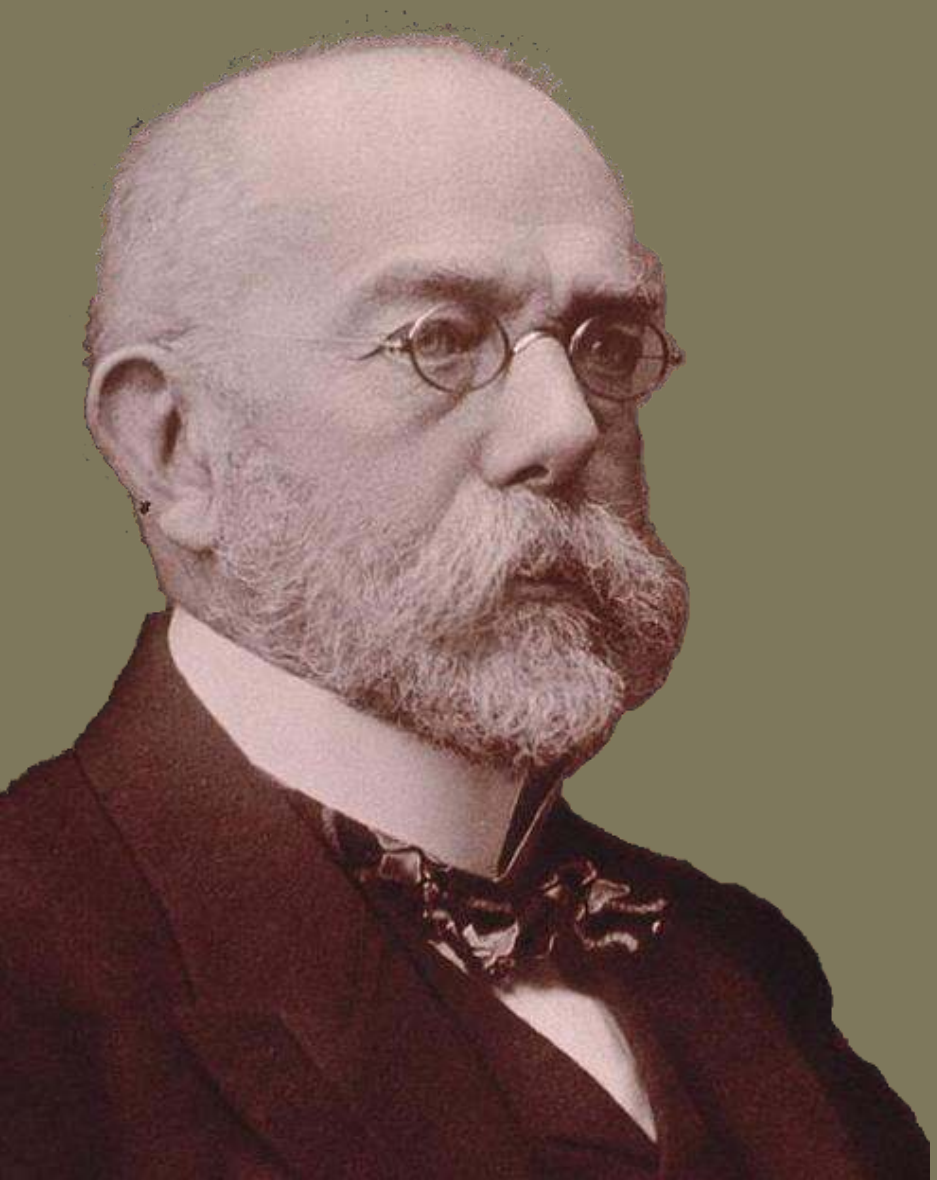
[we're going to talk about this anyway]

Who is there?



What are they doing?

"Who is There": What?



Robert Koch

(1843-1910)

Historical views of microorganisms

Pre-1600s:

Historical views of microorganisms

1670 **van Leeuwenhoek**:
birth of microscopy, discovery of "animalcules"

1774 **Linnaeus** helpfully brands as "chaos"

Late 1800s: **Haeckel**'s "Lebensbaum"

Historical views of microorganisms

1862-1945

Phenotypic classification

Gram staining

Microscopy

Koch's postulates

"During this period, it was widely assumed by bacteriologists that bacteria possessed no species as such…and that bacterial heredity and evolution involved a vague Lamarckian mechanism."

Sapp, J. (2005) Microbial Phylogeny and Evolution: Concepts and Controversies.

Historical views of microorganisms

1946-1977

Bacterial genetics

Bacteria have genes!!

The prokaryote / eukaryote divide

Observation and exploitation of gene transfer mechanisms

Despair about a "natural" classification system

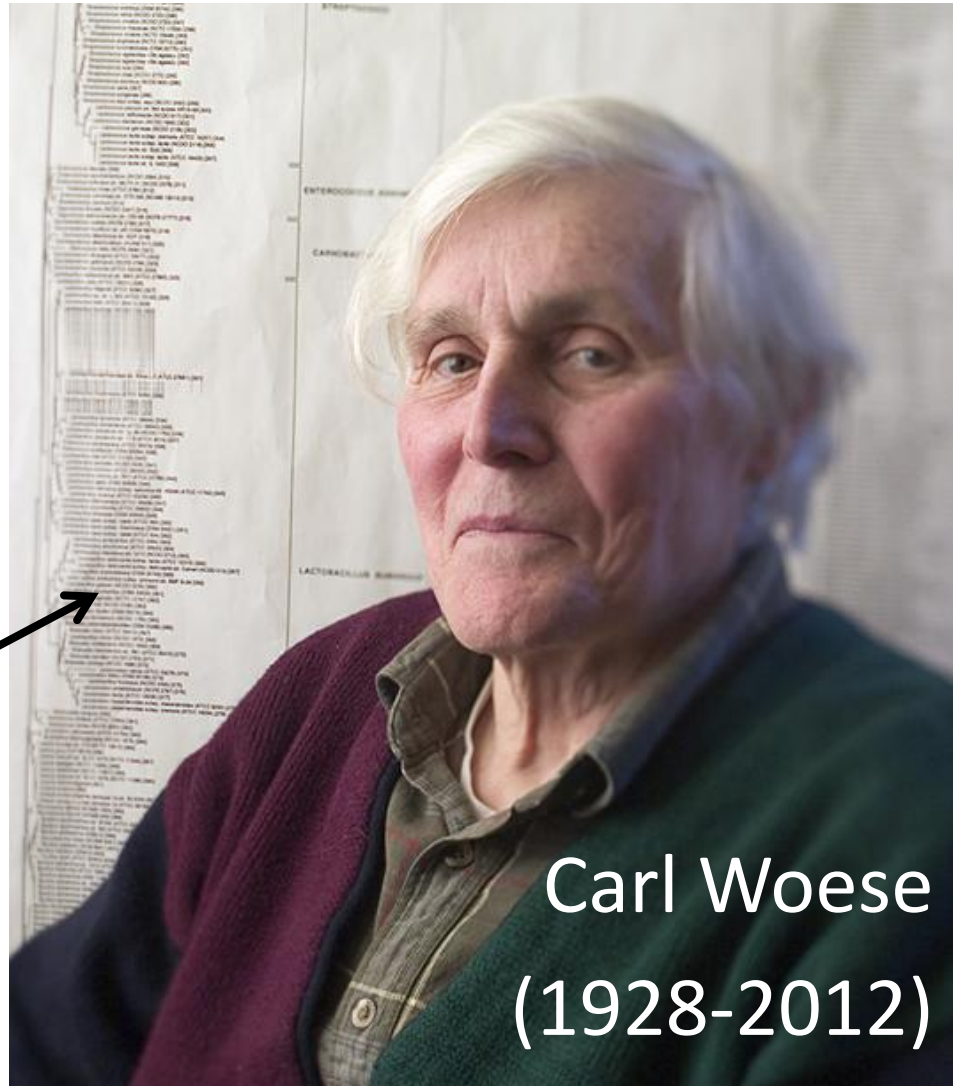
Jacques Monod (1910-1976)

Historical views of microorganisms

1977-1994

Molecular taxonomy and the primacy of "marker genes"

The Answer

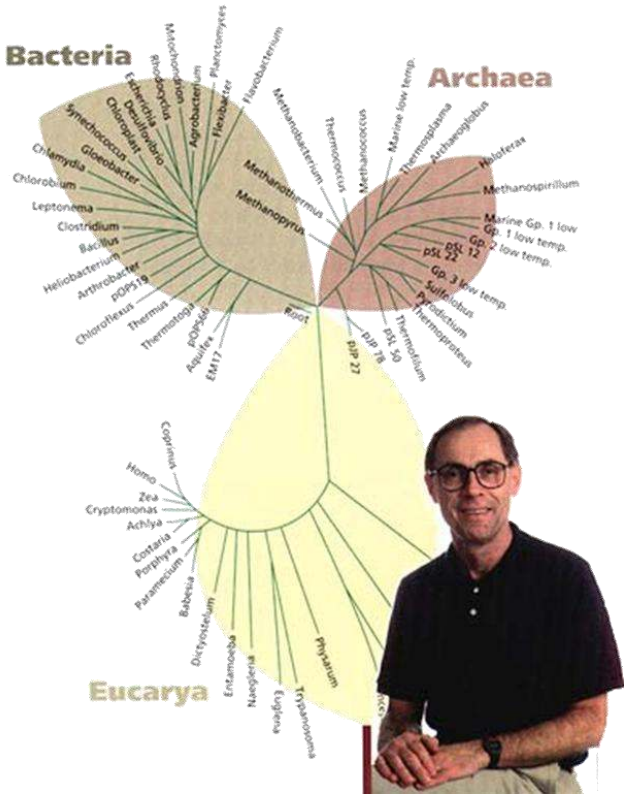


Historical views of microorganisms

1995-present

Genome sequencing: I'm so confused!

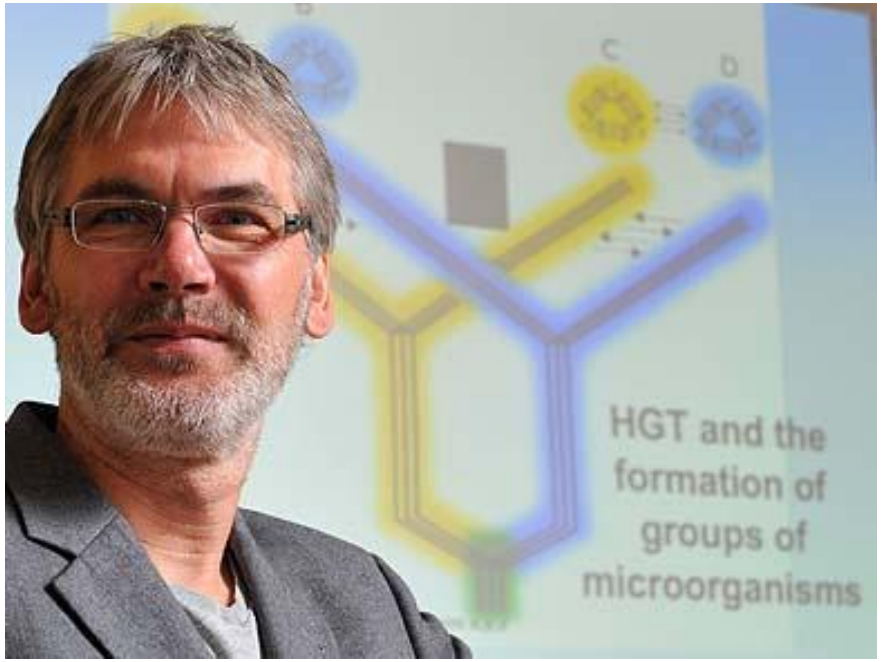
The "Tree of Life"



Norman Pace

vs.

Networks of microbes



J. Peter Gogarten

Prerequisites for "Who is There"

- DEFINITION of species or some other taxonomic / phylogenetic unit
  - Strain
  - Genus, family, etc.
  - Serovar
  - Pathogroup
  - Ecotype
  - Operational Taxonomic Unit (OTU)
  - …

Prerequisites for "Who is There"

- CRITERIA for assigning an organism with a set of observations to a group
  - Morphology
  - Physiology
  - Immunochemical properties
  - **Genetic features**

Koch's Postulates (1890)

A particular "parasite" causes a given disease if…

(i) The parasite occurs in every case of the disease in question and under circumstances which can account for the pathological changes and clinical course of the disease.

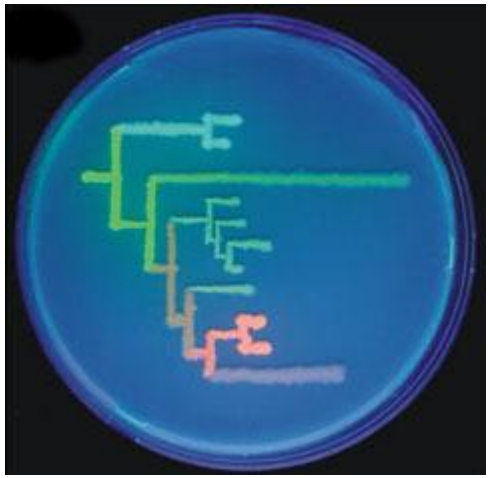
(ii) The parasite occurs in no other disease as a fortuitous and nonpathogenic parasite.

(iii) After being fully isolated from the body and repeatedly grown in pure culture, the parasite can induce the disease anew"

(iv) *"… a requirement to reisolate the microbe from the experimentally inoculated host"*

Culturing and the "Great Plate Count Anomaly"

- Staley and Konopka (1985)
- < 1% (or 0.1%, …) or microorganisms can be grown in pure culture
- Depends on conditions, what's needed, etc…
- What comes up may not be the most abundant / important microorganisms

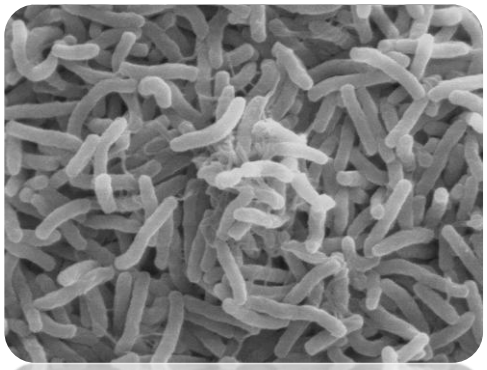


Ugalde et al. *Science (*2004)

Why?

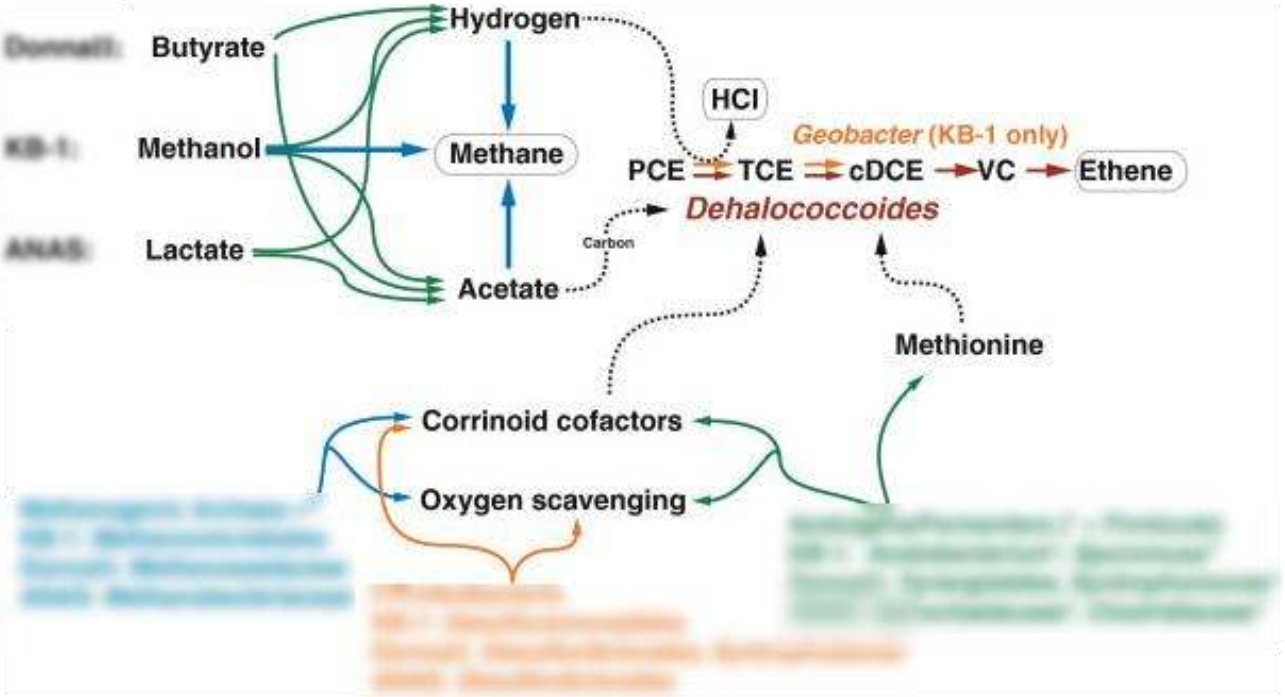
- Natural settings can be very difficult to recreate

