Molecular methods to assess microbial diversity

Microbial Ecology
BSCI 464/MEES 698

Overview

- Sampling
- Traditional Molecular methods
- Metagenomics

Why using molecular methods?

- Uncultured microorganisms comprise the majority of the planet’s biological diversity
- Microorganisms represent two of the three domains of life
- 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
Molecular methods to assess microbial diversity

- Most basic questions in microbial ecology:
- What microbes are present in a community?
- How many of each microbe are present?
- How do the populations change in space and time?

Analytical Methods:

- DGGE
- Metagenomics
- SS ribosomal DNA gene community analysis
- Fluorescent in situ hybridization (FISH)
- Amplified Ribosomal DNA Restriction Analysis (ARDRA)
- Terminal-Restriction Fragment Polymorphism (T-RFLP, T-RFLP) patterns
- Pulsed-field gel electrophoresis (PFGE) analysis

Sample Acquisition

- Research vessels, UVCOL 520
- Surface-deployed devices for water and sediment collection
- Submersibles
- Rapid processing, careful storage are essential to ensure high-quality data

DNA Extraction

- Approach 1: Separate cells from "environment" (e.g., water: filtration, centrifugation; soil: sonicating, detergents, centrifugation)

- Approach 2: Lysed cells "in situ" (e.g., soil: enzymatic lysis, bead-beating; sponge: lyophilization, enzymatic lysis, bead-beating)

- Necessary to test and optimize
- Objective is to obtain high-quality DNA from all components of the community.
- Suitable substrate for PCR
PCR Amplification

Selective amplification of specific gene (e.g. 16S rRNA gene) greatly enhances sensitivity.

Many Problems
- Multiple copies of rRNA (e.g. E. coli, Pseudomonas putida, R. solani, C. neoformans)
- Formation of chimeras (Check-Chimeras)
- Differential amplification depending on primer design
- Semi-quantitative at best; subsequent quantification using fluorescent in situ hybridization (FISH)
- Non-PCR-based methods??

16S ribosomal RNA-based Community Analysis

- ca 1% of microbes in marine samples are culturable using standard methods
- 16S ribosomal RNA gene sequence: good for phylogenetic analysis
- Gives indication of microbial diversity

16S rRNA gene-based community analysis
**Fluorescence In Situ Hybridization (FISH)**

![Fig. 3](image_url)

Fluorescence microphotograph of benthic microbes from Punto Sette. Scale bar 2 µm. 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.

**RFLP methods**

**Restriction Fragment Length Polymorphism**

Rapid screening of many clones in 16S rDNA gene library

Very useful to reduce sequencing costs in libraries from "low-diversity" communities

*"Finger-print" of microbial communities*

- ARDRA Amplified Ribosomal DNA Restriction Analysis
- T-RFLP Terminal Restriction Fragment Length Polymorphism
How does it work?

16S rDNA gene amplification (PCR)

P. putida 16S gene

E. coli 16S gene

Restriction digest of 16S rDNA gene

E. coli 16S rRNA gene

P. putida 16S rRNA gene

HapIII

HapIII

Figure 1. Restriction patterns obtained after restriction digestion with CfoI, AluI, MboI, RsaI and MspI for amplified 16S rDNA of different Acinetobacter species.

http://leuven.rug.ac.be/~mvaneech/ARDRA/Acinetobacter.html
Community Profiling by ARDRA

1. Collect samples
2. Selective culture
3. Extract DNA
4. Create PCR-amplified DNA library
5. RFLP screen for unique clones
6. Sequence analysis of unique clones
7. Amplify rDNA
8. Terminal-Restriction Length Fragment Polymorphism (T-RFLP)
9. Community fingerprinting

T-RFLP community fingerprinting

Population Profiling by Terminal RFLP
Fig. 4. Comparison between T-RFLP fragments from the clone library and their occurrence in T-RFLP fingerprints from the complex archaeal community.

