• Genetic information and phylogenetic markers (G5)
• Phylogenetic trees (G6)
• Ribosomal RNA gene cloning and sequencing (G7)
• Survey of methods for fingerprinting microbial communities (G8)
Microbial Ecology: What is it we really want to know??

• Who is there?
• How many of each “type” are present at any given time and location?
• What are they doing (and, what resources are they using to do it)?
• How fast are they doing it?
Molecular Microbial Ecology

Ecology without cultivation

• Use molecular biology techniques to:
  - Identify microbial “species” & describe distributions
  - Predict metabolic capacities from genetic information
  - Link microbial “species” with system function; assign functional roles
Genetic Information and Phylogenetic Markers (G5)
Why classify through DNA?

- Present in all known organisms. DNA stores information in genes.
- Genes are discreet sequences of nucleotides. The 4 nucleotides in DNA are adenine (A), thymine (T), guanine (G), and cytosine (C) [AGCT]
- Information in genes is transcribed (written) into ribonucleic acids (RNA). These contain uracil (U) instead of thymine (T) [AGCU]
- The ‘message’ in the RNA is read (translated) and proteins synthesized in ribosomes
Gene Density in the Bacterium *Mycobacterium tuberculosis*
• In a perfect world, we would manipulate and compare whole genomes to determine relationships among microbes
• Not yet feasible
• Must substitute a phylogenetic marker: a gene who’s sequence is used to infer phylogenetic relationships among microbes
• MAJOR ASSUMPTION: gene phylogeny more or less reflects the evolutionary history of the microbes possessing the gene of interest
**Ribosomes**
- Sites of protein synthesis
- Usually 10,000 to 20,000 ribosomes per cell
- May occupy 25% of cell volume and use 90% of the cell’s energy
- Differentiated on the basis of size into large and small subunits

**Ribosomal RNA and taxonomy**
- Ribosomes are critical to cell function, so their structural RNAs should not evolve rapidly - a sequence change may disable the ribosome
- Ribosomal gene sequences are conserved, i.e., they do not change much over time
Why the 16S Ribosomal RNA?

- Universally conserved: possessed by all cellular forms of life (Eukarya, Archaea, and Bacteria)
- With very few exceptions, is not laterally transferred
- Mosaic of highly conserved and highly variable regions – broad range of utility in phylogenetic analyses
- Substitutions generally considered to be neutral (or close to it)
- Conserved secondary structure useful for detecting PCR and sequencing artifacts/errors
- Estimated divergence rate: 1% per 50 million years (can vary by an order of magnitude.............)
Ribosomal RNA and Taxonomy

• In the late 1970s, Carl Woese and colleagues studied relationships among prokaryotes through the comparison of rRNA gene sequences.

• Surprise... Not all prokaryotes are related! Archaea sequences were as unrelated to the Bacteria as they are to eukaryotes (Eukarya).

• We now routinely investigate phylogenetic relationships between prokaryotes by comparing nucleotide sequences (AGCT) of their 16S (small subunit) rRNA genes.
Isolation of bacterium from intestine of warm-blooded animal

- Obtain pure culture
- Gram reaction
- Gram-negative
- Rod-shaped
- Facultative
- Ferments lactose, producing acids and gas

Perform biochemical tests:
(Positive: indole, methyl red, mucate; negative: citrate, Voges-Proskauer, H₂S)

Escherichia coli

Organism

- Isolate DNA
- 16S rRNA gene
- Heat to separate strands; add specific primers
- Primer extension with DNA polymerase
- Repeat above steps for many PCR cycles to yield multiple copies of 16S ribosomal RNA gene
- Run agarose gel and check for correct sized PCR product
- Purify PCR product
- Sequence by Sanger method
RNA gene sequencing and taxonomic groups

Organisms group together if their 16S rRNA sequences are most similar.

