Molecular methods to assess microbial diversity

Microbial Ecology
BSCI 464/MEES 698

- Microorganisms represent two of the three domains of life
- Uncultured microorganisms comprise the majority of the planet's biological diversity
- In many environments, as many as 99% of the microorganisms cannot be cultured by standard techniques

Therefore, culture-independent methods are essential to understand the genetic diversity, population structure, and ecological roles of the majority of microorganisms.

Why using molecular methods?

What are the most basic questions in microbial ecology?

- What microorganisms are present in the community?
- How many of each microorganism are present?
- How do the population changes in space and time?
- What is the ecological role of microbial populations?
- How do microorganisms interact with their environment?
- What do microorganisms do?
What molecular methods to assess microbial diversity?

- Based on 16S ribosomal DNA gene (and possibly other genes)
  - 16S rDNA libraries and sequencing
    - Amplified ribosomal DNA restriction analysis (ARDRA)
    - Terminal-restriction length fragment polymorphism (T-RFLP)
    - Denaturing gradient gel electrophoresis (DGGE)
- Fluorescence in situ hybridization (FISH)
- Pulse field gel electrophoresis analysis (PFGE)
- Metagenomics

16S ribosomal DNA-based community analysis

- <1% of microorganisms in marine samples are culturable using standard methods
- 16S rDNA gene sequences are good for phylogenetic analysis
- Gives indication of microbial diversity

16S rDNA gene-based community analysis

1. DNA isolation from microbial niches
2. PCR amplification with 16S rDNA primers
3. Ligate PCR products into a cloning vector
4. Clone constructs into E. coli
5. Pick colonies, prepare plasmid DNA and sequence insert
6. Analyze sequences
What molecular methods to assess microbial diversity?

- Based on 16S ribosomal DNA gene (and possibly other genes)
  - 16S rDNA libraries and sequencing
  - Amplified ribosomal DNA restriction analysis (ARDRA)
  - Terminal-restriction length fragment polymorphism (T-RFLP)
  - Denaturing gradient gel electrophoresis (DGGE)

- Fluorescence in situ hybridization (FISH)
- Pulse field gel electrophoresis analysis (PFGE)
- Metagenomics

Denaturing Gradient Gel Electrophoresis

Principle of DGGE

Separation based on the electrophoretic mobility of partially melted DNA molecules in polyacrylamide gels (decreased compared to helical form)

Once domain with lowest melting temperature melts, transition from helical to melted form stops migration

 Addition of GC-clamp increases sensitivity

Very useful for visualizing several predominant PCR products

Rapid assessment of many samples
DGGE Analysis

DNA molecules melt in discrete domains

Perpendicular gel

Direction of electrophoresis

Parallel gel

Community Profiling with DGGE
Microorganisms differ in the number of rrn operons in their genomes (1 to 15). The number of rrn operons is positively correlated with growth rate. Slow-growing bacteria would be poorly represented in 16S rRNA libraries generated by PCR. PCR biases are a problem with 16S rDNA and PCR methods.

What molecular methods to assess microbial diversity?

- Based on 16S ribosomal DNA gene (and possibly other genes)
  - 16S rDNA libraries and sequencing
  - Amplified ribosomal DNA restriction analysis (ARDRA)
  - Terminal-restriction length fragment polymorphism (T-RFLP)
  - Denaturing gradient gel electrophoresis (DGGE)
- Fluorescence in situ hybridization (FISH)
- Pulse field gel electrophoresis analysis (PFGE)
- Metagenomics

Fluorescence In Situ Hybridization (FISH)

1. Fix sample: stabilizes macromolecules, prevent cell lysis and permeabilize cell walls.
2. Hybridize with fluorescently-labeled oligonucleotide probe; washing step removes unbound probe molecules.
3. Visualize with epifluorescence microscopy.
Fluorescence In Situ Hybridization (FISH)

Fig. 3. Epifluorescence microphotograph of benthic microbes from Punto Sette. Scale bar 5µm, green Thermococcales detected by OregonGreen-labeled Tcoc164, red archaea detected by Cy3-labeled Arch917 and Arc344, blue DAPI-stained cells. Brightness, shape, and size of the luminescent objects were used to distinguish cells from other particles.