- Some organisms grow slooooooowly in the best of circumstances



Aaaaaand…

Dependence on other microbes



Hug et al., *BMC Genomics* (2012)

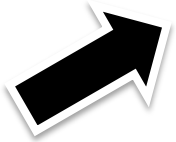
Original recipe makes 1 - 9 inch 3 layer cake Change Servings

- 2 1/8 cups all-purpose flour
- 2 cups white sugar
- 3/4 cup unsweetened cocoa powder
- 1 1/2 teaspoons baking powder
- 3/4 teaspoon baking soda
- 3/4 teaspoon salt
- 3 eggs
- 1 cup milk
- 1/2 cup vegetable oil
- 1 tablespoon vanilla extract
- 2 (20 ounce) cans pitted sour cherries
- 1 cup white sugar
- 1/4 cup cornstarch
- 1 teaspoon vanilla extract
- 3 cups heavy whipping cream
- 1/3 cup confectioners' sugar

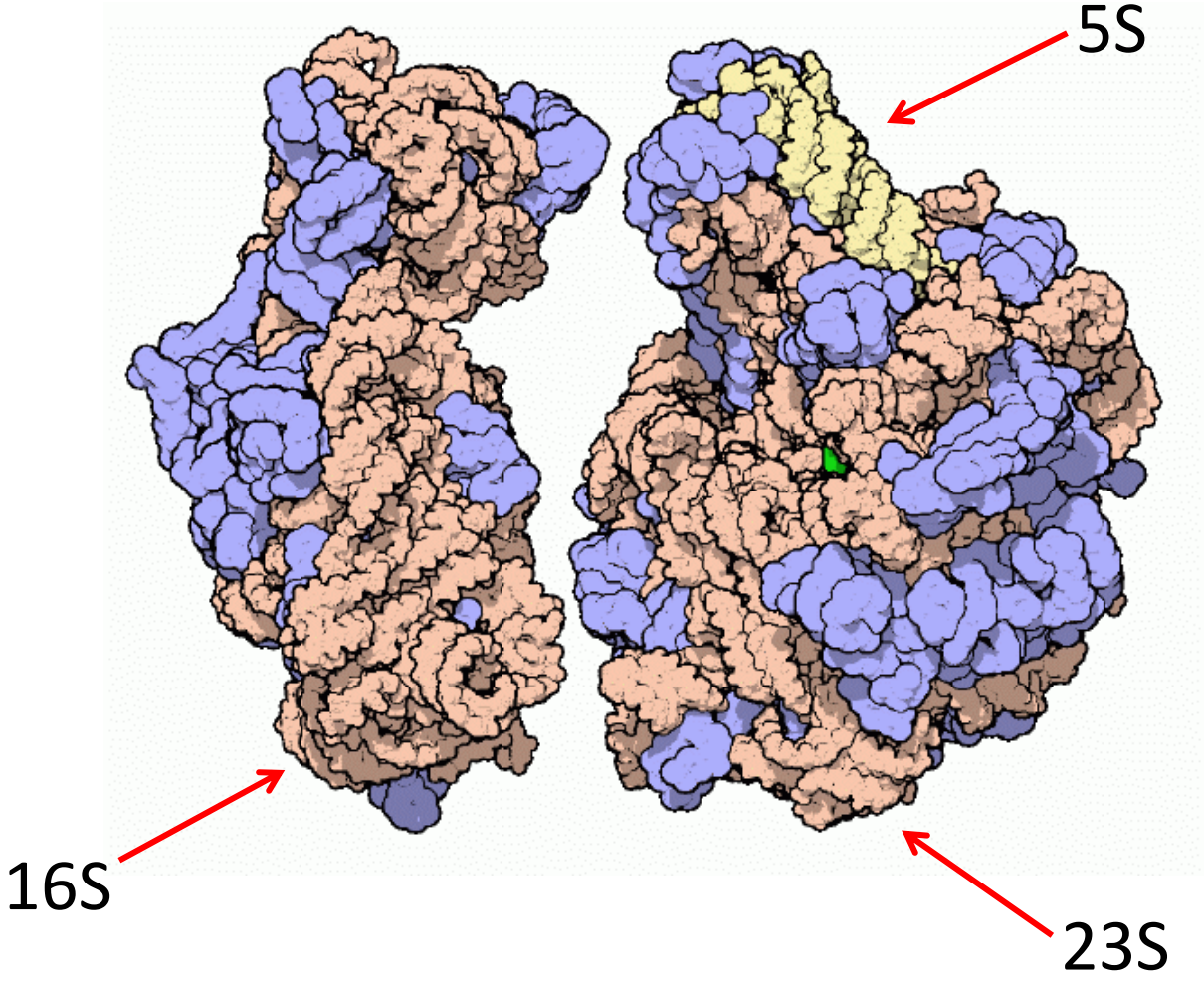
Greer, L. (2003) *Allrecipes.com*

Original recipe makes 1 - 9 inch 3 layer cake Change Servings

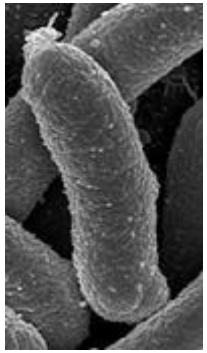
- 2 1/8 cups all-purpose flour
- [redacted]
- 3/4 cup unsweetened cocoa powder
- [redacted]
- [redacted]
- 1 tablespoon vanilla extract
- [redacted]
- 1/4 cup cornstarch
- [redacted]



Can we have our cake?

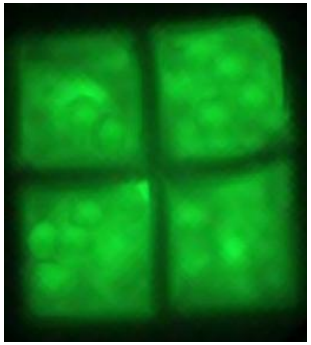


Ribosomal RNA genes: taxonomic and phylogenetic markers

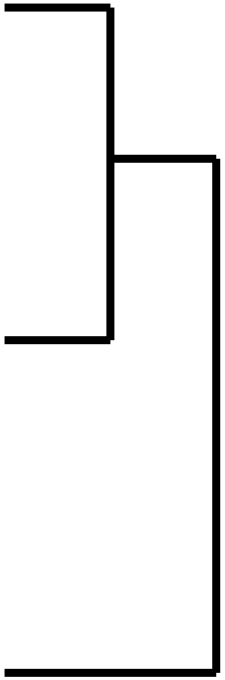


GGATCCAGGCTCTGCTAACCAGGGCGGCGATCTCGGCTGGGCTACACCAG...

TGAACTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCT...



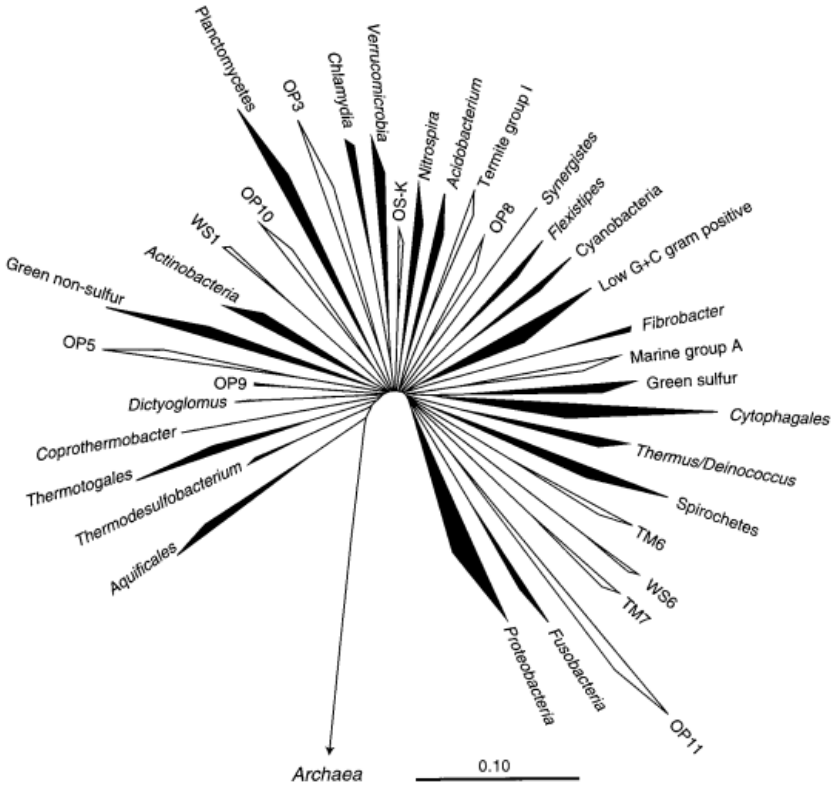
ATTCCGGTTGATCCTGCCGGAGGCCATTGCTATCGGAGTCCGATTTAGCC...



What's so special about ribosomal RNA genes?

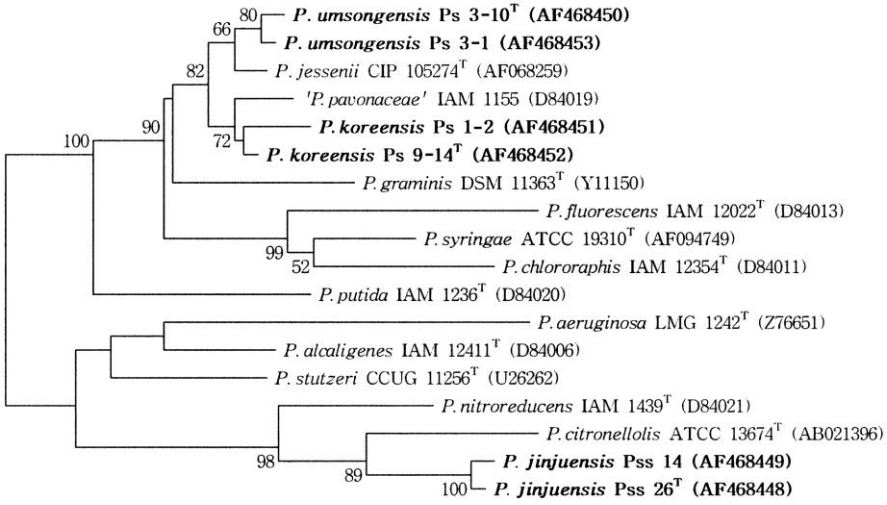
- Everyone has them
- They evolve slowly overall
- But have different rates in different regions (e.g., 'variable' regions)
- The 16S rRNA gene is generally accepted as "long enough" to yield accurate phylogenetic trees

All Bacteria



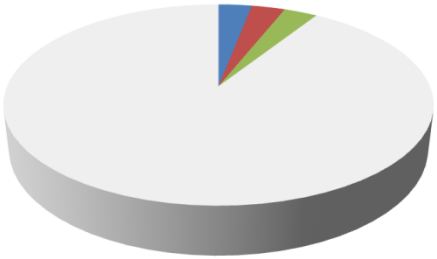
Hugenholtz et al., *J Bacteriol* (1998)

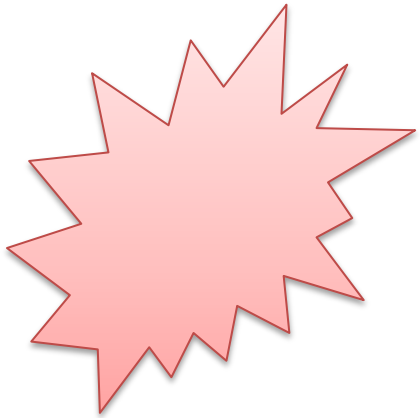
Pseudomonas

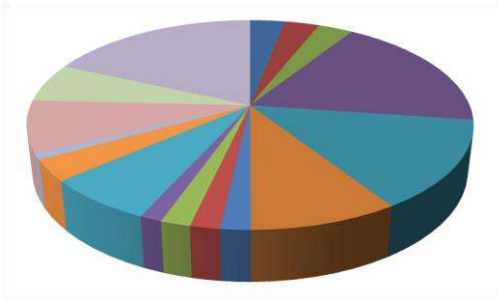


Kwon et al., *IJSEM* (2003)

How does this help us?



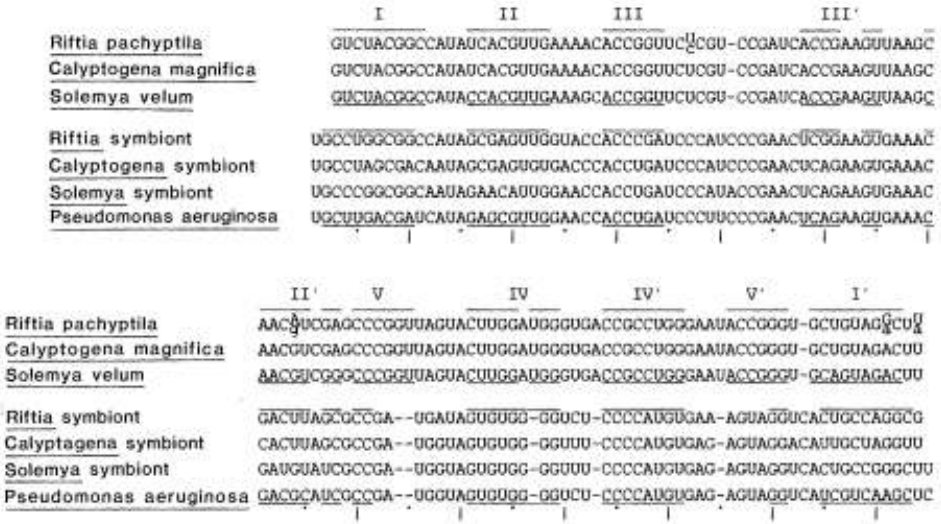


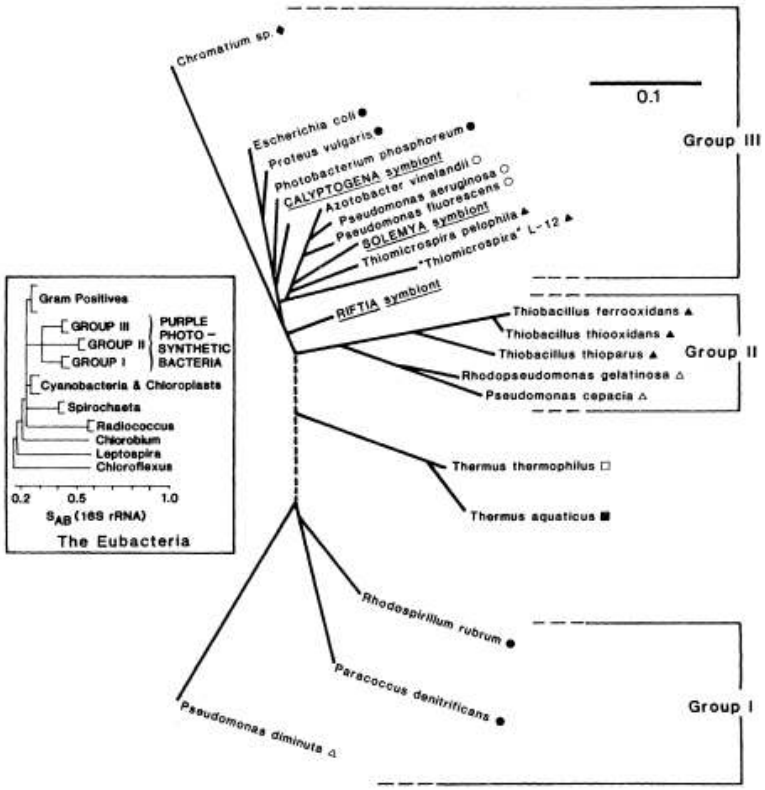


Analysis of Hydrothermal Vent–Associated Symbionts by Ribosomal RNA Sequences

David A. Stahl
David J. Lane
Gary J. Olsen
Norman R. Pace*

"5S rRNA was used because it is relatively easily isolated and analyzed and because its sequence for some 200 organisms and organelles are available for comparison."





