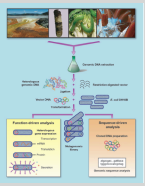
# Molecular methods to assess microbial diversity

**Microbial Ecology**
**BSCI 464/MEES 698**



## Why using molecular methods?

- 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*

## 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?

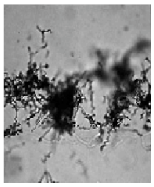
.....

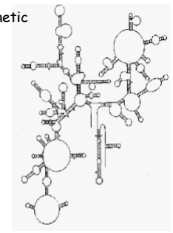
## What molecular methods to assess microbial diversity?

- 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

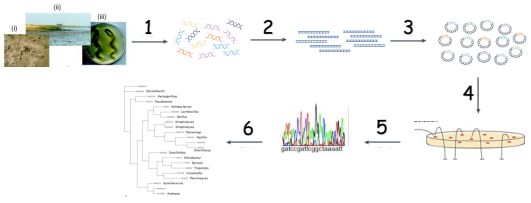
## 16S ribosomal DNA-based community analysis

- <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

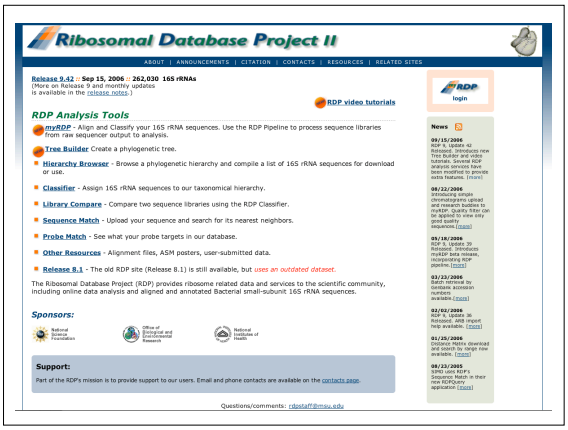




## 16S rDNA gene-based community analysis




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 Molecule in polyacrylamide gels (decreased compared to helical form)

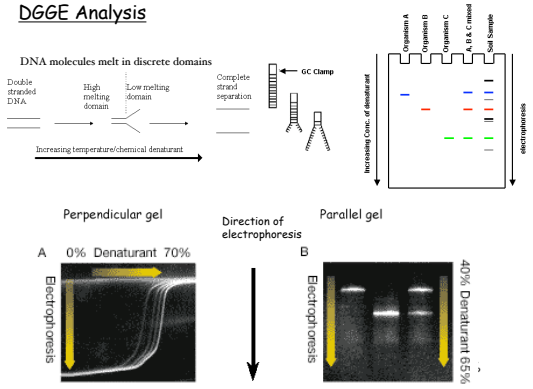
Once domain with lowest melting temperature melts, transition from helical to melted form stops migration

Addition of GC-clamp increases sensitivity

Very useful for visualizing several predominant PCR products

Rapid assessment of many samples

## DGGE Analysis



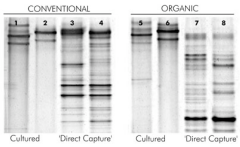

## DGGE

DGGE enables rapid comparison of many samples

Bands can be excised and sequence

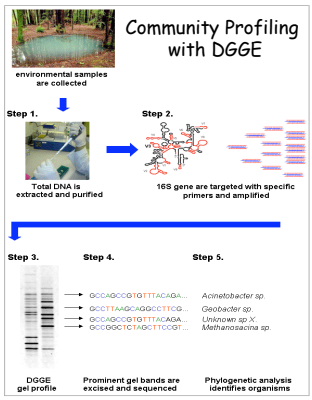
Good for simple communities or checking dominant organisms

PCR-based




*(From Andrea Ottesen)*

## Community Profiling with DGGE



## Problems with 16S rDNA and PCR methods

- microorganisms differ in the number of rrn operons in their genomes (1 to 15)
- number of rrn operons is positively correlated with growth rate

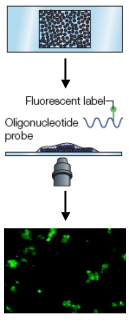
    *then slow-growing bacteria would be poorly represented in 16S rRNA libraries generated by PCR*