Differences suggest more phylogenetic distance, and the further apart the organisms are in an evolutionary sense: more distant from their last common ancestor.
Phylogenetic Trees (G6)
(a) Cells → Isolate DNA

(b) DNA → Gene encoding ribosomal RNA

(c) DNA Sequencing

(d) Sequence analysis: AGTCGCTAG, ATTCGTTAG, AGCCGT TAG

(e) Generate phylogenetic tree
PCR is a way of finding a needle in a haystack and subsequently producing a pile of needles from the hay.

For example, we are looking for a specific 300 bp strand of DNA in a sequence of ~3,000,000,000 bp

If you want to manipulate or see DNA, you need LOTS of it!
Polymerase Chain Reaction (PCR)

Denature DNA (1 double strand to 2 single strands) and attach primers

Enzyme (DNA Polymerase) creates complementary strands from free nucleotides

Repeat

Geometric increase in PCR product

30 cycles of PCR creates $2^{30}$ copies of the original template > 1,000,000,000 copies
Cells → Isolate DNA → PCR → Gene encoding ribosomal RNA → DNA sequencing

DNA sequencing:
- AGTCGCTAG
- ATTCGTAG
- AGCCGTTAG

Sequence analysis → Generate phylogenetic tree

(e)
• Compare sequences ‘pair-wise’
• Trees are drawn based on degree of similarity between sequences
• The exact branching order depends on the algorithm (converts similarity to evolutionary distance)
• Trees only indicate potential relationships (gene tree vs. organismal tree)

A is more closely related to C than to B or D

A and C (as well as B and D) diverged more recently than A and B
A. Sequences

sequence A  ACGC GTTG GCG ATG GCA AC
sequence B  ACGC GTTG GCG ACG GTA AT
sequence C  ACGCATTTGA ATGATGATAAT
sequence D  ACACATTGAGTGATAATAAT

B. Distances between sequences, the number of steps required to change one sequence into the other.

n_{AB}  3
n_{AC}  7
n_{AD}  8
n_{BC}  6
n_{BD}  7
n_{CD}  3

C. Distance table

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>3</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>B</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>C</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

D. The assumed phylogenetic tree for the sequences A-D showing branch lengths. The sum of the branch lengths between any two sequences on the trees has the same value as the distance between the sequences.
Figure 6.5. Structure of evolutionary trees.
Table 6.2. Number of possible evolutionary trees to consider as a function of number of sequences

<table>
<thead>
<tr>
<th>Taxa or sequence no.</th>
<th>No. of rooted trees</th>
<th>No. of unrooted trees</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>105</td>
<td>15</td>
</tr>
<tr>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>7</td>
<td>10,395</td>
<td>954</td>
</tr>
</tbody>
</table>

![Diagram of evolutionary tree with taxa A, B, C, and D]
Ribosomal RNA Gene Cloning and Sequencing From Environmental Samples (G7)
## Cultivability of Microorganisms From Different Natural Environments

<table>
<thead>
<tr>
<th>Habitat</th>
<th>Cultivability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seawater</td>
<td>0.001 – 0.01</td>
</tr>
<tr>
<td>Freshwater</td>
<td>0.01 - 1</td>
</tr>
<tr>
<td>Unpolluted estuarine waters</td>
<td>0.1 - 3</td>
</tr>
<tr>
<td>Sediment</td>
<td>~0.25</td>
</tr>
<tr>
<td>Soil</td>
<td>~0.3</td>
</tr>
</tbody>
</table>
Are the microorganisms that are easily cultured in the laboratory using traditional techniques:

representative of the numerically dominant species in native bacterioplankton communities, but possess low cultivation efficiencies (e.g. a high incidence of inactive cells?),

OR

are a small subset of the community and the cells not accounted for (>99.9% of the total) are completely different taxa/species?
Molecular Microbial Ecology: The Ribosomal RNA Approach

- Laboratory of Norm Pace (Pace et al. 1986)
- Clone and sequence small subunit (16S) ribosomal RNA genes from the natural environment to determine phylogenetic diversity of microorganisms present
- Serves as a phylogenetic marker for comparing microorganisms (cultured isolates and environmental gene clones)
- Giovannoni et al. 1990: first study to PCR-amplify 16S ribosomal RNA genes from a natural environment (Sargasso Sea seawater)
RNA Gene Cloning and Sequencing from Seawater