Characterizing the unculturable
Science, 1984

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, June 1985, p. 1379–1384
0099-2240/85/061379-06$02.00/0

Vol. 49, No. 6

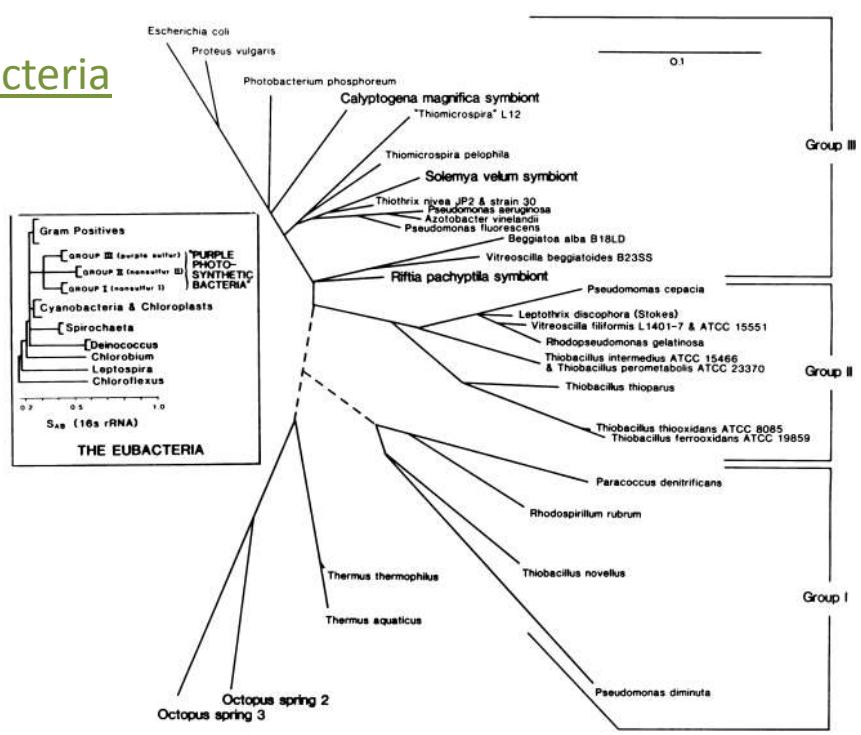
Characterization of a Yellowstone Hot Spring Microbial Community by 5S rRNA Sequences

DAVID A. STAHL,† DAVID J. LANE, GARY J. OLSEN, AND NORMAN R. PACE*

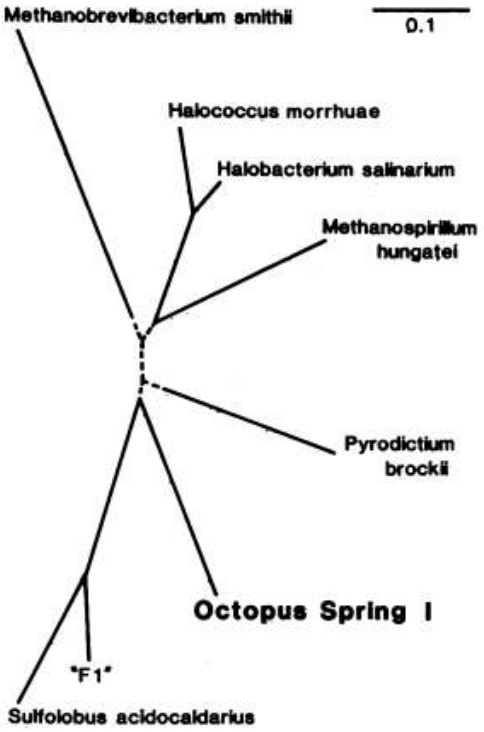
Department of Biology and Institute for Molecular and Cellular Biology, Indiana University, Bloomington, Indiana 47405

Received 30 November 1984/Accepted 8 March 1985

Bacteria



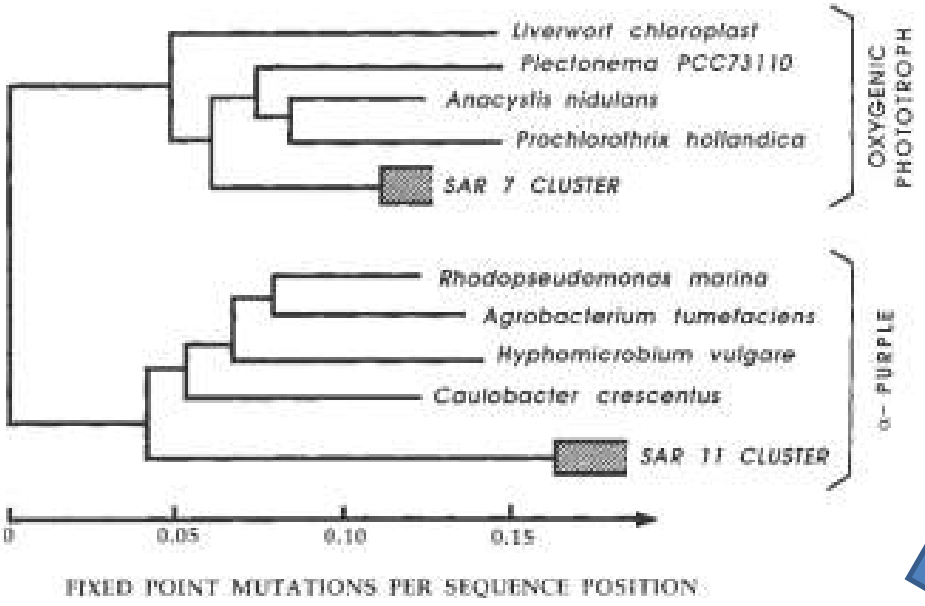
Archaea



Profiling the microbial community

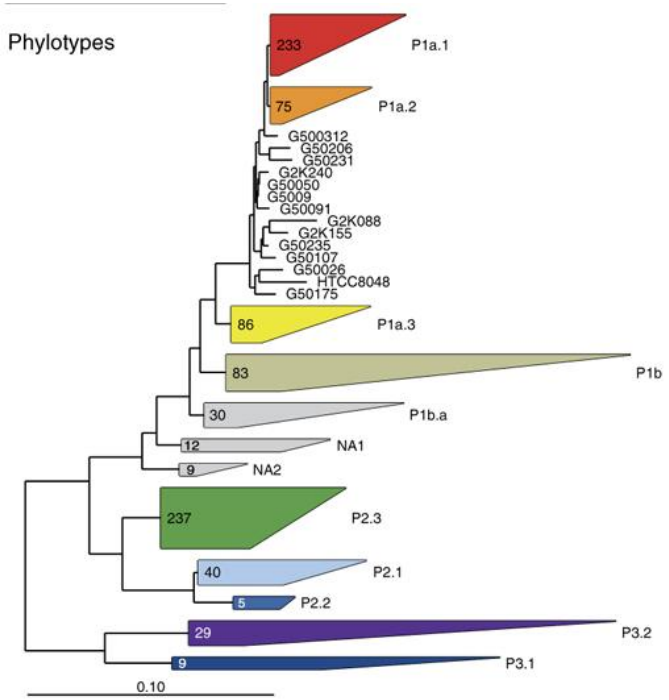
Genetic diversity in Sargasso Sea bacterioplankton

Stephen J. Giovannoni, Theresa B. Britschgi, Craig L. Moyer & Katharine G. Field

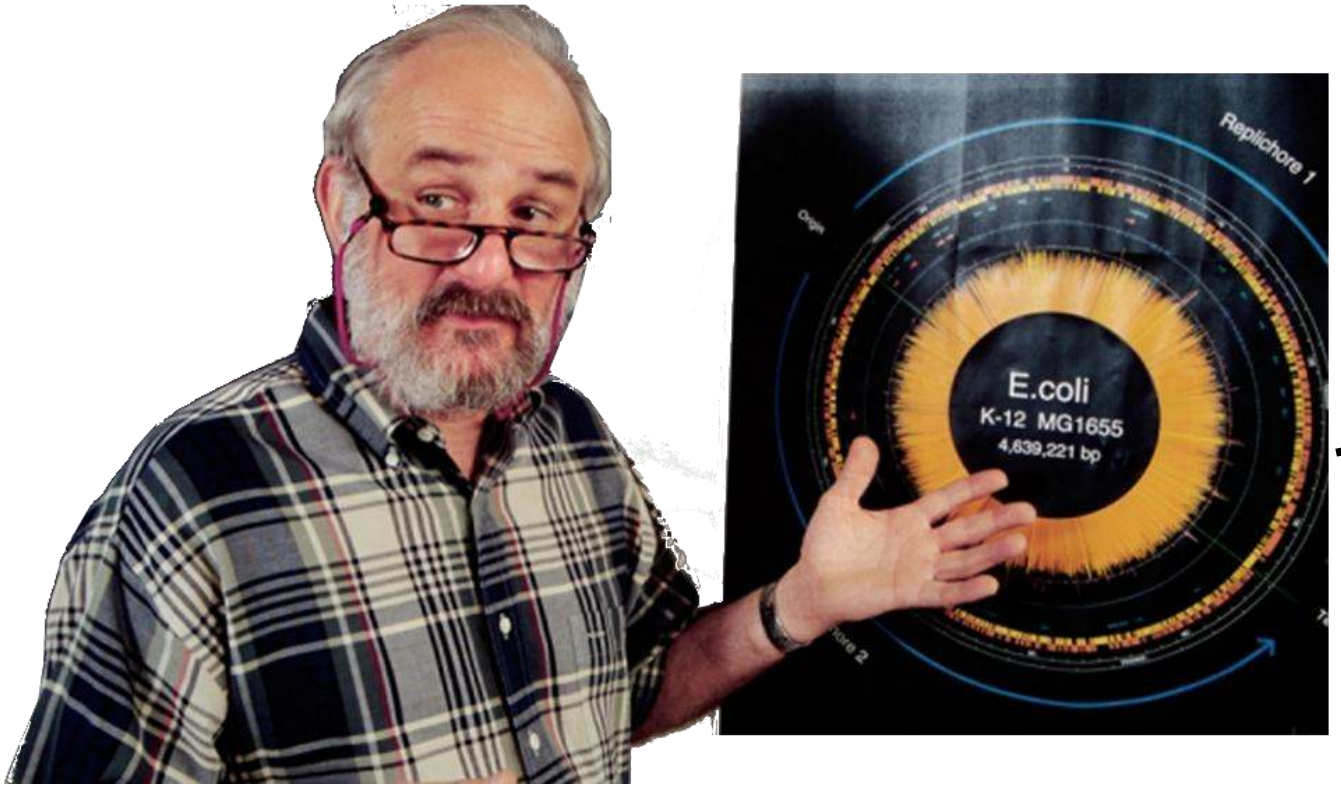


Nature, 1990

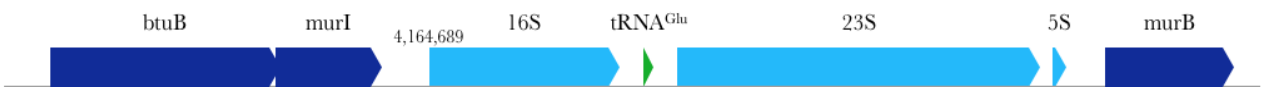
Brown et al., *Mol Syst Biol*, 2012



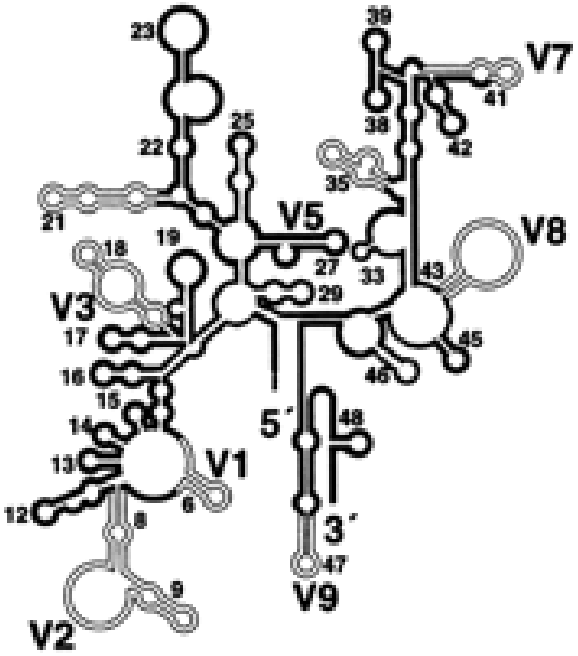
Using taxonomic / phylogenetic marker genes



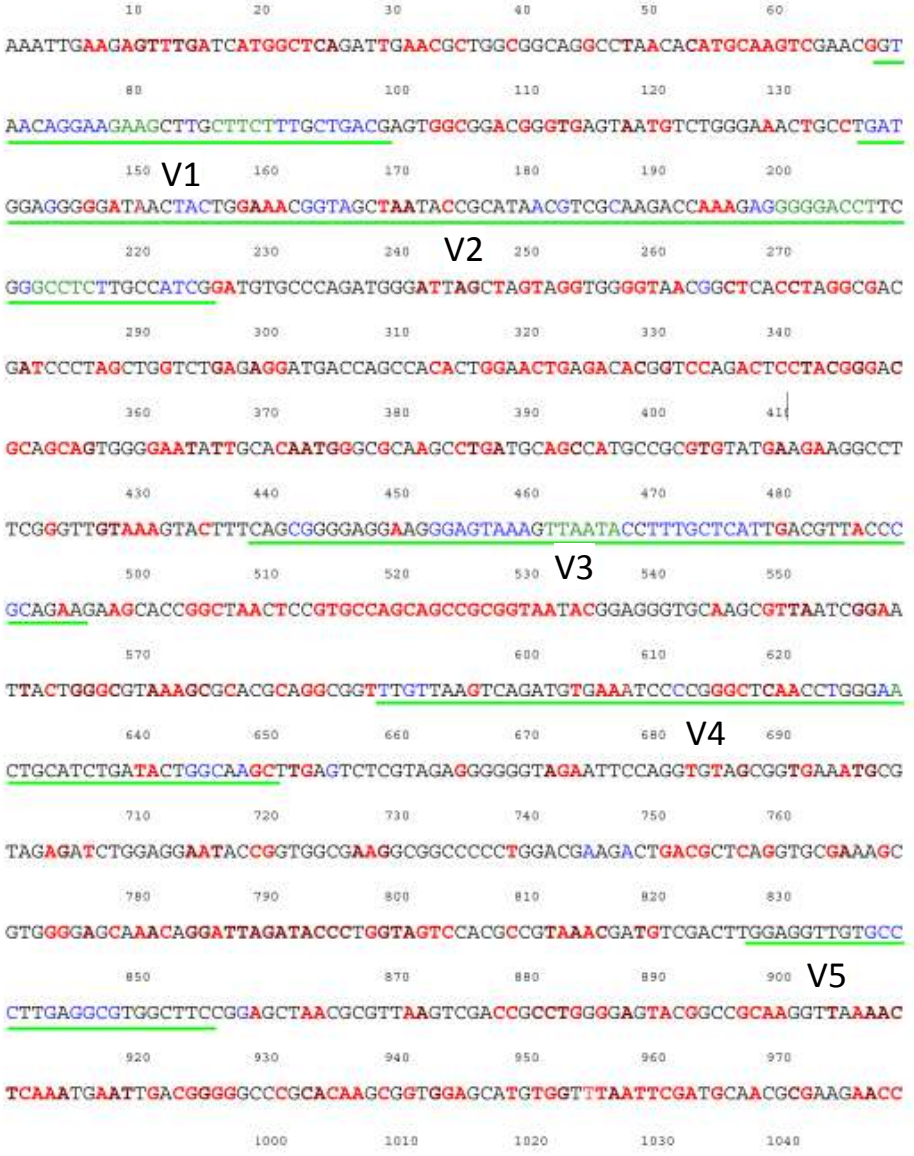
Blattner et al., *Science* 1997



http://www.cbs.dtu.dk



Tortoli 2003 *Clin Microbiol Rev*



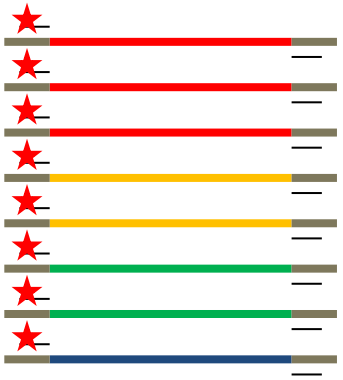
Baker et al 2003 *J Microbiol Meth*

Assessing diversity using 16S

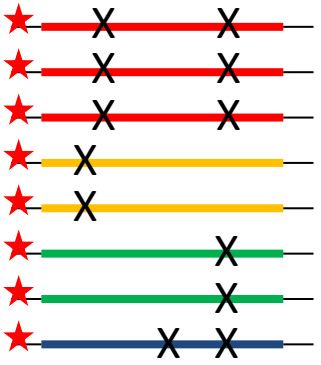
Terminal restriction fragment length polymorphism

T-RFLP

DNA sample in clone library

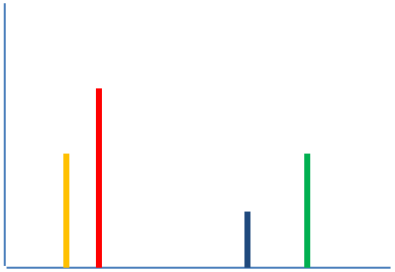


PCR amplify
5′ end fluorescently labelled

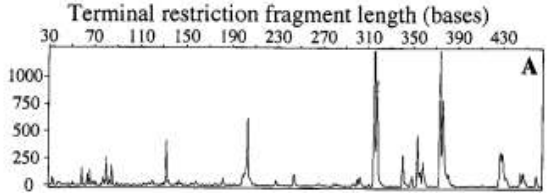


Restriction digest

Variable-length fragments



Electropherogram



Profile of a wastewater community
(Liu et al., *Appl Environ Microbiol*, 1997)