Forward (21F-FAM) and reverse (958R-JOE) end-labeled T-RFLP fingerprints of a complex archaeal community from which the clone library was constructed. From the PCR amplified community, 16S rRNA genes were cut with the restriction enzymes HhaI, RsaI and HaeIII. The fragments of the complex community matching in the respective forward or reverse fragment to clones in our library are numbered. The corresponding T-RFLP OTUs affiliations to a certain peak are shown in the shaded areas. Peaks not found in the clone library are indicated by the asterisks.

Moeseneder et al., J. Microbiol. Meth. 2001

---

Principle of DGGE

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

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

Addition of GC-clamp increased sensitivity

Very useful for visualizing several predominant PCR products

Rapid assessment of many samples

Denaturing Gradient Gel Electrophoresis

---

DGGE Analysis

DNA denaturing analysis of discrete domains

Direction of electrophoresis

Perpendicular gel

Parallel gel

100% 0%

40% 60%
Community Profiling with DGGE

- microorganisms differ in the number of rrr operons in their genomes (1 to 15)
- number of rrr operons is positively correlated with growth rate
  
  then slow-growing bacteria would be poorly represented in 16S rRNA libraries generated by PCR

- PCR biases
Pulsed Field Gel Electrophoresis

Direct visualization of large DNA molecules or fragments

PFGE fingerprinting of viral communities

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

PFGE of viral communities from the Chesapeake Bay

Warren et al. 1999. AEM 65:231-240
Lipid profiling

Metagenomics: Potential for Exploring Microbial Diversity

Problem: The vast majority of environmental microorganisms cannot be cultured but they are the most abundant organisms on the planet and are important in planetary scale anthropogenic changes

Metagenomic a key technology

- to better understand the physiology of non-cultured microbes and their significance in biogeochemical cycles
- for the development of novel biotechnological and pharmaceutical products (enzymes, natural products)

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

Riesenfeld et al., 2004
Genomics that does not require an initial culturing step

Streit and Schmitz, 2004

(a) a large amount of DNA must be isolated and cloned from a small sample
(b) many clones and sequences must be processed to provide meaningful data
(c) lognormal-type population distributions make it difficult to represent the minor species from a sample

Size of Metagenomes

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

Size of Inserts

Delong, 2005
Phylogenetic Anchors

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

- screen a metagenomic library for a phenotype and then attempt to determine the phylogenetic origin of the cloned DNA
- 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
- sequence the entire metagenome and identify interesting genes and phylogenetic anchors in the resulting reconstructed genomes

Function then Phylogeny

<table>
<thead>
<tr>
<th>Function</th>
<th>Phylogeny</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathway 1</td>
<td>Clonal group A</td>
</tr>
<tr>
<td>Pathway 2</td>
<td>Clonal group B</td>
</tr>
<tr>
<td>Pathway 3</td>
<td>Clonal group C</td>
</tr>
</tbody>
</table>

Table 1: Metagenomics-driven bacterial functional consortia

<table>
<thead>
<tr>
<th>Environment</th>
<th>Microorganisms</th>
<th>Functional Traits</th>
<th>Phylogenetic Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil</td>
<td>Bacteria</td>
<td>Growth inhibition</td>
<td>Gamma-Proteobacteria</td>
</tr>
<tr>
<td>Water</td>
<td>Archaea</td>
<td>Oxidation</td>
<td>Methanobacteria</td>
</tr>
<tr>
<td>Marine</td>
<td>Eubacteria</td>
<td>Photosynthesis</td>
<td>Alphaproteobacteria</td>
</tr>
</tbody>
</table>