- Filter seawater through a 0.2 µm pore-sized filter (or less) to harvest bacterioplankton cells
- Extract genomic DNA from the bacterioplankton collected on the filter
- Use the polymerase chain reaction and conserved oligonucleotide primers to amplify 16S ribosomal RNA genes from the mixed community DNA (1 genome = at least one 16S rRNA gene)
- Clone the PCR product into a plasmid vector
- Sequence clone inserts (one clone = one 16S rRNA gene in the original community)
- Sequence analysis
Environmental Clone Library

Extract DNA

Grow "environmental clones" on plates

Transform E. coli cells with plasmids

PCR 16S rRNA genes with bacterial specific primers

Ligate 16S rRNA genes into plasmids
Environmental Clone Libraries

• Advantages
  - Can get full-length sequence data for phylogenetic characterization
  - Easy to do
  - Libraries can be screened in several ways

• Disadvantages
  - Expensive to sequence
  - Difficult to analyze many different samples
Link Environmental Gene Clones With Microbes of Known Characteristics
Graphs Depicting the Number of 16S rRNA Gene Sequences Published in GenBank Since 1993

Rappé and Giovannoni 2003
Survey of methods for fingerprinting microbial communities (G8)
Microbial community

Extract DNA

Total community DNA

PCR

Amplify 16S rRNA genes using general or specific primers

Sample 1 2 3 4

All 16S rRNA genes

Excise bands and clone 16S rRNA genes

Different 16S rRNA genes

Sample 1 2 3 4

DGGE

Excise bands

Sequence (a)
Terminal Restriction Fragment Length Polymorphism analysis (T-RFLP)


Distinguishes organisms using length variability of one end of a gene after it is cut into several fragments with restriction endonucleases
Restriction Endonucleases

- Known as “molecular scissors”
- There are thousands of known restriction endonucleases, all of which recognize and bind to very specific base sequences in dsDNA
- Type II endonucleases: Cut dsDNA somewhere within the recognition site
- T-RFLP generally uses Type II restriction enzymes that recognize sequences of four nucleotides (“4-cutters”)
- Several enzymes are often used
Two Type II restriction endonucleases that recognize sequences of 4 nucleotides:
Terminal Restriction Fragment Length Polymorphism Analysis (TRFLP)

1. Extract DNA
2. PCR 16S rRNA genes with labelled primer
3. Cut DNA with Restriction Enzymes
4. Gel Electrophoresis of terminal restriction fragments
Restriction fragments are usually analyzed on an automated DNA sequencer.

Only labeled fragments are detected.

Standards are also run to accurately size the DNA fragments.
T-RFLP

- Advantages
  - Relatively easy to do
  - Results can be banked for future comparisons

- Limitations
  - Less sensitive phylogenetic resolution than sequencing
  - Each fragment length can potentially represent a diversity of microorganisms
  - Cannot directly sequence restriction fragments, making identification indirect
Denaturing Gradient Gel Electrophoresis (DGGE)


Distinguishes organisms using the variable melting characteristics of dsDNA with different nucleotide sequences
Denaturing Gradient Gel Electrophoresis (DGGE): Fingerprinting bacterial communities

1. Extract DNA
2. PCR 16S rRNA genes with bacterial specific primers containing GC-clamp
3. Electrophoresis at constant temperature on acrylamide gel containing a vertical gradient of denaturants (urea and formamide)
4. GC-CLAMP
5. 16S-rRNA

DNA stops moving on gel when it reaches its critical denaturant concentration

20% denaturant

70% denaturant

5' - CGGCCGCGCGCGCGCGCGGCGGGCGGGGGCGGGGCACGGGGGGCCTACGGGAGGCAGCAG~
3' - CGGGCGCGCGCGCGCGCGCGCGCGGCGCGGGCGGGGGCGGGGGGTCGCCCTCAGCGTC~
DGGE gel, Plum Island Sound, July 2000