Denaturing Gradient Gel Electrophoresis

DGGE

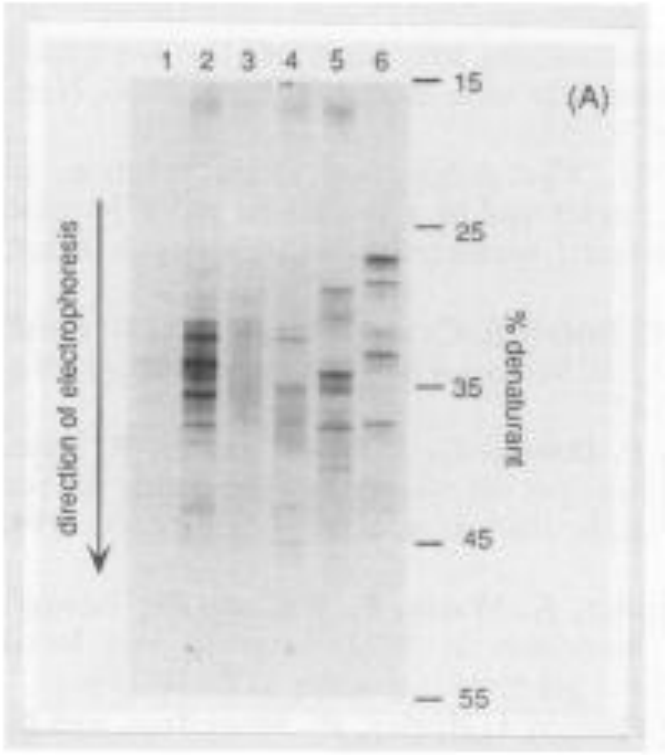
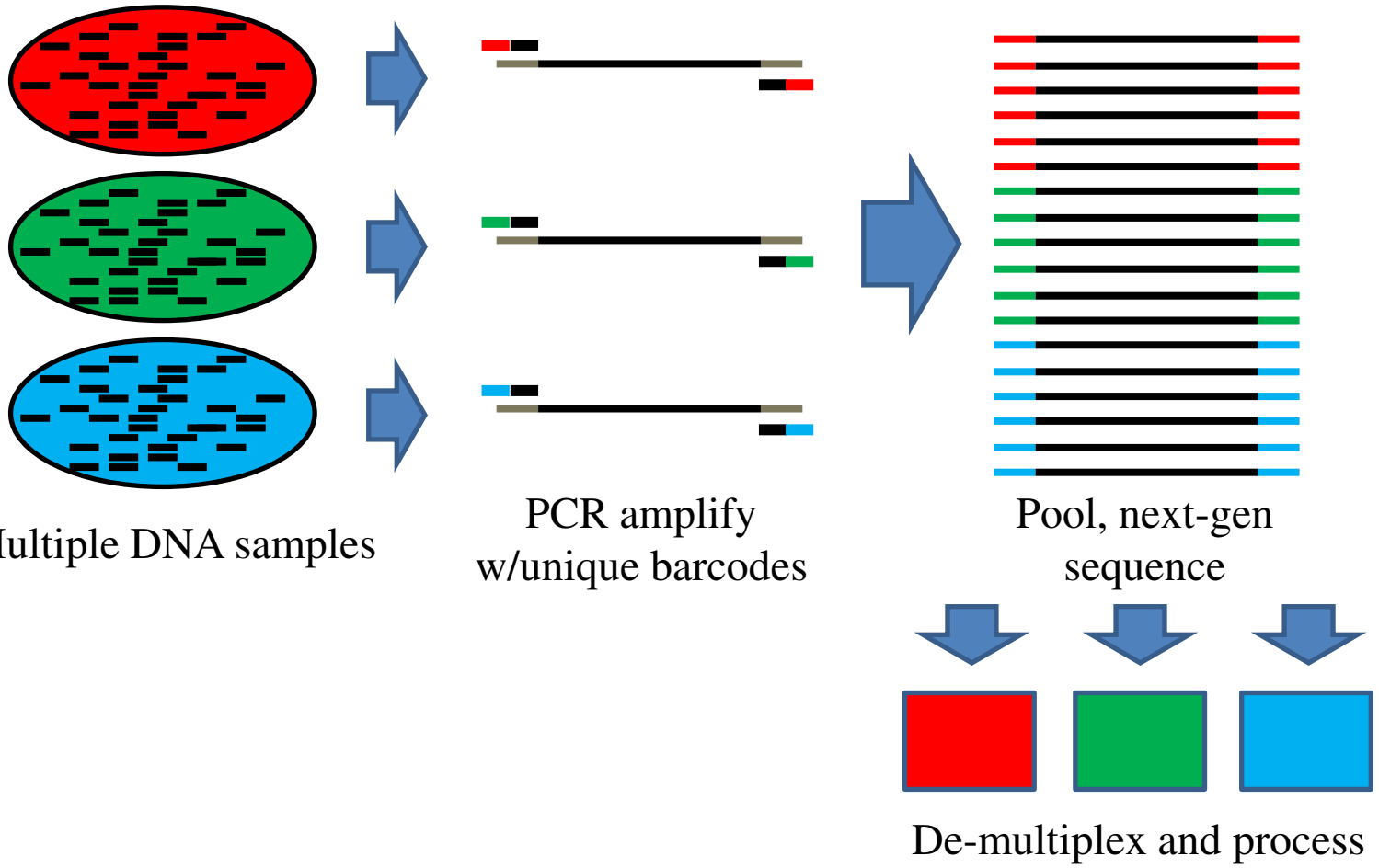


FIG. 7. DGGE analysis of 16S rDNA fragments obtained after enzymatic amplification of genomic DNA from uncharacterized microbial populations and individual bacteria. (A) A negative image of an ethidium bromide-stained parallel DGGE separation pattern of microbial mat samples 1 (lane 1), 2 (lane 2), and 3 (lane 3), a bacterial biofilm grown under aerobic conditions (lane 4), and a bacterial biofilm grown under anaerobic conditions (lane 5) is shown. Lane 6 contains the separation pattern of a mixture of PCR fragments of five individual bacteria, i.e., *D. sapovorans*, *E. coli*, *M. chthonoplastes*, *T. thioparus*, and *D. desulfuricans*. This mixture served as a positive control for the hybridization experiment. (B) The results after hybridization analysis of this DGGE separation pattern and hybridization with a oligonucleotide probe specific for sulfate-reducing bacteria are presented.

Muyzer et al., *Appl Environ Microbiol* 1993

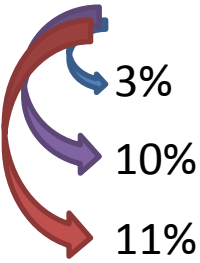
Amplicon sequencing



What to do with a 16S profile

- Available options will depend on how much information you have collected
- We can query or estimate:
  - Richness (# of distinct things)
  - Diversity (richness + *evenness*)
  - The presence of specific sets of lineages
  - The *similarity* between samples
- T-RFLP, DGGE, etc: observations are either identical or maximally dissimilar
- Sequencing gives a great deal more information

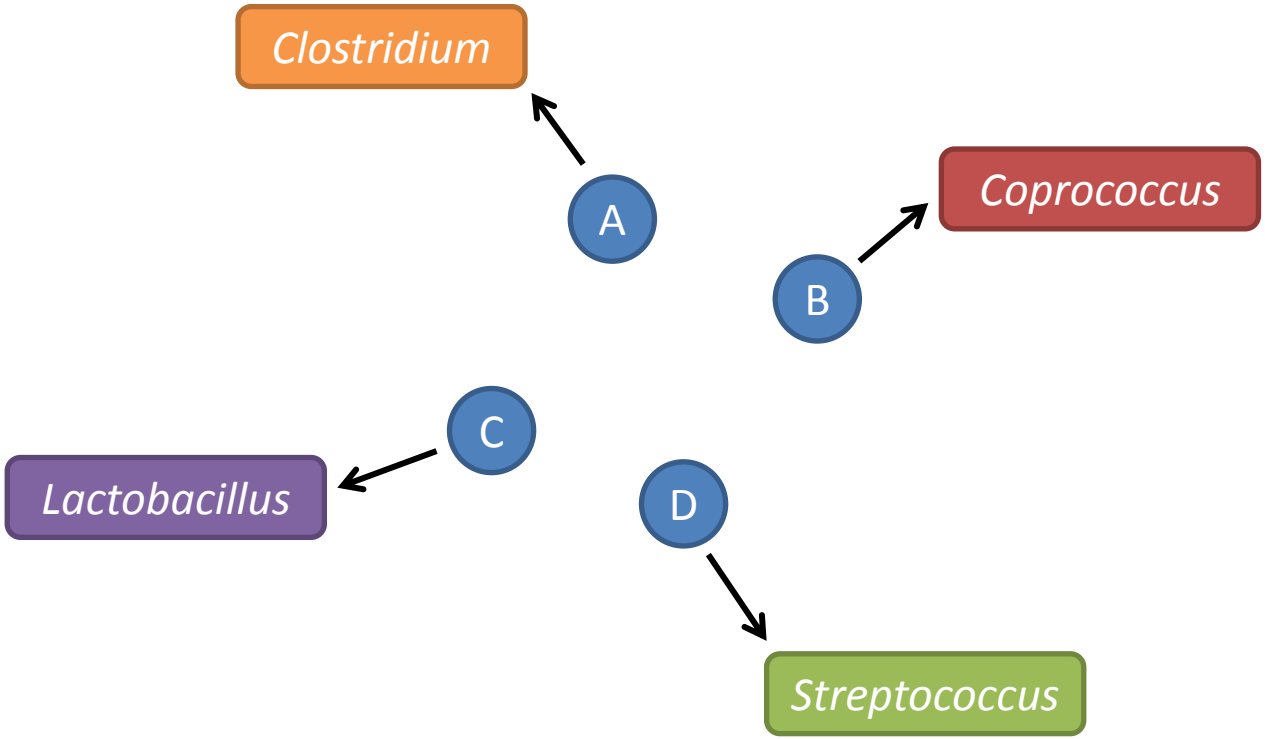
Your 16S data




```
>A
---T---T-G------C----G-C---A-AG-G---AT-G---G----G-CT-C-GC-G--T-C-T-G--A------TT-A--G-AT-A-G--T-TGG-T-G
>B
---T---C-G------C----T-G---A-AA-G---AT-G---G----A-CT-C-GC-G--T-C-T-G--A------TT-A--G-CT-A-G--T-TGG-C-G
>C
-------T-G------G----T-T---T-TA-G---AT-G---G----G-CA-T-GC-G--T-G-G-C--A------TT-A--G-GT-A-G--T-TGG-C-G
>D
---A---C-G------G----T-A---T-GA-G---AT-G---G----A-CC-C-GC-G--T-C-T-G--A------TT-A--G-GT-A-G--T-TGG-T-G
```

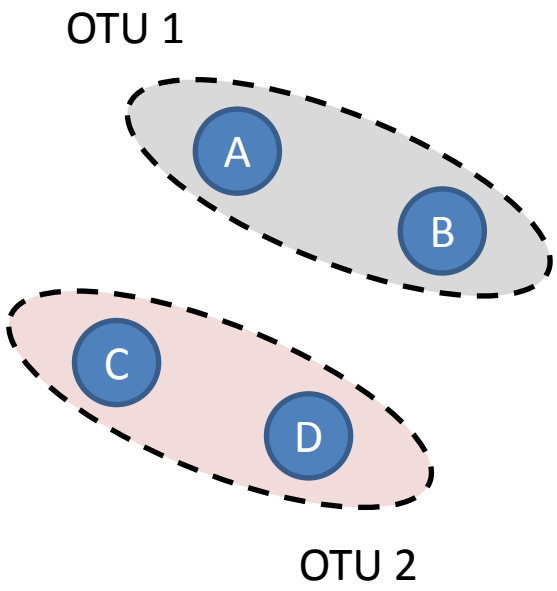
% dissimilarity between sequences

1. Taxonomic assignment (SUPERVISED analysis)



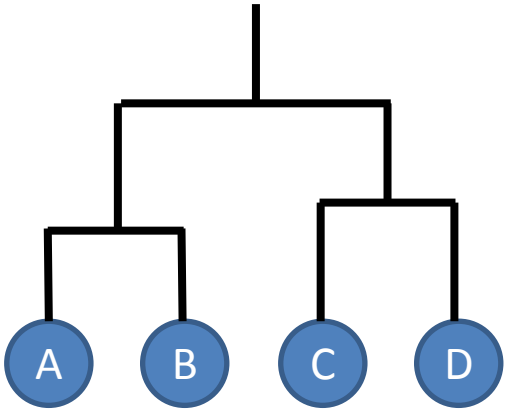
Homology (best BLAST match)
Composition (k-mer similarity)

2. *de novo* OTU clustering (UNSUPERVISED analysis)



Cluster based on sequence identity

3. Phylogenetic tree construction



What's the difference?

Taxonomic assignment uses a reference database, puts names on things (even if names may not be super-informative)

OTU clustering groups similar sequences together, treats each group as a distinct unit of analysis

Phylogenetic analysis considers the *relative* similarity of sequences, allows for more nuanced analysis

Another important distinction

Alpha diversity: consider the properties of each sample separately (genus diversity, OTU richness, etc.)

Beta diversity: consider the **pairwise** similarity between samples (dissimilarity indices, phylogenetic overlap, etc.)

These lead to different types of analysis

What do we do with diversity information?

(15 minutes)

- Alpha-diversity
  - Compare against other information (pH, etc.)
- Beta-diversity
  - Build a matrix
  - Cluster, relate to sample attributes
- Examples coming after the break!

Example #1: Soil, pH, bacteria and fungi

The ISME Journal (2010) 4, 1340–1351

www.nature.com/ismej

ORIGINAL ARTICLE

Soil bacterial and fungal communities across a pH gradient in an arable soil

Johannes Rousk[1,7], Erland Bååth[1], Philip C Brookes[2], Christian L Lauber[3], Catherine Lozupone[4], J Gregory Caporaso[4], Rob Knight[4,5] and Noah Fierer[3,6]

[1]Department of Microbial Ecology, Lund University, Ecology Building, Lund, Sweden; [2]Soil Science Department, Rothamsted Research, Harpenden, Herts, UK; [3]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA; [4]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, USA; [5]Howard Hughes Medical Institute, University of Colorado, Boulder, CO, USA and [6]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA

Big questions:
-What is the influence of soil pH on microbial diversity and abundance?
-Do effects differ between bacteria and fungi?