- FISH: rRNA targeted oligonucleotide probes
- RING-FISH: polynucleotide RNA probes
- FISH-MAR: Fluorescence in situ hybridization-microautoradiography

Pulsed Field Gel Electrophoresis

Direct visualization of large DNA molecules or fragments

PFGE of viral communities from the Chesapeake Bay

Wommack et al. 1999. AEM 65:231-240

A-E: Chesapeake Bay stations

Mw is kilobases
What molecular methods to assess microbial diversity?

- Based on 16S ribosomal DNA gene (and possibly other genes)
  - 16S rDNA libraries and sequencing
  - Amplified ribosomal DNA restriction analysis (ARDRA)
  - Terminal-restriction length fragment polymorphism (T-RFLP)
  - Denaturing gradient gel electrophoresis (DGGE)
- Fluorescence in situ hybridization (FISH)
- Pulse field gel electrophoresis analysis (PFGE)
- Metagenomics

Environment Exploration

- Culture-based
  - heavily biased (<0.1% of microorganisms can be cultured)
  - amenable to many types of analyses
- Directed rDNA analyses and sequencing
  - less biased
  - limited analyses possible
- Random shotgun sequencing: metagenomic
  - "differentially" biased
  - amenable to many types of analyses
  - $$$

Metagenomics: functional and sequence-based analysis of the collective microbial genomes contained in an environmental sample: the microbiome

Riesenfeld et al., 2004
Size of Metagenomes

Constructing metagenomic libraries from environmental samples is conceptually simple but technically challenging.

1. A large amount of DNA must be isolated and cloned from a sample (0.5 to 5 µg minimum; high quality DNA; 200 liters Sargasso Sea water)

2. Many clones and sequences must be processed to provide meaningful data (human gut study: 140,000 sequence reads)

3. Lognormal-type population distributions make it difficult to represent the minor species from a sample

Genomics that does not require an initial culturing step or PCR step

Constructing metagenomic libraries from environmental samples is conceptually simple but technically challenging.

Genomics that does not require an initial culturing step or PCR step

Deling, 2005

Size of Inserts

Size of inserts

Deling, 2005
Pros and cons of small-insert and large-insert libraries

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small-insert library (plasmid)</td>
<td>Large-insert library (nuDNA)</td>
</tr>
<tr>
<td>High copy number allows detection of weakly expressed genes</td>
<td>Large size limits detection of weakly expressed genes</td>
</tr>
<tr>
<td>High copy number allows detection of weakly expressed genes</td>
<td>Large size limits detection of weakly expressed genes</td>
</tr>
<tr>
<td>Large numbers of clones must be screened to obtain positives</td>
<td>Large numbers of clones must be screened to obtain positives</td>
</tr>
<tr>
<td>Clones of a known insert or nuDNA can be sequenced for unknowns in position</td>
<td>Clones of a known insert or nuDNA can be sequenced for unknowns in position</td>
</tr>
<tr>
<td>Not stable for enzymes/activities and pathways that are avoided by large-insert vectors</td>
<td></td>
</tr>
<tr>
<td>Technical ease</td>
<td></td>
</tr>
</tbody>
</table>

The challenge: linking function (phenotype) to phylogeny (identity of the host)

Metagenomics strategies

**Phylogeny then function**
- Screen clones for a specific phylogenetic anchor (e.g., 16S rRNA) or gene and then sequence the entire clone and search for genes of interest among the genes flanking the anchor.

**Function then Phylogeny**
- Screen a metagenomic library for a phenotype and then attempt to determine the phylogenetic origin of the cloned DNA.

**Random sequencing**
- Sequence the entire metagenome and identify interesting genes and phylogenetic anchors in the resulting reconstructed genomes.
Phylogeny then Function

<table>
<thead>
<tr>
<th>Environment</th>
<th>Number of contigs</th>
<th>JSI ORFs</th>
<th>Function size (Mbp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acid mine drainage</td>
<td>JGI</td>
<td>12,822</td>
<td>10.8</td>
<td></td>
</tr>
<tr>
<td>Human gut community</td>
<td>TIGR</td>
<td>50,164</td>
<td>78.8</td>
<td></td>
</tr>
<tr>
<td>Sargasso Sea</td>
<td>Venter Inst.</td>
<td>1,214,207</td>
<td>1,044.70</td>
<td></td>
</tr>
<tr>
<td>Sludge</td>
<td>JGI</td>
<td>34,659</td>
<td>28.3</td>
<td></td>
</tr>
<tr>
<td>Soil</td>
<td>JGI</td>
<td>185,274</td>
<td>152.4</td>
<td></td>
</tr>
<tr>
<td>Whalefall sample</td>
<td>JGI</td>
<td>41,932</td>
<td>31.8</td>
<td></td>
</tr>
<tr>
<td>Global Ocean Sampling</td>
<td>Venter Inst.</td>
<td>5 to 6 M</td>
<td>6.3 Gbp</td>
<td></td>
</tr>
</tbody>
</table>