<table>
<thead>
<tr>
<th>Parker River</th>
<th>0 ppt</th>
<th>0.5 ppt</th>
<th>3 ppt</th>
<th>6 ppt</th>
<th>12 ppt</th>
<th>19 ppt</th>
<th>26 ppt</th>
<th>31 ppt</th>
<th>Coastal Ocean</th>
</tr>
</thead>
<tbody>
<tr>
<td>whole</td>
<td>whole</td>
<td>whole</td>
<td>whole</td>
<td>whole</td>
<td>whole</td>
<td>whole</td>
<td>whole</td>
<td>whole</td>
<td>whole</td>
</tr>
<tr>
<td>&lt;1 μm</td>
<td>&lt;1 μm</td>
<td>&lt;1 μm</td>
<td>&lt;1 μm</td>
<td>&lt;1 μm</td>
<td>&lt;1 μm</td>
<td>&lt;1 μm</td>
<td>&lt;1 μm</td>
<td>&lt;1 μm</td>
<td>&lt;1 μm</td>
</tr>
</tbody>
</table>

Ladder

31 ppt

Ladder

6 ppt

Ladder

12 ppt

Ladder

19 ppt

Ladder

26 ppt

Ladder

31 ppt

Parker River

Ladder
Sequencing DGGE Bands

1. Stab middle of band with pipette tip
2. Swirl tip in PCR mix and amplify
3. Clone PCR products
4. Pick and PCR-amplify 8 clones with dgge primers
5. Screen against natural sample on DGGE gel
6. Sequence DNA in all clones that match bands from the original sample

Some will match perfectly
Some will not match any bands
Some will match other bands
DGGE

• Advantages
  - Very sensitive to variations in DNA sequence
  - Can excise and sequence DNA in bands

• Limitations
  - Somewhat difficult
  - "One band-one species" isn’t always true
  - Cannot compare bands between gels
  - Only works well with short fragments (<500 bp), thus limiting phylogenetic characterization
Variants of DGGE

- Temperature gradient gel electrophoresis (TGGE)
  - Same as DGGE but uses temperature gradient instead of chemical gradient
  - Heuer et al. (1997) AEM 63:3233-3241

- Constant denaturant capillary electrophoresis (CDCE)
  - PCR products migrate through capillaries
  - Lim et al. (2003) AEM 67:3897-3903
Length Heterogeneity PCR (LH-PCR) & Automated Ribosomal Intergenic Spacer Analysis (ARISA)

- LH-PCR

- ARISA
  - Distinguishes organisms using length variability in the spacer region between the genes coding for rRNA small and large subunits

Bacterial rRNA operon
Automated ribosomal intergenic spacer analysis (ARISA)

- Extract DNA
- PCR rRNA intergenic spacer region with labeled primers
- PCR products of variable length
- Gel or Capillary Electrophoresis of terminal restriction fragments

Compare banding patterns among samples
FIG. 2. Partial ARISA profiles of the bacterial communities in Crystal Bog Lake (A), Lake Mendota (B), and Sparkling Lake (C) during the summer of 1998. In each panel, the red and black electropherograms represent duplicate PCRs that were performed on a single sample from each site.
ARISA & LH-PCR

- Advantages
  - Relatively easy to do
  - Results can be stored for future comparisons

- Limitations
  - Less sensitive phylogenetic resolution
  - Each fragment length can represent a diversity of microorganisms
  - Cannot directly sequence fragments
Caveats of PCR-based methods

- Can only amplify genes that match primers
- Amplification can be uneven
  - Under-amplification of abundant taxa
  - Overamplification of rare taxa
- Results are conservatively considered an ‘inventory’ rather than a quantitative measure of the abundant PCR-amplifiable organisms
Variations on PCR-Based Techniques

- Can use rRNA instead of DNA
- Make complementary DNA (cDNA) with reverse transcriptase from extracted rRNA
- rRNA biases analyses towards highly active organisms that have a lot of ribosomes
Oligonucleotide Primer Design