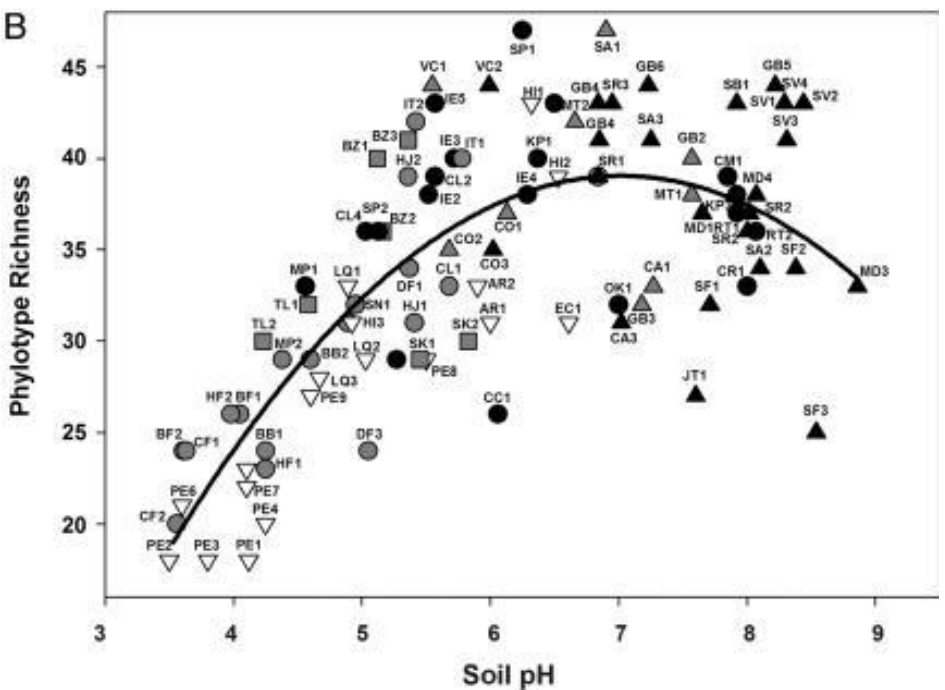
pH and microbial diversity

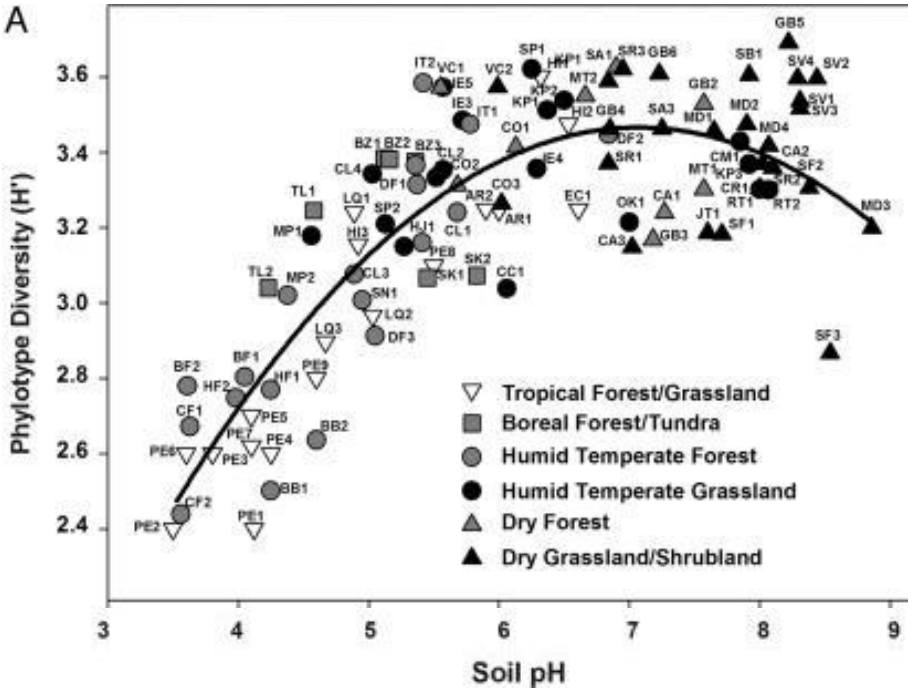
From Fierer and Jackson, *PNAS* (2006)

OTUs estimated with T-RFLP

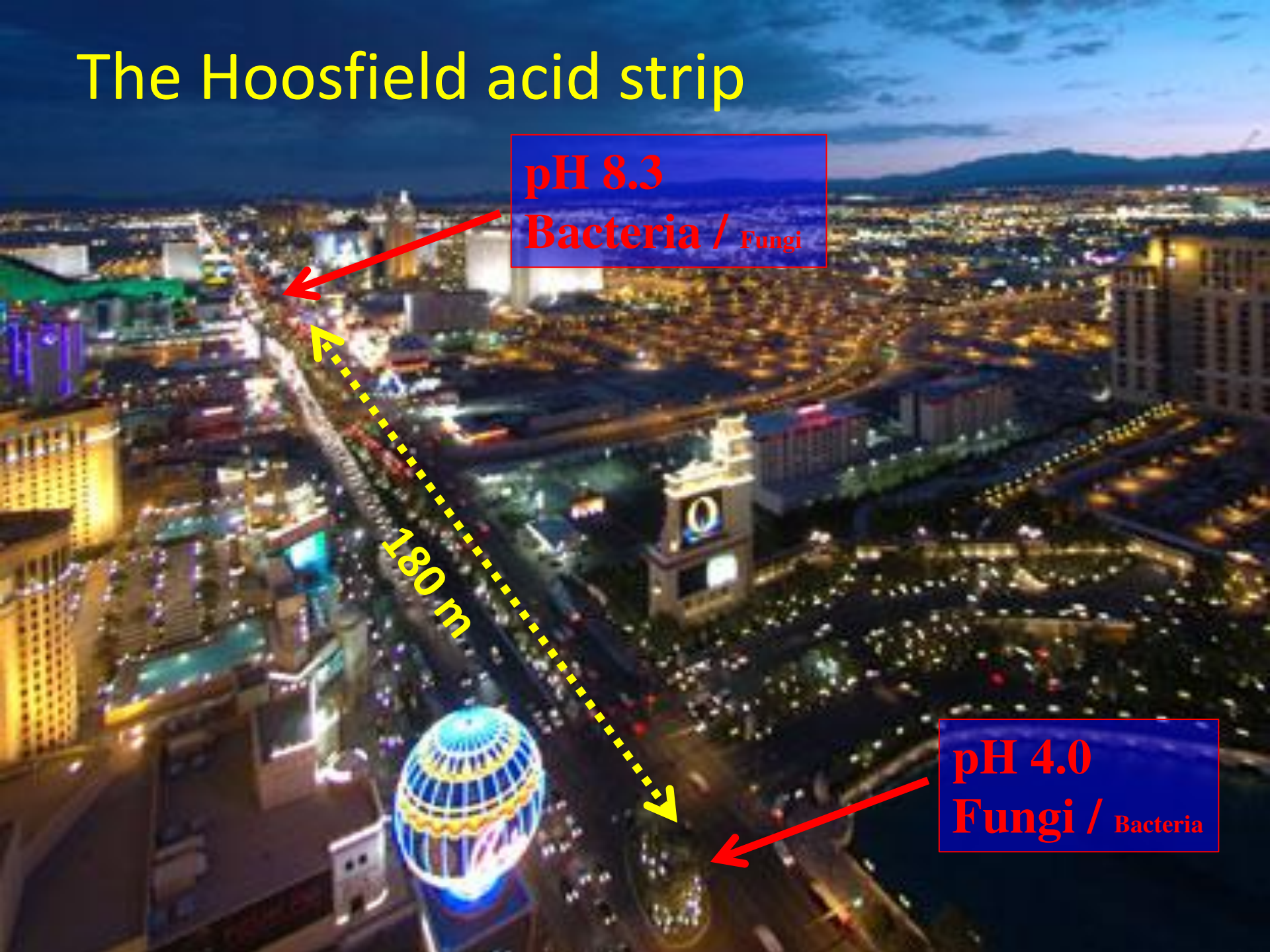
$$H' = -\sum_{i=1}^{R} p_i \log p_i$$

pH was the only property to display a strong relationship





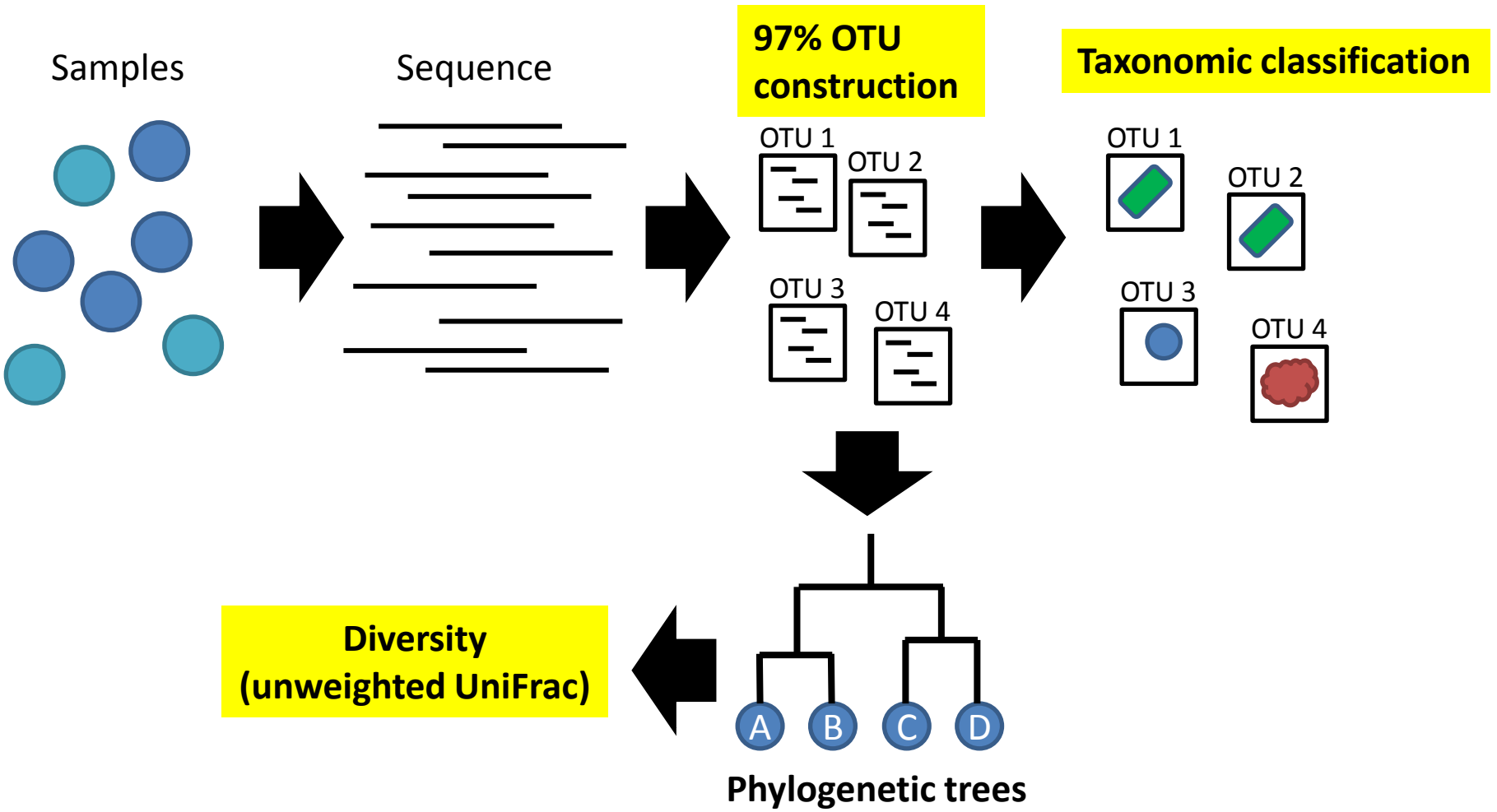
The Hoosfield acid strip



- Previous work based on T-RFLP and fatty acid composition was informative, but failed to give taxonomic resolution

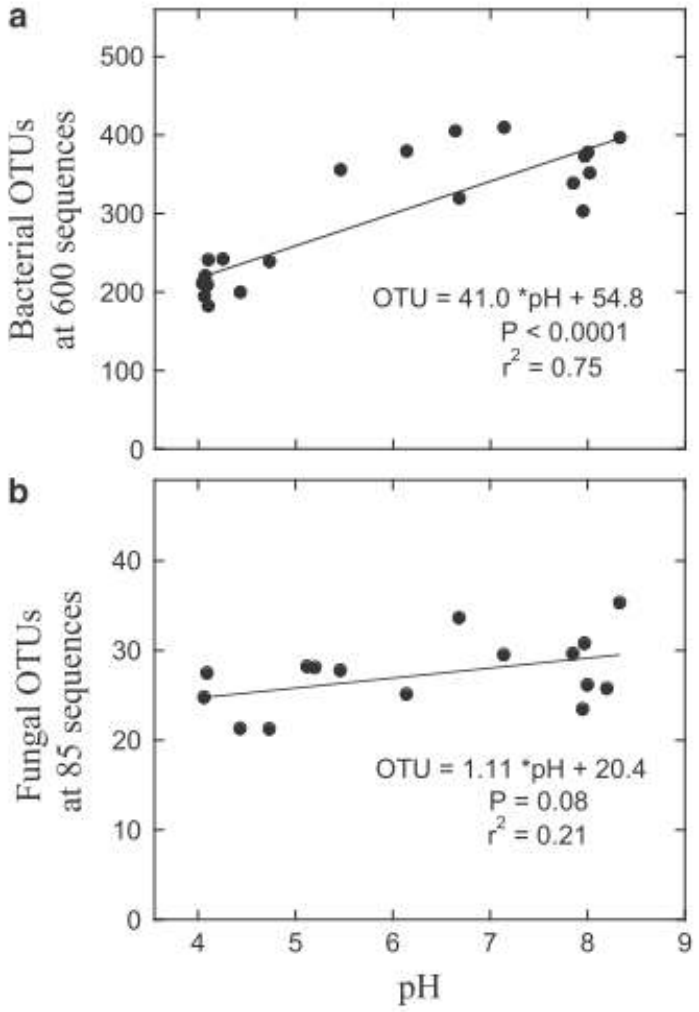
- Now: **barcoded pyrosequencing** of bacterial 16S and fungal 18S

- Advantage of Hoosfield: pH gradient, most everything else is well controlled


Samples
Sequence
97% OTU construction
Taxonomic classification
OTU 1
OTU 2
OTU 3
OTU 4
OTU 1
OTU 2
OTU 3
OTU 4
A
B
C
D
Phylogenetic trees
Diversity (unweighted UniFrac)

How many OTUs?

Richness vs. pH



Beta-diversity: measuring the similarity of communities

Nonphylogenetic

Jaccard similarity: qualitative

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

= 1 (perfectly similar)

Bray-Curtis dissimilarity: qualitative

$$BC(A,B) = \frac{\sum_k |x_{Ak} - x_{Bk}|}{\sum_k |x_{Ak} + x_{Bk}|}$$

J Dissimilarity = 1 - 1 = **0**

BC dissimilarity = 6/20 = **0.3**

Beta-diversity: measuring the similarity of communities

Nonphylogenetic

Jaccard similarity: qualitative

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

= 1 (perfectly similar)

Bray-Curtis dissimilarity: qualitative

$$BC(A,B) = \frac{\sum_k |x_{Ak} - x_{Bk}|}{\sum_k |x_{Ak} + x_{Bk}|}$$

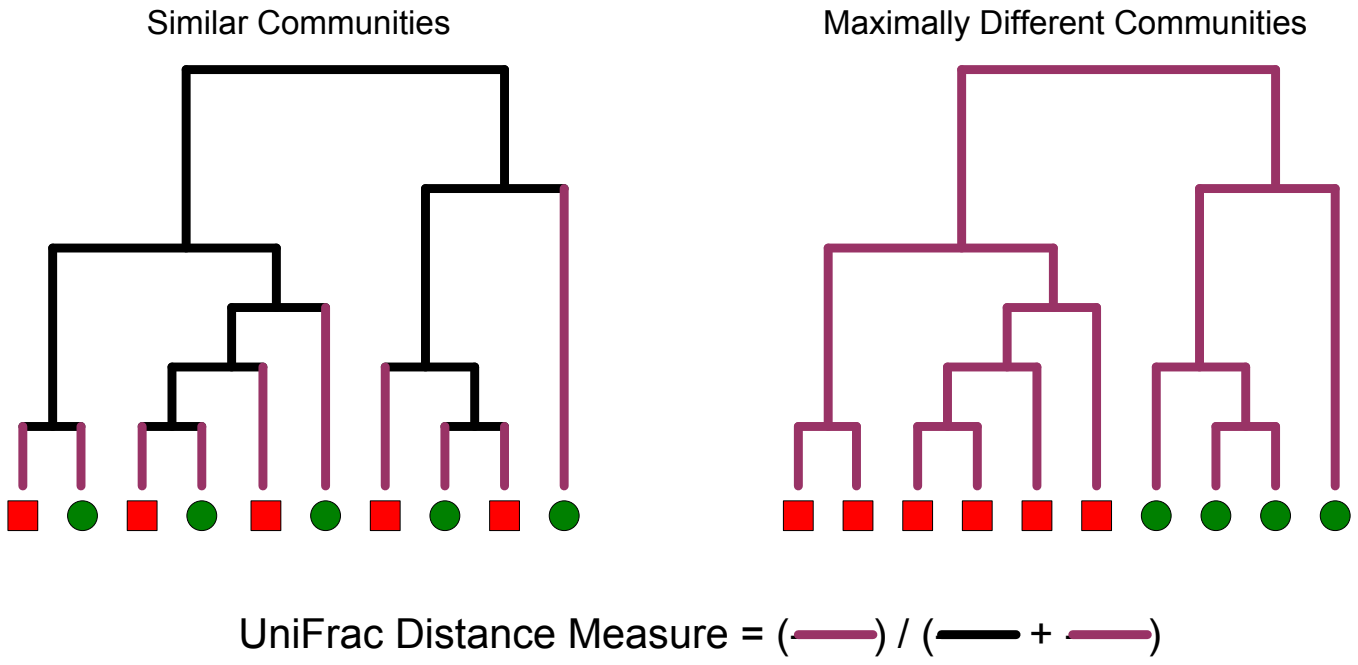
J Dissimilarity = 1 - 1 = **0**

BC dissimilarity = 6/20 = **0.3**

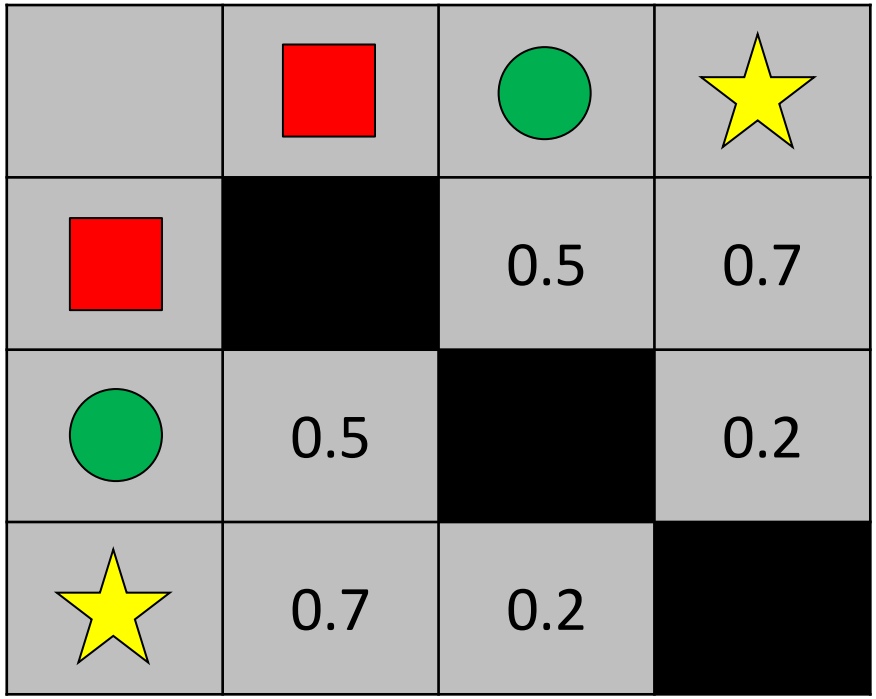
Beta-diversity: measuring the similarity of communities

