Rob's various things

Contact details:

beiko@cs.dal.ca

@rob_beiko
“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."

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

Carl Woese
(1928-2012)
Historical views of microorganisms
1995-present

Genome sequencing: I'm so confused!
The "Tree of Life"

vs.

Networks of microbes

Norman Pace

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

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)
<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 1/8 cups all-purpose flour</td>
<td>1/2 cup vegetable oil</td>
</tr>
<tr>
<td>2 cups white sugar</td>
<td>1 tablespoon vanilla extract</td>
</tr>
<tr>
<td>3/4 cup unsweetened cocoa powder</td>
<td>2 (20 ounce) cans pitted sour cherries</td>
</tr>
<tr>
<td>1 1/2 teaspoons baking powder</td>
<td>1 cup white sugar</td>
</tr>
<tr>
<td>3/4 teaspoon baking soda</td>
<td>1/4 cup cornstarch</td>
</tr>
<tr>
<td>3/4 teaspoon salt</td>
<td>1 teaspoon vanilla extract</td>
</tr>
<tr>
<td>3 eggs</td>
<td>3 cups heavy whipping cream</td>
</tr>
<tr>
<td>1 cup milk</td>
<td>1/3 cup confectioners' sugar</td>
</tr>
</tbody>
</table>
Can we have our cake?
Ribosomal RNA genes: taxonomic and phylogenetic markers

GGATCCAGGCTCTGCTAAACCAGGGCGGATCTCGGCTGGCTACACCAG...

TGAACTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCT...

ATTCCGGTTGATCCTGCCGGAGCCATTGCTATCGGAGTCCGATTTAGCC...
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


Pseudomonas

How does this help us?
"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
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

Profiling the microbial community
Genetic diversity in Sargasso Sea bacterioplankton

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


Nature, 1990
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

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

Multiple DNA samples

PCR amplify w/unique barcodes

Pool, next-gen sequence

De-multiplex and process
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

% dissimilarity between sequences

3% 10% 11%

> A

> B

> C

> D
1. Taxonomic assignment (SUPERVISED analysis)

- **Clostridium**
  - A
  - B
- **Coprococcus**
  - C
  - D
- **Lactobacillus**
- **Streptococcus**

Homology (best BLAST match)
Composition (k-mer similarity)
2. *de novo* OTU clustering

(unsupervised analysis)

Cluster based on sequence identity

OTU 1

A
B

OTU 2

C
D

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!
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

pH was the only property to display a strong relationship

\[ H' = - \sum_{i=1}^{R} p_i \log p_i \]
The Hoosfield acid strip

pH 8.3
Bacteria / Fungi

pH 4.0
Fungi / Bacteria

180 m
• 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

Diversity (unweighted UniFrac) → Phylogenetic trees
How many OTUs?
Richness vs. pH

![Bacterial OTUs vs. pH](chart_a)

**Bacterial OTUs at 600 sequences**

- OTU = 41.0 \* pH + 54.8
- \( P < 0.0001 \)
- \( r^2 = 0.75 \)

![Fungal OTUs vs. pH](chart_b)

**Fungal OTUs at 85 sequences**

- OTU = 1.11 \* pH + 20.4
- \( P = 0.08 \)
- \( r^2 = 0.21 \)
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** = *Unique*

\[
\text{UniFrac Distance Measure} = \frac{\text{Similar Communities}}{\text{Maximally Different Communities}}
\]
### Beta-diversity matrix

<table>
<thead>
<tr>
<th></th>
<th>Red</th>
<th>Green</th>
<th>Yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green</td>
<td>0.5</td>
<td></td>
<td>0.7</td>
</tr>
<tr>
<td>Yellow</td>
<td>0.5</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.2</td>
<td></td>
</tr>
</tbody>
</table>

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

\[ PCo1 = -0.03 \cdot (pH - 5.7)^2 + 0.2 \cdot pH - 1.0 \]
\[ P < 0.0001 \]
\[ r^2 = 0.994 \]

\[ PCo1 = 0.08 \cdot pH - 0.49 \]
\[ P = 0.001 \]
\[ r^2 = 0.55 \]
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

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

Trevor D. Lawley¹*, Simon Clare¹*, Alan W. Walker¹*, Mark D. Stares¹, Thomas R. Connor¹, Claire Raisen¹, David Goulding¹, Roland Rad¹, Fernanda Schreiber¹, Cordelia Brandt¹, Laura J. Deakin¹, Derek J. Pickard¹, Sylvia H. Duncan², Harry J. Flint², Taane G. Clark³, Julian Parkhill¹, Gordon Dougan¹
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

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

MixB:
- *Staphylococcus warneri*
- *Enterococcus hirae*
- *Lactobacillus reuteri*
- *Anaerostipes sp. nov.*
- *Bacteroidetes sp. nov.*
- *Enterorhhabdus sp. nov.*
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
Basing diversity on a tree?

Wu et al. (2009) Science
Software and resources for community analysis (a sampling)

- Ribosomal Database Project: [http://rdp.cme.msu.edu/](http://rdp.cme.msu.edu/)
- Express Beta Diversity: [http://kiwi.cs.dal.ca/Software/ExpressBetaDiversity](http://kiwi.cs.dal.ca/Software/ExpressBetaDiversity)
THE END
of
Who is There
Rob
will return in
What they are Doing