http://img.jgi.doe.gov/cgi-bin/m/main.cgi

Function then Phylogeny

<table>
<thead>
<tr>
<th>Environment</th>
<th>Number of contigs</th>
<th>JSI ORFs</th>
<th>Function size (Mbp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acid mine drainage</td>
<td>JGI</td>
<td>12,822</td>
<td>10.8</td>
<td></td>
</tr>
<tr>
<td>Human gut community</td>
<td>TIGR</td>
<td>50,164</td>
<td>78.8</td>
<td></td>
</tr>
<tr>
<td>Sargasso Sea</td>
<td>Venter Inst.</td>
<td>1,214,207</td>
<td>1,044.70</td>
<td></td>
</tr>
<tr>
<td>Sludge</td>
<td>JGI</td>
<td>34,659</td>
<td>28.3</td>
<td></td>
</tr>
<tr>
<td>Soil</td>
<td>JGI</td>
<td>185,274</td>
<td>152.4</td>
<td></td>
</tr>
<tr>
<td>Whalefall sample</td>
<td>JGI</td>
<td>41,932</td>
<td>31.8</td>
<td></td>
</tr>
<tr>
<td>Global Ocean Sampling</td>
<td>Venter Inst.</td>
<td>5 to 6 M</td>
<td>6.3 Gbp</td>
<td></td>
</tr>
</tbody>
</table>

Random sequencing

<table>
<thead>
<tr>
<th>Microbiome</th>
<th>Seq. center</th>
<th>Est. ORFs</th>
<th>Genome size (Mbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acid mine drainage</td>
<td>JGI</td>
<td>12,822</td>
<td>10.8</td>
</tr>
<tr>
<td>Human gut community</td>
<td>TIGR</td>
<td>50,164</td>
<td>78.8</td>
</tr>
<tr>
<td>Sargasso Sea</td>
<td>Venter Inst.</td>
<td>1,214,207</td>
<td>1,044.70</td>
</tr>
<tr>
<td>Sludge</td>
<td>JGI</td>
<td>34,659</td>
<td>28.3</td>
</tr>
<tr>
<td>Soil</td>
<td>JGI</td>
<td>185,274</td>
<td>152.4</td>
</tr>
<tr>
<td>Whalefall sample 1</td>
<td>JGI</td>
<td>41,932</td>
<td>31.8</td>
</tr>
<tr>
<td>Global Ocean Sampling</td>
<td>Venter Inst.</td>
<td>5 to 6 M</td>
<td>6.3 Gbp</td>
</tr>
</tbody>
</table>

JGI: DOE joint Genome Institute
TIGR: The Institute for Genome Research
Function-driven versus sequence-driven strategies

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function-driven screening method</td>
<td>Maximum yield of cloned genes can be recovered</td>
</tr>
<tr>
<td>Full-length gene products are obtained</td>
<td>Dependence on expression of cloned genes to the bacterial host</td>
</tr>
<tr>
<td>Dependence on the design of the plasmid vector used for insertion of the cloned gene</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sequence-driven screening method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independence of the cloned gene and the bacterial host used</td>
</tr>
<tr>
<td>The cloned gene are inserted into the gene</td>
</tr>
<tr>
<td>No product is produced and no enzymes are used for different targets, for example, mRNA probes can be used</td>
</tr>
<tr>
<td>No selection for functional gene products</td>
</tr>
</tbody>
</table>

Impact On Public Databases

As of April 1, 2004, 5% of GenBank was from the Sargasso Sea scaffold collection.

A BLAST analysis of one sequence read from their collection against GenBank will often identify 50 similar DNA fragments of no known function that are all from the Sargasso Sea.

Finding matches to sequences from the Sargasso Sea is more likely to be due to the abundance of sequences from this study than to ecological similarities.