Phylogenetic: e.g., Unweighted UniFrac

- Black = shared
- Purple = **Uni**que

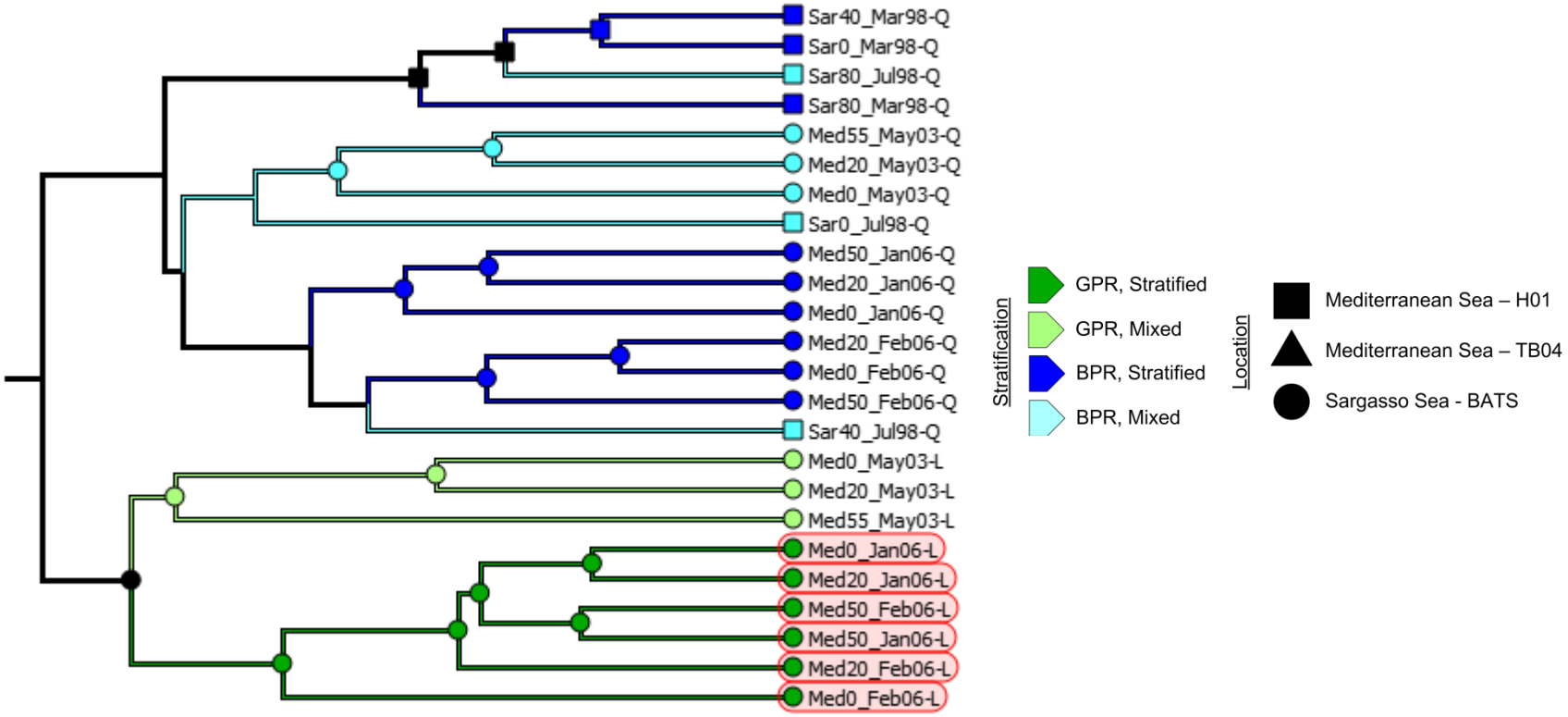


Beta-diversity matrix

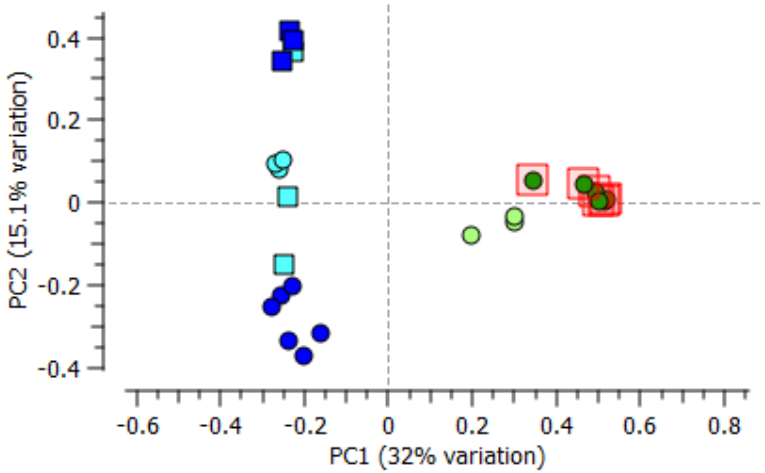


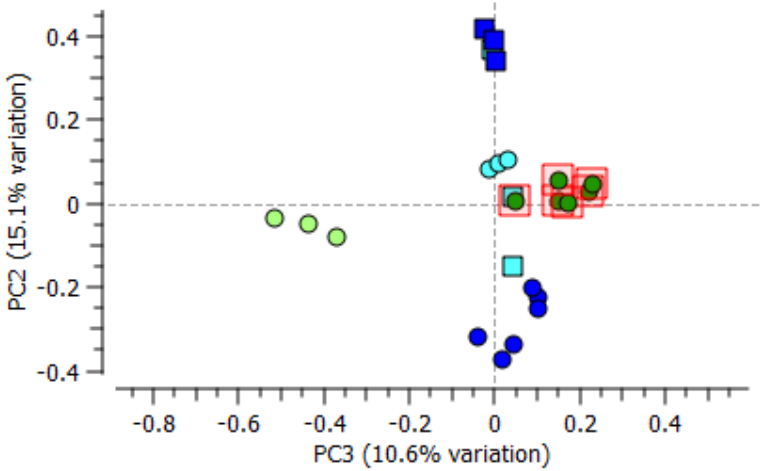
What can we do with a matrix?

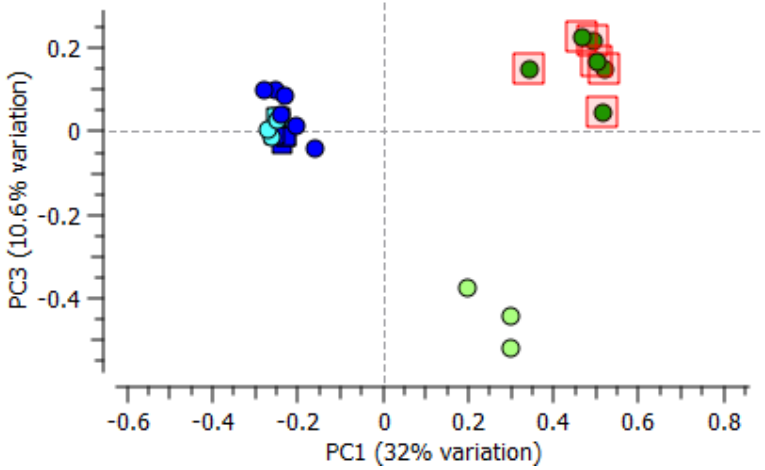
Cluster sites based on similarity blue vs green-absorbing proteorhodopsins



Ordination (PCoA)

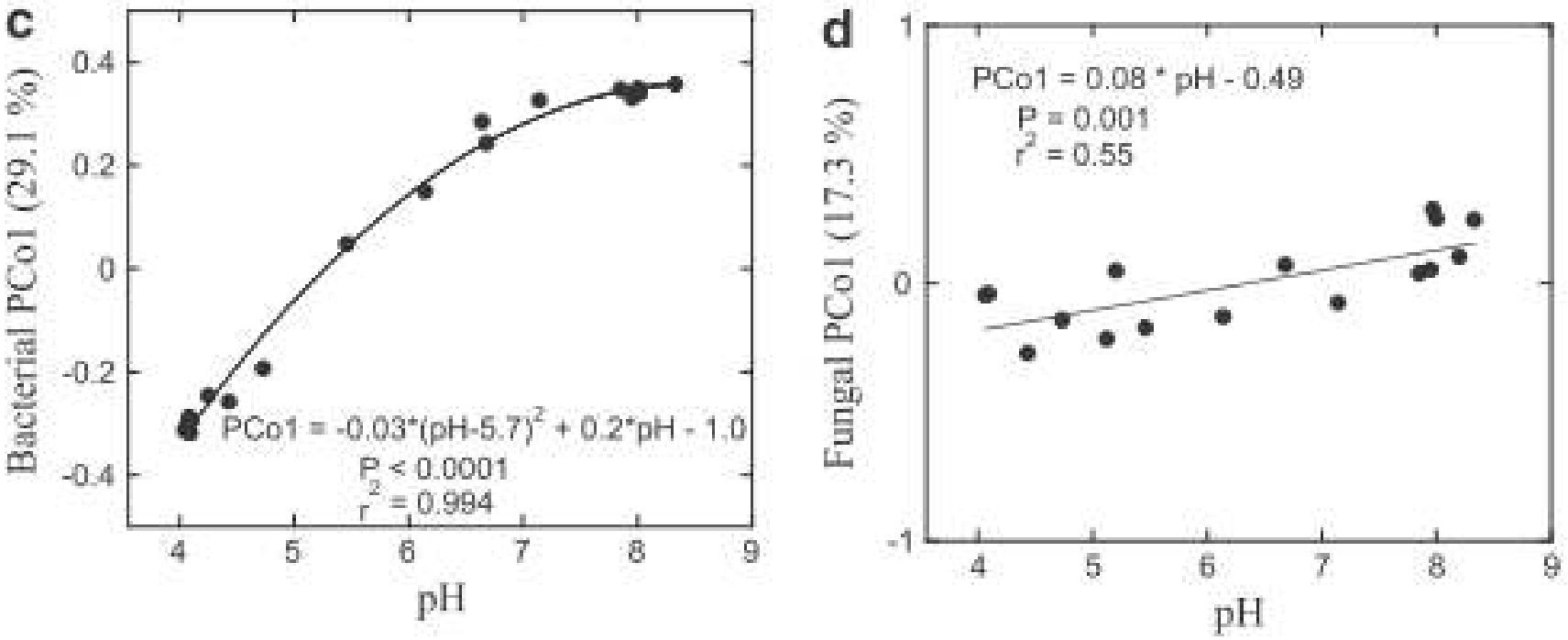




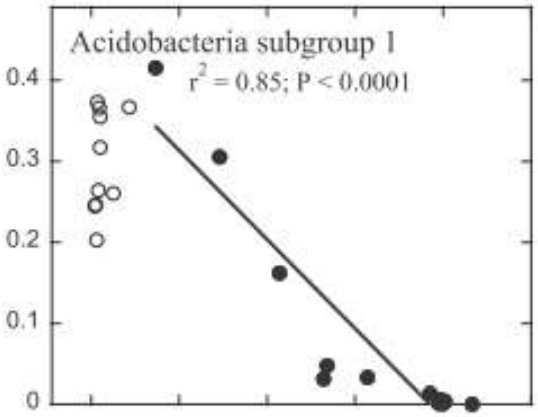


Fierer results:

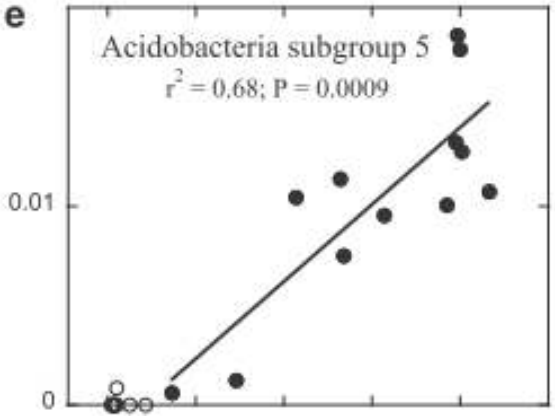
UniFrac ordination vs. pH



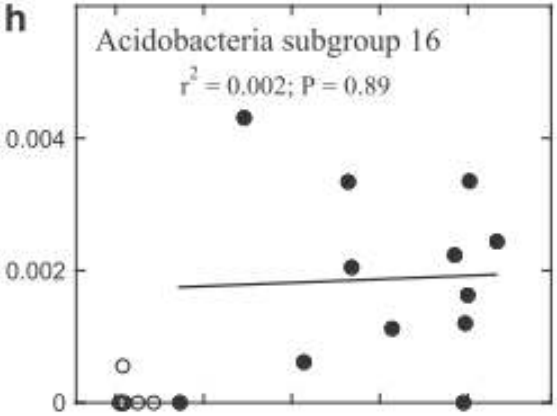
Taxonomy: Acidobacteria subgroups



Intuitive



Counterintuitive



Meh

Conclusions

- Bacteria are more strongly influenced by pH than are fungi
  - Wider tolerance ranges for specific species / communities?
- Communities from similar pH levels tend to be more similar
- Different taxonomic groups go sharply up or down in response to pH, others are unaffected
  - Although there could still be gradients within a particular group (ecological non-homogeneity)

Example #2: Gut microbiomes, *C. difficile* and bacteriotherapy

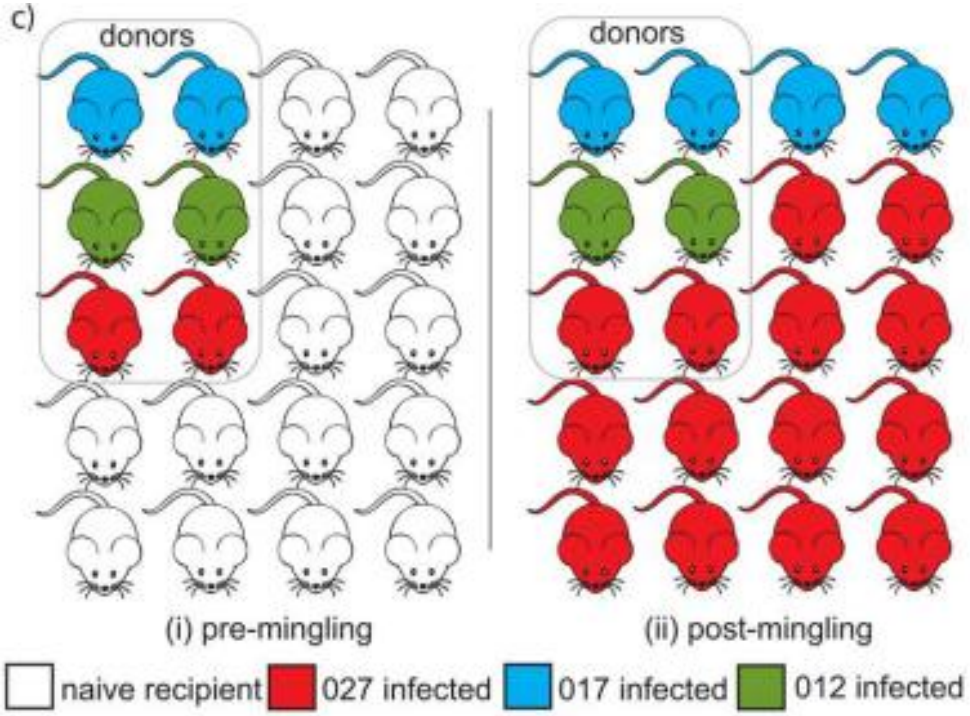
OPEN ACCESS Freely available online

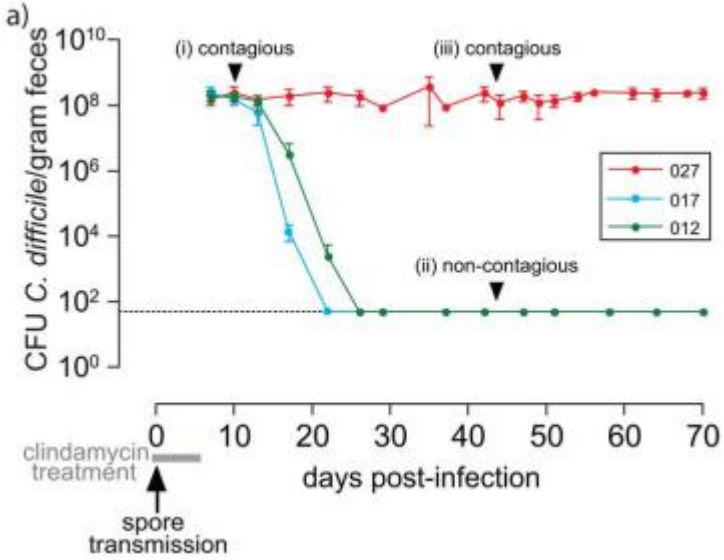
PLOS | PATHOGENS

Targeted Restoration of the Intestinal Microbiota with a Simple, Defined Bacteriotherapy Resolves Relapsing *Clostridium difficile* Disease in Mice