- PCR biases

## 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)

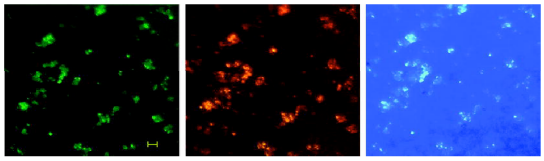


Fig. 3 Epifluorescence microphotograph of benthic microbes from Punto Sette. Scale bar 2μm, Green *Thermococcales* detected by OregonGreen-labeled Tcoc164, red archaea detected by Cy3-labeled Arch917 and Arc344, blue DAPI-stained cells. Brightness, shape, and size of the luminescent objects were used to distinguish cells from other particles.

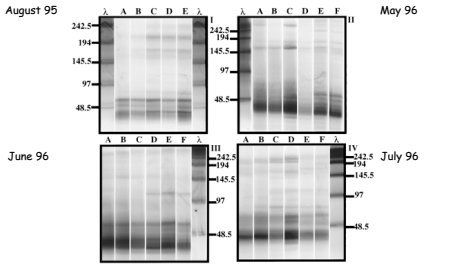
- FISH: rRNA targeted oligonucleotide probes
- RING-FISH: polynucleotide RNA probes
- FISH-MAR: fluorescence in situ hybridization-microautoradiography

Rusch and Amend, Extremophiles 2004

## 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-F: Chesapeake Bay stations
Mwm in kilobases

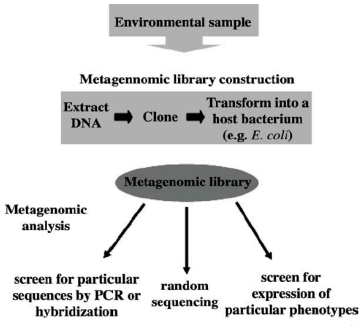
## What molecular methods to assess microbial diversity?

- 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

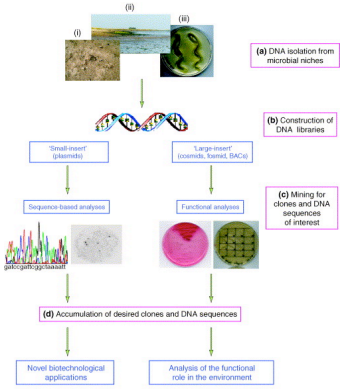
Constructing metagenomic libraries from environmental samples is conceptually simple but technically challenging

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)

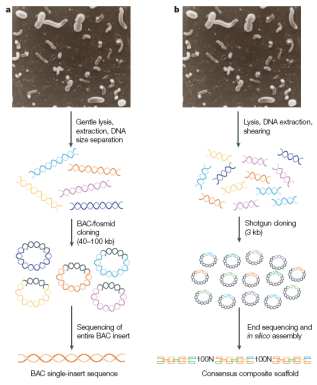
3- lognormal-type population distributions make it difficult to represent the minor species from a sample

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




Streit and Schmitz, 2004

## Size of Inserts




Delong, 2005

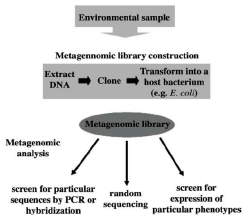
## Pros and cons of small-insert and large-insert libraries

| Advantages | Disadvantages |
|---|---|
| ***Small-insert library (plasmids)*** | |
| High copy number allows detection of weakly-expressed foreign genes | Small insert size |
| Expression of foreign genes from vector promoters is feasible | Large numbers of clones must be screened to obtain positives |
| Cloning of sheared DNA or soil DNA contaminated with matrix substances is possible | Not suitable for cloning of activities and pathways that are encoded by large gene clusters |
| Technically simple | |
| ***Large-insert library (cosmids, fosmids, BACs)*** | |
| Large insert size | Low copy-number might prevent detection of weakly-expressed foreign genes |
| Small numbers of clones can be screened to obtain positives | Limited expression of foreign genes by vector promoters |
| Suitable for cloning of enzyme activities and pathways that are encoded by large gene clusters | Requires high-molecular soil DNA of high purity for library construction |
| Suitable for partial genomic characterization of uncultured soil microorganisms | Technically difficult |

BACs, bacterial artificial chromosomes.