Riesenfeld et al., 2004
Phylogeny then Function

Riesenfeld et al., 2004

<table>
<thead>
<tr>
<th>Environment</th>
<th>Insert size (kb)</th>
<th>Sequence reads</th>
<th>Base pairs of sequence</th>
<th>Goal</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fossil</td>
<td>LASE*</td>
<td>532</td>
<td>~37,000</td>
<td>Random viral clone sequencing</td>
<td>(15)</td>
</tr>
<tr>
<td>Marine</td>
<td>LASE*</td>
<td>3601</td>
<td>~7.40,000</td>
<td>Random viral clone sequencing</td>
<td>(16)</td>
</tr>
<tr>
<td>Drinking water network biofilm</td>
<td>Phantasm and Cosmid</td>
<td>2496</td>
<td>2 $\times$ 10$^6$</td>
<td>Random-clone sequencing</td>
<td>(105)</td>
</tr>
<tr>
<td>Acid mine drainage</td>
<td>3.2 $\times$ 10$^6$</td>
<td>76.2 $\times$ 10$^6$</td>
<td>Reconstruct genome of microbial community</td>
<td>(114)</td>
<td></td>
</tr>
<tr>
<td>Marine</td>
<td>2.6 $\times$ 10$^6$</td>
<td>1.63 $\times$ 10$^7$</td>
<td>&quot;Pilot study&quot; of large-scale whole community sequencing</td>
<td>(115)</td>
<td></td>
</tr>
</tbody>
</table>

*RASE: Unique amplified genome library
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.

It is critical that users of the databases are aware that 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.

Integrating Metagenomics and Community Ecology

Microscopy

direct observation of microorganisms in situ with FISH, RING-FISH, and FISH-MAR

Visualization of uncultured archaea in various environments by fluorescence in situ hybridization (FISH). Schleper et al., 2005.
**Microscopy**

direct observation of microorganisms in situ with FISH, RING-FISH, and FISH-MAR

**SIP-enabled metagenomics** (stable-isotope probing)

SIP is used to identify microorganisms in environmental samples that use a particular growth substrate: (1) incorporation of a substrate that is highly enriched in a stable isotope, such as 13C, and (2) identification of active microorganisms by selective recovery and analysis of isotope-enriched cellular components.

---

**Integrating Metagenomics and Community Ecology**

DNA-based stable isotope probing (SIP) and 13C-phospholipid fatty acids (PLFA) analyses

---

**DNA-based stable isotope probing (SIP)**

- Incorporate a substrate that is highly enriched in a stable isotope (e.g., 13C).
- Identify active microorganisms by selective recovery and analysis of isotope-enriched cellular components.

**13C-phospholipid fatty acids (PLFA)**

- Analyze lipids from microorganisms to assess their metabolic activity and community composition.

---

**FISH**

- Fluorescence in situ hybridization
- Microautoradiography

**Radioisotopes**

- H-acetate, 14C-pyruvate, 14C-butyrate, 14C-bicarbonate

---

**Stable Isotope Probing (SIP)**

- Use a growth substrate highly enriched in a stable isotope (e.g., 13C).
- Identify active microorganisms by selective recovery and analysis of isotope-enriched cellular components.

**DNA-based stable isotope probing (SIP)**

- Analyze lipids from microorganisms to assess their metabolic activity and community composition.

---

**13C-phospholipid fatty acids (PLFA)**

- Analyze lipids from microorganisms to assess their metabolic activity and community composition.

---

**Dumont and Murrel, 2005**
Microscopy
Direct observation of microorganisms in situ with FISH, RING-FISH, and FISH-MAR.

SIP-enable Metagenomics (stable-isotope-probing)
SIP is used to identify microorganisms in environmental samples that use a particular growth substrate: (1) incorporation of a substrate that is highly enriched in a stable isotope, such as $^{13}$C, and (2) identification of active microorganisms by selective recovery and analysis of isotope-enriched cellular components.

Community Genomics
Comparative metagenomics of microbial communities
Integrating community genomics and functional assays in situ

Comparative Metagenomics of Microbial Communities

Fig. 3. Functional profiling of microbial communities. Two-way clustering of samples and encoded functions based on relative enrichment of KEGG functional processes. The 15 most discriminating processes are highlighted.

Tringe et al., 2005

Integrating community genomics and functional assays in situ

Allen and Banfield, 2005
Industrial enzymes — from the metagenome to applications and processes.

'Lfull-cycle' approach to characterizing microorganisms in their natural settings without the need for cultivation.