Trevor D. Lawley[1*], Simon Clare[1,2], Alan W. Walker[1,2], Mark D. Stares[1], Thomas R. Connor[1], Claire Raisen[1], David Goulding[1], Roland Rad[1], Fernanda Schreiber[1], Cordelia Brandt[1], Laura J. Deakin[1], Derek J. Pickard[1], Sylvia H. Duncan[2], Harry J. Flint[2], Taane G. Clark[3], Julian Parkhill[1], Gordon Dougan[1]

How to make mice sick





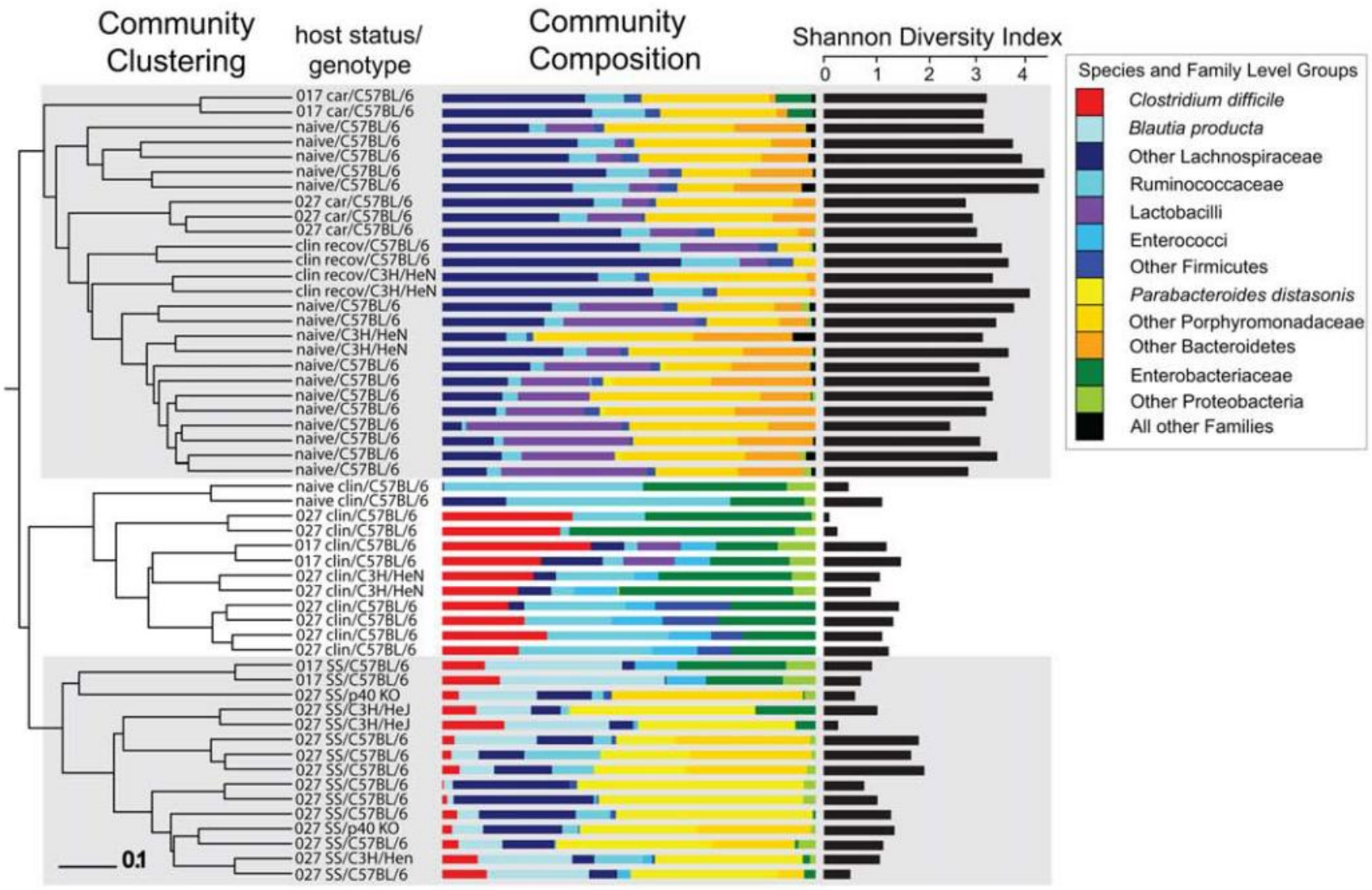
Donors infected with several strains
Recipients hit with clindamycin and exposed to donors

How to make them better - maybe

Bacterial culture

Assessing microbial communities

- Amplfy V2-V5 region of 16S
- Taxonomic assignment using best BLAST matching to reference database
- 98% OTUs:
  - Alpha-diversity (Shannon)
  - Beta-diversity (Bray-Curtis)
  - **Cluster** by beta-diversity

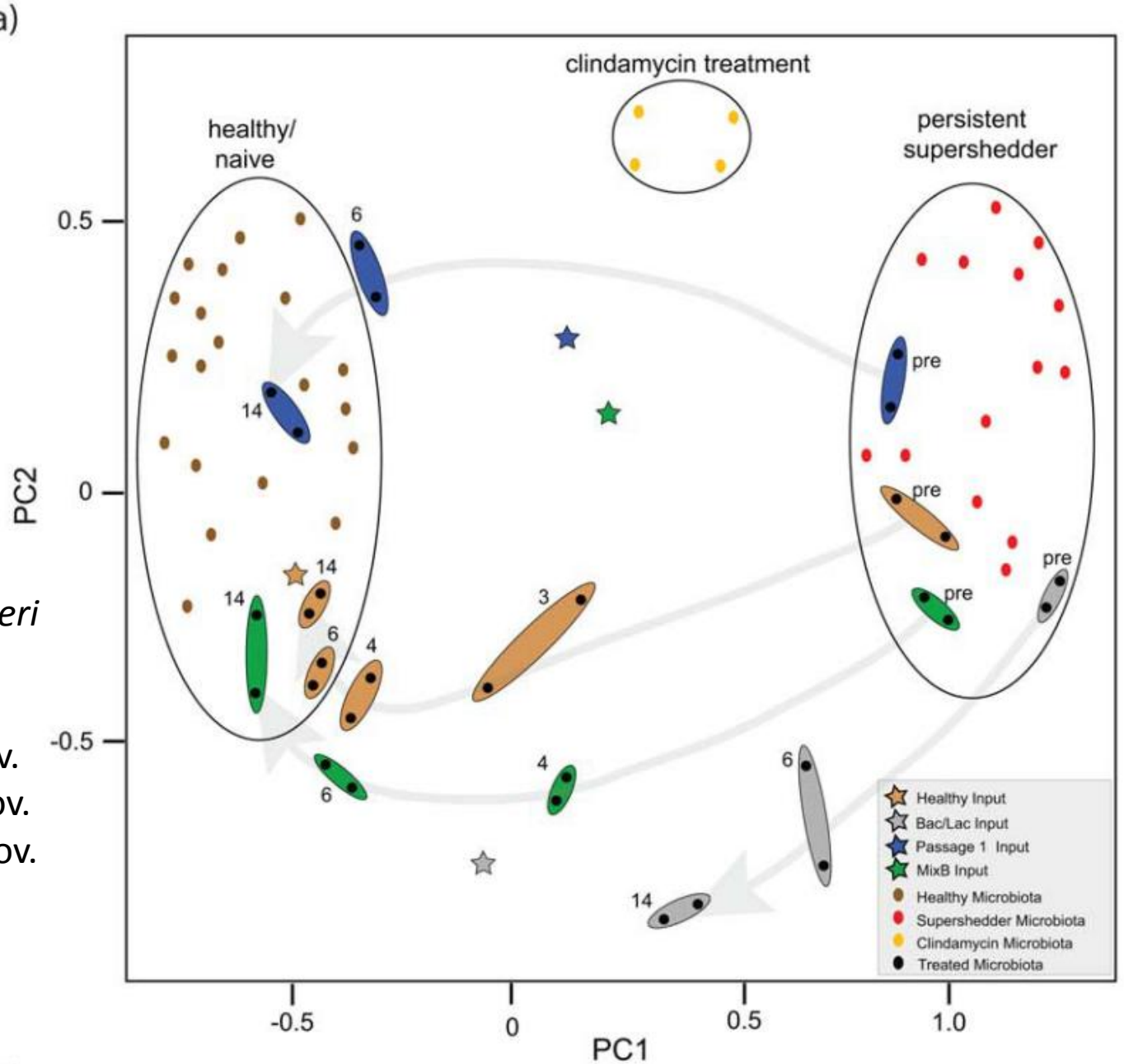


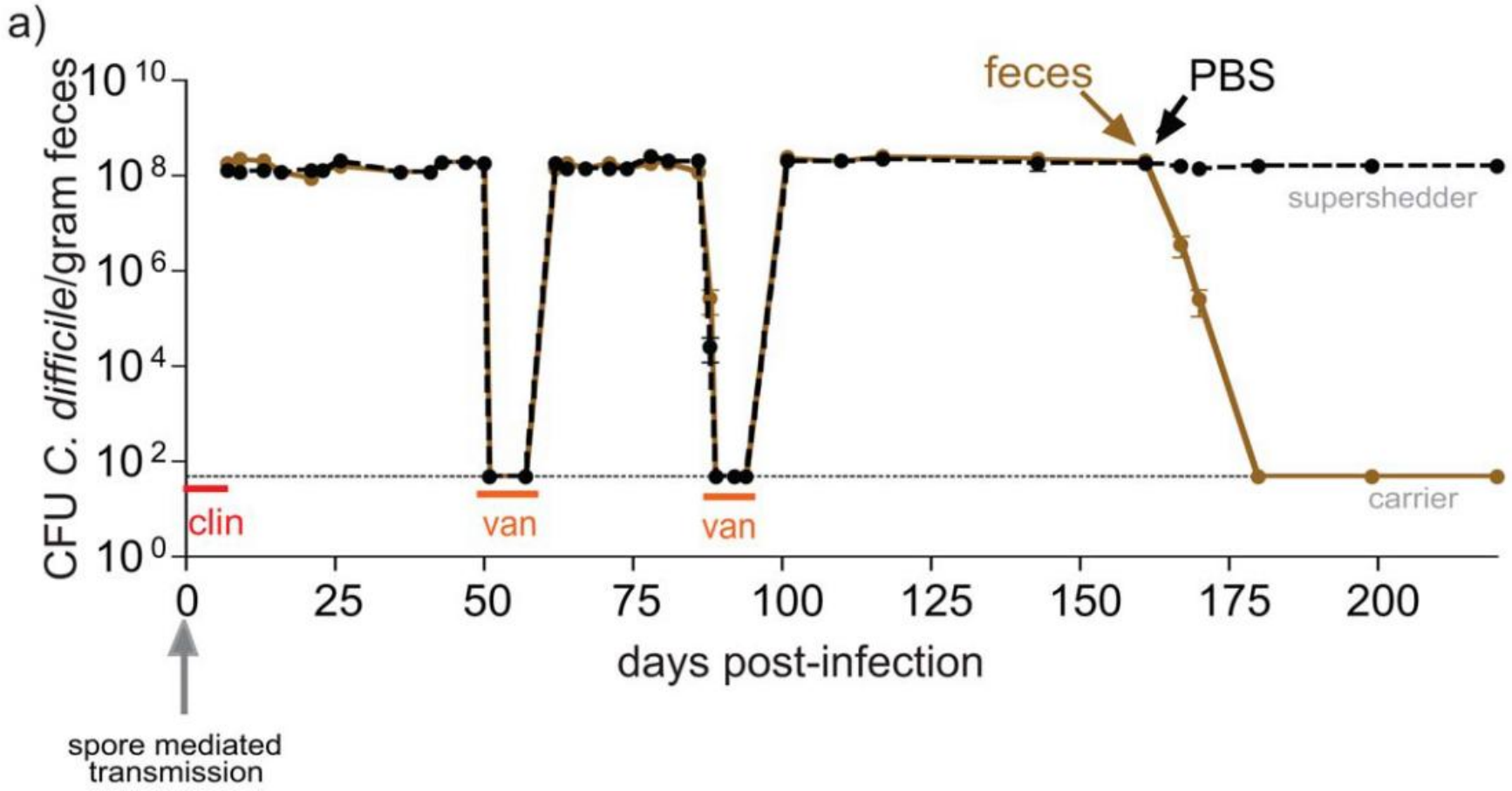
Beta-diversity

Alpha-diversity

97% OTUs: PCA from OTU-by-sample matrix

MixB:
-*Staphylococcus warneri*
-*Enterococcus hirae*
-*Lactobacillus reuteri*
-*Anaerostipes* sp. nov.
-*Bacteroidetes* sp. nov.
-*Enterorhabdus* sp. nov.

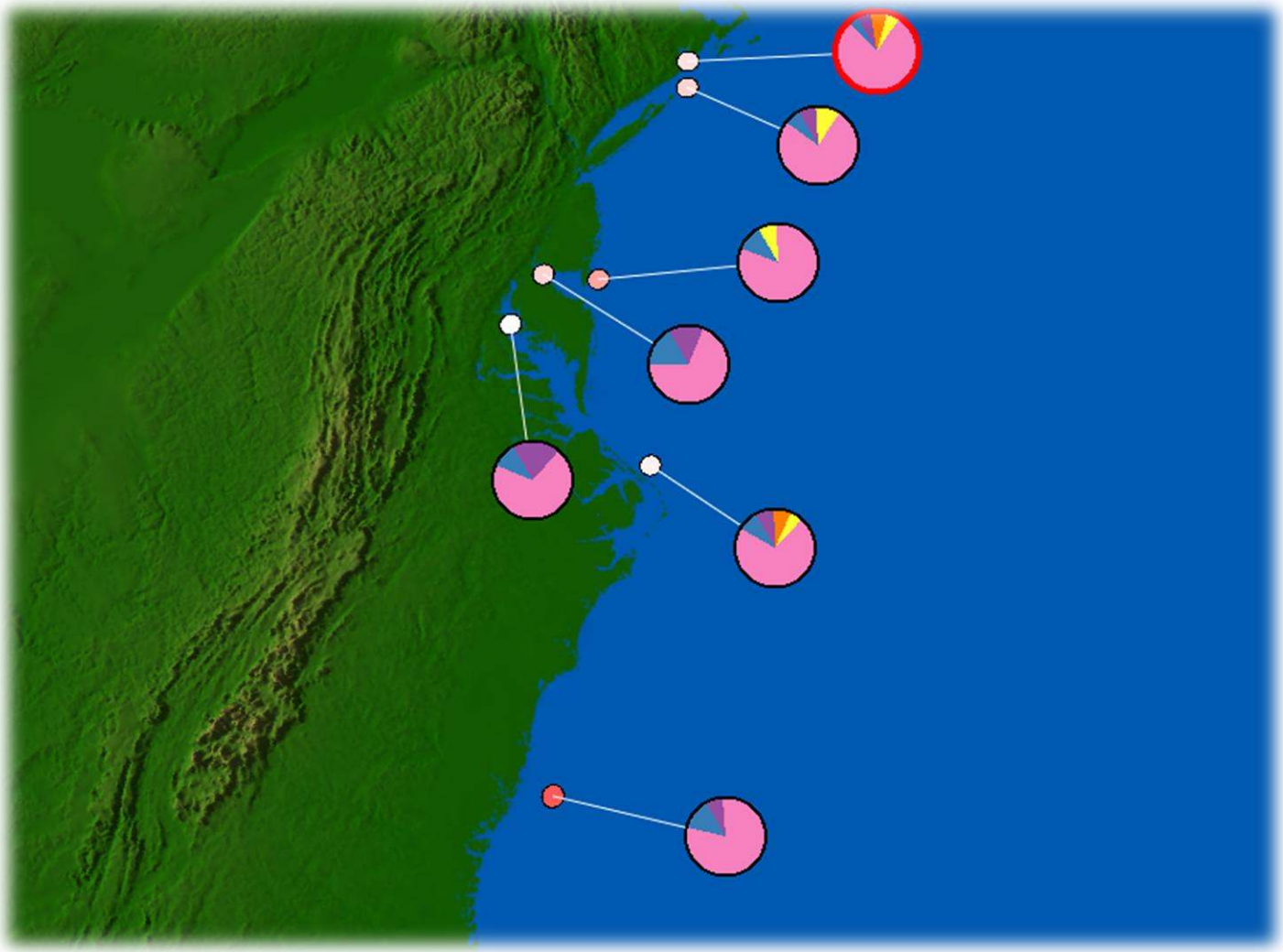



a)
CFU C. difficile/gram feces
10 10
10 8
10 6
10 4
10 2
10 0
feces
PBS
supershedder
carrier
clin
van
van
0
25
50
75
100
125
150
175
200
days post-infection
spore mediated transmission