Daniel, 2005

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




Riesenfeld et al., 2004

## 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 2 Metagenomics discovery of homologues of targeted genes or gene families

| Environment | Number of clones | Insert size (kb) | Total DNA (Gb) | Genes of interest | Reference |
|---|---|---|---|---|---|
| Marine | 6240 | 80 | | 16S rRNA<br>Archaea 16S rRNA<br>Photosystem II (*psbA*) | (5)<br>(9)<br>(126) |
| Marine | 7200 | 40 | 2.9 | Archaea 16S rRNA | (6) |
| Marine | 32,000 | 10–20 kb | 0.48 | 16S rRNA | (106) |
| Polychaete symbionts | n.s. | Fosmid | — | 16S rRNA | (18) |
| Marine | 5000 | 50 | 0.25 | Polyketide synthase | (22) |
| Marine | n.s. | BAC | — | Proteorhodopsin | (25) |
| Sediment | n.s. | Fosmid | — | Archaea 16S rRNA methyl coenzyme M reductase A | (43) |
| Tubeworm symbiont | 1500 | Fosmid | ~0.06 | Histidine protein kinase | (55) |
| Soil | 3600<br>24,600 | 27<br>45 | 1.2 | 16S rRNA | (63, 96) |
| Marine | 6107 | 35–40 | | 16S rRNA | (64) |
| Beetle symbionts | n.s. | Fosmid | — | Polyketide synthase | (81) |
| Beetle and sponge symbionts | n.s. | Cosmid | — | Polyketide synthase | (82) |
| Soil | 25,278 | 35–40 | 0.95 | 16S rRNA | (84) |
| Soil | 56,000 | 33–44 | 2.2 | Acidobacteria 16S rRNA | (85) |
| Sponge symbionts | n.s. | Fosmid | — | *RadA* | (100) |
| Sponge symbionts | n.s. | 40 | — | Archaea 16S rRNA | (101) |
| Sponge symbionts | n.s. | 40 | — | Archaea 16S rRNA | (103) |
| River biofilm | n.s. | 40 | — | Hybridization analysis | (107) |
| Marine | 3552 | 40 | 0.14 | Archaea 16S rRNA<br>Planctomycetes 16S rRNA | (110)<br>(116) |

Riesenfeld et al., 2004

## Function then Phylogeny

TABLE 1 Metagenomics discovery based on functional screens

| Environment | Number of clones | Insert size (kb) | Total DNA (Gb) | Activity of interest | Reference |
|---|---|---|---|---|---|
| Soil | n.s. | Cosmid | — | Fatty acid enol esters | (11) |
| Soil | n.s. | Cosmid | — | Pigments | (12) |
| Soil | 700,000 | Cosmid | ~26 | Antimicrobials | (13) |
| Soil | n.s. | Cosmid | — | Fatty acid enol esters | (14) |
| Marine | 825,000 | Plasmid | ~4.0 | Chitinases | (21) |
| Feces and soil | 4 × 6000–35,000 | 30–40 | ~3 | Biotin biosynthesis | (31) |
| Anaerobic digester | 15,000 | 1–12 | 0.10 | Cellulases<br>4-hydroxybutyrate utilization | (48)<br>(49) |
| Soil | 3 × ~300,000 | 5-8 kb | 5.9 | Lipases<br>Antiporter activity | (50)<br>(69) |
| Soil & river sediment | 1 × 80,000<br>2 × 240,000 | 3–5 kb | 2.2 | Dehydratase | (59) |
| Soil & river sediment | 4 × 100,000 | 3–6 kb | 1.8 | Alcohol oxidoreductase | (60) |
| Soil | 3 × 320,000–580,000 | 3–5 kb | 6.2 | Carbonyl formation | (61) |
| Soil | $1.5 \times 10^7$ | 37 kb | 560 | Antimicrobial | (68) |
| Human mouth | 450 | Plasmid | ~0.001 | Antibiotic resistance | (27) |
| Various | n.s. | n.s. | n.s. | Amylase | (91) |
| Soil | 4 × 58,000–650,000 | 3–4 kb | 4.2 | Antibiotic resistance<br>Antimicrobials and novel enzymes | (92)<br>(96) |
| Soil | 3648<br>24,576 | 27<br>44.5 | 1.2 | Antimicrobials | (36) |
| Soil | n.s. | 30 kb | — | Novel biocatalysts | (117) |
| Soil | n.s. | n.s. | — | Antimicrobials | (119) |
| Geothermal sediment | 37,000 | 5 | 0.2 | Pigments | (120) |