Metagenomic Analysis of the Human Distal Gut Microbiome

Steven B. Gill,1,2,* Weihe Peng,1 Robert I. Delong,1 Paul S. Eckburg,1,3,4 Peter J. Brownfield,1 Buck T. Samuel,3,4 Jeffrey I. Gordon,3 David A. Relman,1,3,4 Claire M. Fraser-Liggett1,2,9 Karen E. Nelson1

The human intestinal microbiota is composed of 10^12 to 10^14 microorganisms whose collective genome ("microbiome") contains at least 100 times as many genes as our own genome. We studied 78 million base pairs of genome DNA sequence and 16S rRNA polymerase chain reaction (PCR)-amplified 16S ribosomal DNA sequence obtained from the fecal genomes of two healthy adults. Using metabolic function analyses of identified genes, we compared our human genome with the average content of previously sequenced microbial genomes. Our microbiome has significantly enriched metabolism of glycans, amino acids, and nematocidal methanogenic and 2-methyl-4-cyanoethyl-4-aminobutyrate pathways-mediated biosynthesis of vitamins and isoprenoids. Thus, humans are superorganisms whose metabolism represents an amalgamation of microbial and human attributes.

www.sciencemag.org  SCIENCE VOL 312 2 JUNE 2006
Human subjects

- Fecal specimens from 2 healthy humans
  (subject 7 and 8)
- Ages 28 and 37
- Female and male
- Vegetarian diet and unrestricted diet
- No antibiotic used for 1 year

The use of human subjects was approved by the Stanford University Administrative Panel on Human Subjects in Medical Research

(Gill et al., 2005)

Human gut microbiome statistics

<table>
<thead>
<tr>
<th></th>
<th>Subject 7</th>
<th>Subject 8</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequencing reads</td>
<td>65,059</td>
<td>74,462</td>
<td>139,521</td>
</tr>
<tr>
<td>16S rRNA clones sequenced</td>
<td>3,514</td>
<td>3,601</td>
<td>7,115</td>
</tr>
</tbody>
</table>

1 microbial genome equivalent

Total sequences: 78.7 Mbp
Open reading frames: ~50,000

(Gill et al., 2005)

Human gut microbiome phylogenetic

<table>
<thead>
<tr>
<th></th>
<th>Shotgun 16S rRNA (%)</th>
<th>Blast x best hit (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subject 7</td>
<td>Subject 8</td>
</tr>
<tr>
<td>Bifidobacteriales (Actinobacteria)</td>
<td>21.2</td>
<td>3.7</td>
</tr>
<tr>
<td>Clostridiales (Firmicutes)</td>
<td>47.3</td>
<td>74.1</td>
</tr>
<tr>
<td>Bacteroides (Bacteroidetes)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Methanobacteriales (Archaea)</td>
<td>12.3</td>
<td>11.05</td>
</tr>
</tbody>
</table>

- Bifidobacteriales: mostly Bifidobacterium longum, lactic acid bacteria
- Firmicutes and Bacteroidetes found as dominant bacterial divisions in human gut; discrepancy in this study regarding Bacteroidetes - biases associated with fecal lysis - less abundant in those subjects
- Archaea were mostly Methanobacteriales, more than half were Methanobrevibacter smithii

(Gill et al., 2005)
Human gut microbiome new findings

- Large number of Archaea
- Antibiotic resistance genes
  - residual microorganisms
  - antibiotics from nutrition
- Mobile genetic elements influence bacterial diversity
- Enrichment of a specific pathway for the biosynthesis of vitamins and isoprenoids (MEP: 2-methyl-D-erythritol 4-phosphate)

PhyloChip:

- High-density oligonucleotide microarray designed to detect and quantify all known prokaryotic 16S rRNA gene sequences.
- Affymetrix platform
- Based on the 16S rRNA gene
- 500,000 probes
- >100,000 unique database sequences (both archaeal and bacterial), totaling almost 9,000 distinct taxonomic groups
- Each group is assayed by a set of 11 or more perfectly matching probes each with a corresponding mismatch control probe
- Used to characterize complex environments such as soil, aquifers, and urban air samples

Gary Andersen, Terry Hazen and Eoin Brodie - Lawrence Berkeley National Laboratory
GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes

- 24,243 50-mer oligonucleotides
- >10,000 genes
- >150 functional groups
  - N, C, S, P cycling
  - Metal reduction and resistance
  - Organic contaminant degradation

GeoChip
- Provide information on biogeochemical processes and functional activities of microbial communities relevant to human health, agriculture, energy, global climate change, ecosystem management, and environmental cleanup and restoration.
- Provide direct linkages of microbial genes/populations to ecosystem processes and functions.

Where to?
- Comparative metagenomics of microbial communities
- Environmental proteomic studies
- Integration of community genomics and functional assays in situ