Summary

- *C. difficile* 027-infected mice exhibit dysbiosis, and are "super-shedders"
- Antibiotic treatment opens the door for latent or environmental *C. diff* to take over
- *Bacteroides* + *Lactobacillus* is not sufficient to restore a healthy state
- A mix of 6 commensals, or an entire "healthy" community, is
- Mouse models allow the testing of reduced mixtures of bacteria (i.e., fewer opportunistic pathogens than a typical full poop sample)

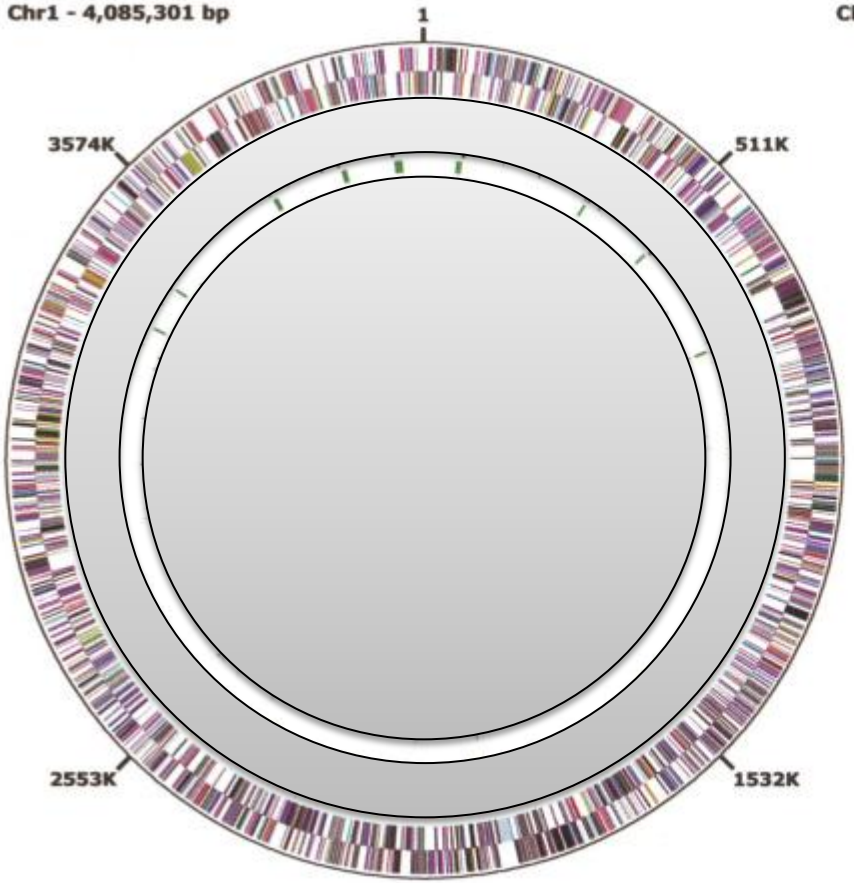
Example 3:
Global Ocean Sampling expedition



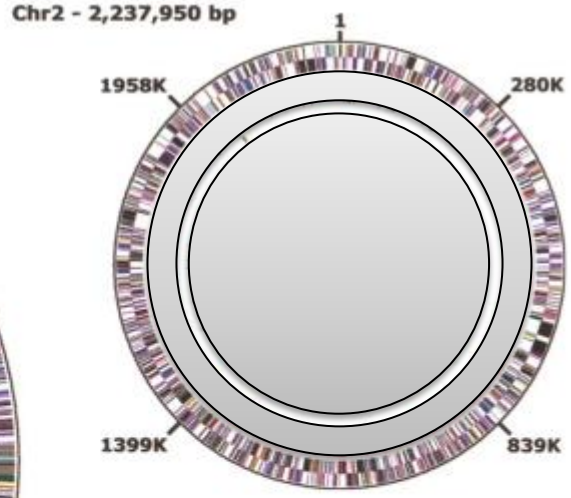
(GenGIS demo)

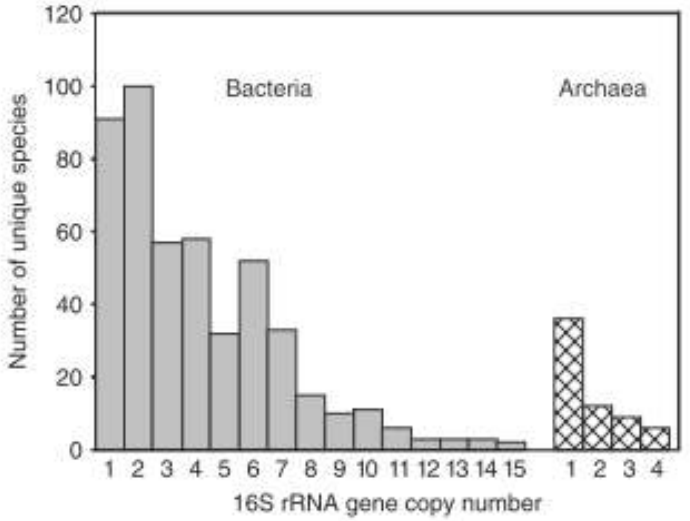
Why 16S is not perfect

Multiple marker genes



Vezzi et al. (2005) *Science*

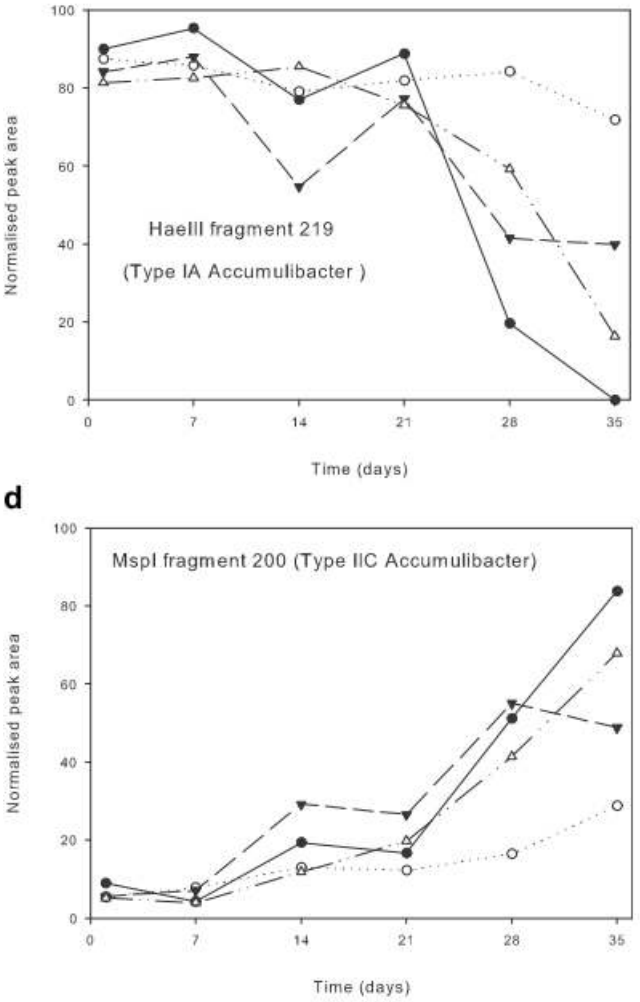




Lee et al. (2009) *Nucleic Acids Res*

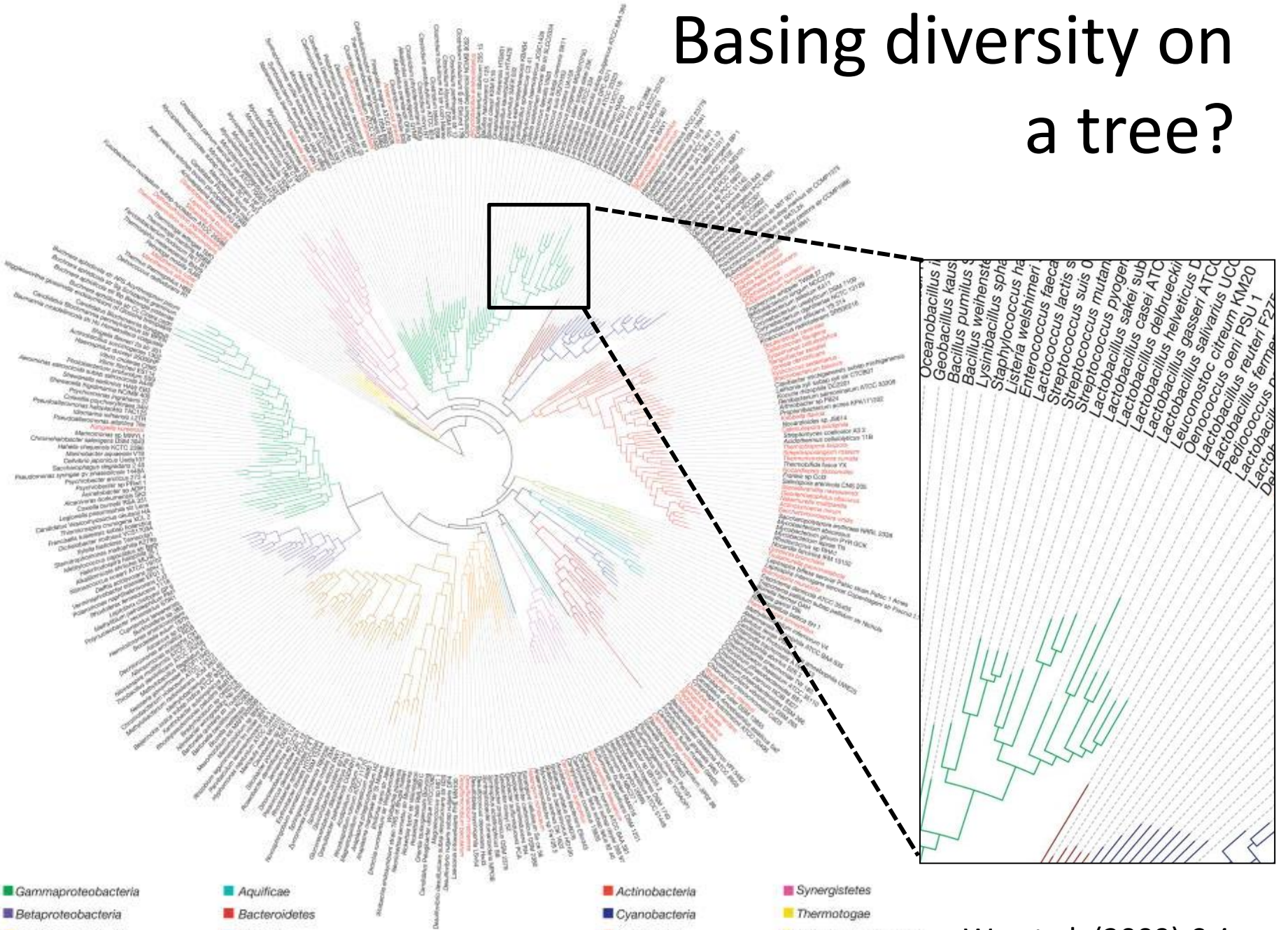
Short fragments

- May give insufficient information for taxonomic assignment
- Evolve too rapidly to accurately represent distantly related taxa
- Evolve too *slowly* to distinguish closely related strains



A. phosphatis and *ppk* as marker
Slater et al. (2010) *Water Res*

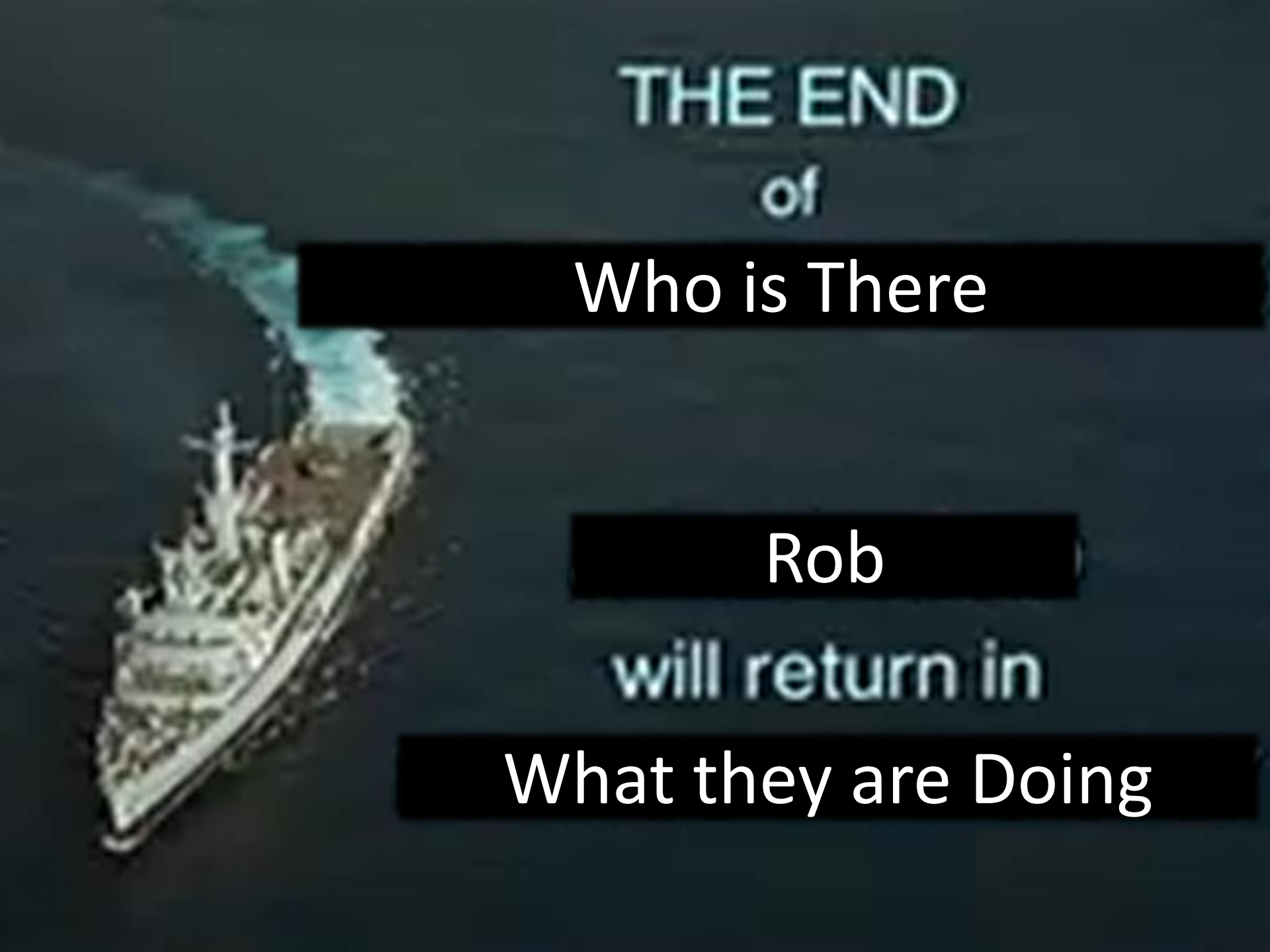
Basing diversity on a tree?



Wu et al. (2009) *Science*

Software and resources for community analysis (a sampling)

- QIIME: http://qiime.org/
- mothur: http://www.mothur.org/
- Ribosomal Database Project: http://rdp.cme.msu.edu/
- Express Beta Diversity: http://kiwi.cs.dal.ca/Software/ExpressBetaDiversity


THE END
of
Who is There
Rob
will return in
What they are Doing