Riesenfeld et al., 2004

## Random sequencing

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

JGI: DOE joint Genome Institute
TIGR: The Institute for Genome Research

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

## Function-driven versus sequence-driven strategies

| Advantages | Disadvantages |
|---|---|
| ***Function-driven screening method*** | |
| Completely novel genes can be recovered | Dependent on expression of the cloned genes by the bacterial host |
| Selects for full-length genes | Requires production of a functional gene product by the bacterial host |
| Selects for functional gene products | Dependent on the design of a simple activity-based screening strategy |
| | |
| ***Sequence-driven screening method*** | |
| Independent of expression of the cloned genes by the bacterial host used | Recovered genes are related to known genes |
| Similar screening strategies can be used for different targets, for example, colony hybridization and PCR | Partial genes can be cloned |
| | Not selective for functional gene products |

Daniel, 2005

## 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 R. Gill,[1*‡] Mihai Pop,[1†] Robert T. DeBoy,[1] Paul B. Eckburg,[2,3,4]
Peter J. Turnbaugh,[5] Buck S. Samuel,[5] Jeffrey I. Gordon,[5] David A. Relman,[2,3,4]
Claire M. Fraser-Liggett,[1,6] Karen E. Nelson[1]

The human intestinal microbiota is composed of $10^{13}$ to $10^{14}$ microorganisms whose collective genome ("microbiome") contains at least 100 times as many genes as our own genome. We analyzed ~78 million base pairs of unique DNA sequence and 2062 polymerase chain reaction–amplified 16*S* ribosomal DNA sequences obtained from the fecal DNAs 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 xenobiotics; methanogenesis; and 2-methyl-D-erythritol 4-phosphate pathway–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

| | Subject 7 | Subject 8 | Total |
|---|---|---|---|
| Sequencing reads | 65,059 | 74,462 | 139,521 |
| 16S rRNA clones sequenced | 3,514 | 3,601 | 7,115 |

1 microbial genome equivalent

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

(Gill et al., 2005)

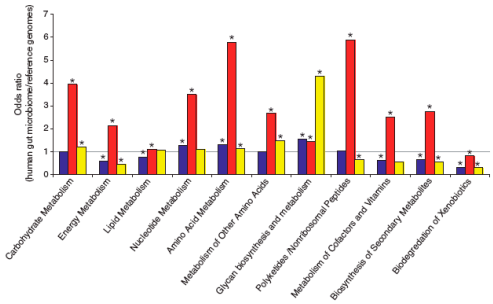
## Human gut microbiome phylogenetic

| | Shotgun 16S rRNA (%) | | Blast x best hit (%) | |
|---|---|---|---|---|
| | Subject 7 | Subject 8 | Subject 7 | Subject 8 |
| **Bifidobacteriales** (Actinobacteria) | 21.2 | 3.7 | 21.8 | 5.2 |
| **Clostridiales** (Firmicutes) | 47.3 | 74.1 | 17.5 | 27.1 |
| **Bacteroides** (Bacteroidetes) | 0 | 0 | 2.7 | 3.4 |
| **Methanobacteriales** (Archaea) | 12.3 | 13.05 | 9 | 7.8 |

- 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)

## Metabolic functions enriched* or underrepresented* in the human gut microbiome




blue: bacterial genomes; red: human genome; archaeal genomes

(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)

(Gill et al., 2005)

## PhyloChip:

High-density oligonucleotide microarray designed to detect and quantify all known prokaryotic 16S rRNA gene sequences.