Marinobacter aquaeolei
Marinobacter articus
Marinobacter hydrocarbonoclasticus
Oceanospirillum japonicum
Oceanospirillum linum
Oceanospirillum multiglobuliferum
Thiothrix nivea

Marinobacter probe:
5'-GCTGCCCTCCCAGTGAGAGT-3'

5' 228 245 3'
gatcatccgcg...ACTCTACGGGAGGCAGC...agagacgcttg

gagactttgtcg...ACTCTACGGGAGGCAGC...agagacgggta

gatcatccgcg...ACTCTACGGGAGGCAGC...agagacgcttg

5' 228 245 3'
gagactttgtcg...CCTCTACGCAGGGCAAA...agagacgggta

gatcatccgcg...ACTCTACCGGAGGCAGC...agagacgcttg

5' 228 245 3'
gatcatccgcg...ACTCTACCCGAGGCAGC...agagacgcttg

5' 228 245 3'
gatcatccgcg...ACTCTACCGGAGGCAGC...agagacgcttg
DNA is generally double stranded; forms when two complementary strands are parallel.

Nucleotide bases align in such a DNA molecule, with A to T, and G to C.
The Great Plate Count Anomaly

Plate Counts

Direct Counts
“The Great Plate Count Anomaly”
Staley & Konopka, 1985

- There are a LOT of bacteria
- Most bacteria will not grow on a petri plate
- Who are all these bacteria?

Figure 1  Isopleth graph of total acridine orange direct microscopic count of bacteria (cells per ml $\times 10^{-5}$) from Lake Washington during 1977 (32).

Acridine Orange Direct counts (AODC)

2x10^6 cells/ml

Viable plate counts

1x10^3 cells/ml

Figure 2  Isopleth graph of the viable plate count of bacteria (cells per ml $\times 10^{-3}$) from Lake Washington during 1977 (32).
PCR – the polymerase chain reaction
“Xeroxing DNA”

1 copy

Cycle 1
- dNTPs, Buffer, salts, Taq polymerase

Cycle 2
- 2 copies

Cycle 3
- 4 copies

Cycle 35
- 68,719,476,736 copies in ~ 2 hrs

2 copies

4 copies

8 copies
16S Ribosomal RNA Secondary Structural Model
Terminal Restriction Fragment Length Polymorphism (T-RFLP)

16S rDNA

Terminal Restriction Fragments

Terminal Restriction Fragments
Restriction Fragment Length Polymorphism Analysis (RFLP) on an Agarose Gel
Environmental Clone Libraries

- Original technique used to identify the uncultivated majority
- DNA-based studies using PCR & 16S rRNA
  - Giovannoni et al. (1990) Nature 345:60-63
### (a) Sequence alignment and analysis

<table>
<thead>
<tr>
<th>Organism</th>
<th>Sequence</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>CGUAGAACCCUGGAC</td>
<td>For A → B, three differences occur out of a total of twelve; thus $\frac{3}{12} = 0.25$</td>
</tr>
<tr>
<td>B</td>
<td>CCUAGAGCUGGCG</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>CCAAGACUGGGC</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>GCUGAGUGUGGCC</td>
<td></td>
</tr>
</tbody>
</table>

### (b) Calculation of evolutionary distance

<table>
<thead>
<tr>
<th>Evolutionary distance</th>
<th>Corrected evolutionary distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_D$ A → B</td>
<td>0.25 0.30</td>
</tr>
<tr>
<td>$E_D$ A → C</td>
<td>0.33 0.44</td>
</tr>
<tr>
<td>$E_D$ A → D</td>
<td>0.42 0.61</td>
</tr>
<tr>
<td>$E_D$ B → C</td>
<td>0.25 0.30</td>
</tr>
<tr>
<td>$E_D$ B → D</td>
<td>0.33 0.44</td>
</tr>
<tr>
<td>$E_D$ C → D</td>
<td>0.33 0.44</td>
</tr>
</tbody>
</table>

### (c) Phylogenetic tree