Affymetrix plateform


PhyloChip boasts a lot of analytical power in a small package.

- based on the 16S rRNA gene
- 500,000 probes
- >30,000 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 sample

*Gary Andersen, Terry Hazen and Eoin Brodie - Lawrence Berkely National Laboratory*

The ISME Journal (2007) 1, 67–77

www.nature.com/ismej

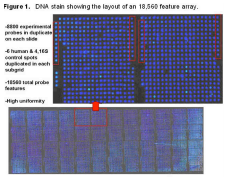
ORIGINAL ARTICLE

# GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes

Zhili He[1,2], Terry J Gentry[2,3], Christopher W Schadt[2], Liyou Wu[1,2], Jost Liebich[2,5], Song C Chong[4], Zhijian Huang[2,6], Weimin Wu[4], Baohua Gu[2], Phil Jardine[2], Craig Criddle[4] and Jizhong Zhou[1,2]

[1]Department of Botany and Microbiology, Institute for Environmental Genomics, University of Oklahoma, Norman, OK, USA; [2]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA; [3]Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA and [4]Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA

**Figure 1.** DNA stain showing the layout of an 18,560 feature array.




24,243 50-mer oligonucleotides

> 10,000 genes

> 150 functional groups
- N, C, S, P cycling
- metal reduction and resistance
- organic contaminant degradation

The ISME Journal (2007) 1, 67–77

www.nature.com/ismej

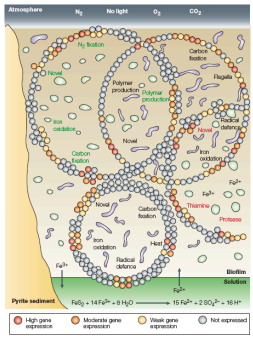
**Table 1** Summary of the numbers of probes by functional gene category on the GeoChip

| *Gene category* | *Unique probes* | *Group probes* | *Total* |
|---|---|---|---|
| Nitrogen cycling | 3896 | 1414 | 5310 |
| Nitrogen fixation | 1225 | 0 | 1225 |
| Denitrification | 1805 | 501 | 2306 |
| Nitrification | 251 | 96 | 347 |
| Nitrogen mineralization | 615 | 817 | 1432 |
| Carbon cycling | 3708 | 891 | 4599 |
| Carbon fixation | 739 | 279 | 1018 |
| Cellulose degradation | 1213 | 72 | 1285 |
| Lignin degradation | 423 | 90 | 513 |
| Chitin degradation | 651 | 93 | 744 |
| Methane production | 194 | 243 | 437 |
| Methane oxidation | 243 | 93 | 336 |
| Others | 245 | 21 | 266 |
| Sulfate reduction | 1286 | 329 | 1615 |
| Phosphorus utilization | 89 | 56 | 145 |
| Metal reduction and resistance | 4039 | 507 | 4546 |
| Arsenic resistance | 775 | 102 | 877 |
| Cadmium resistance | 237 | 45 | 282 |
| Chromium resistance | 303 | 16 | 319 |
| Mercury resistance/reduction | 439 | 109 | 548 |
| Nickel resistance | 135 | 5 | 140 |
| Zinc resistance | 110 | 18 | 128 |
| Other metals and metalloids | 1450 | 182 | 1632 |
| Other metal resistance/reduction | 590 | 30 | 620 |
| Contaminant degradation | 6941 | 1087 | 8028 |
| BTEX and related aromatics | 3852 | 324 | 4176 |
| Chlorinated aromatics | 79 | 11 | 90 |
| Nitroaromatics | 134 | 18 | 152 |
| Heterocyclic aromatics | 176 | 37 | 213 |
| Hydrocarbons (e.g. PAHs) | 597 | 144 | 741 |
| Polychlorinated biphenyls (PCBs) | 303 | 85 | 388 |
| Chlorinated solvents | 178 | 54 | 232 |
| Pesticides | 504 | 54 | 558 |
| Other chemicals & by-products | 1297 | 160 | 1457 |
| Others | 21 | 0 | 21 |
| Total | 19 959 | 4284 | 24 243 |

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

Allen and Banfield, 2005