<table>
<thead>
<tr>
<th>Date</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 14</td>
<td>Jim Brown</td>
</tr>
<tr>
<td>Jan 28</td>
<td>Youwen Pan</td>
</tr>
<tr>
<td>Feb 11</td>
<td>Dilan Weerakoon</td>
</tr>
<tr>
<td>Feb 25</td>
<td>Rebecca Weingarten</td>
</tr>
<tr>
<td>Mar 10</td>
<td>Jae Lee</td>
</tr>
<tr>
<td>Mar 24</td>
<td>Drew Devine</td>
</tr>
<tr>
<td>Apr 7</td>
<td>Justin Bradshaw</td>
</tr>
<tr>
<td>Apr 21</td>
<td>open date</td>
</tr>
</tbody>
</table>

www.mbio.ncsu.edu/MJC
If you are a speaker...

- Give me the info on the paper you’ve chosen AT LEAST one week in advance

- Prepare your presentation. If you plan to use my computer, use Powerpoint or Keynote format and be sure to include any accessory files (movies, &c)

- Show up at 8:45 the morning of your presentation with either your laptop or your presentation on a USB drive.

- Afterwards, give me a copy of your presentation
Microbial Population Structures in the Deep Marine Biosphere

John D. Moyer,1 David G. Mark Welch,2 Hillary G. Morrison,3 Jason M. Huse,4 Philipp A. Nolte,5 Gordon P. Bemfica,6 Michael J. O’Neill7

The microbial potential of environmental DNA sequences for modeling microbial ecosystems depends on accurate assessments of population structure, including diversity (richness) and relative abundance estimates. Here we report on the extraction of 500,000 microbial small-subunit ribosomal RNA amplicons from two deep-sea hydrothermal vents on the Juan de Fuca Ridge. We used a tag sequencing strategy that combines the use of amplicons of the V6 hyper-variable region of the 16S rRNA gene with parallel sequencing of high-throughput DNA libraries. This approach allowed us to identify 6.1 million predicted proteins (1.8 gigabases) from two vent communities, and that different patterns of evenness for both high- and low-abundance taxa may be important in defining microbial ecosystem dynamics. Furthermore, we discovered that hundreds of thousands of sequences will be necessary to capture the vast diversity of microbial communities at two neighboring hydrothermal vents by examining the sequences of more than 900,000 microbial small-subunit ribosomal RNA amplicons. The two vent communities have different microbial population structures, and that different patterns of evenness for both high- and low-abundance taxa may be important in defining microbial ecosystem dynamics.

References and Notes

30. We thank T. F. Hansen, Ø. H. Holen, A. J. van Noordwijk, M. Th. Hanson, and N. E. Hansen for suggestions. The study was supported by grants from the Natural Environment Research Council (UK) and the Research Council of Norway (contract number 170526/220).
1. Proof of principle for the method

2. How big is microbial diversity?

3. What is the community composition of these environments?

4. How similar are the microbial communities in these two nearby similar environments?
Approach

1. Sample
2. Isolate DNA
3. rRNA PCR
4. Sequence
5. Divide into groups
6. Phylogenetic analysis
Approach

Traditional

Sample

Isolate DNA

rRNA PCR

Sequence

Divide into groups

Phylogenetic analysis

This paper
Approach

Traditional
- Isolate DNA
- rRNA PCR
- Sequence
- Divide into groups
- Phylogenetic analysis

This paper
- Sample
- Isolate DNA
- rRNA PCR
- Sequence
- Divide into groups
- Phylogenetic analysis

Entire ssu-rRNA 1000-1500bp
Variable region 6 50-75bp
Approach

Traditional

Entire ssu-rRNA
1000-1500bp

Sequencing
hundreds of sequences

This paper

Sample

Isolate DNA

rRNA PCR

Variable region 6
50-75bp

Sequence

454 Pyrosequencing
hundreds of thousands of sequences

Divide into groups

Phylogenetic analysis
Approach

Sample

Isolate DNA

rRNA PCR

Sequence

Divide into groups

Phylogenetic analysis

Traditional

This paper

Entire ssu-rRNA

1000-1500bp

Variable region 6

50-75bp

454 Pyrosequencing

hundreds of thousands of sequences

Sequencing

hundreds of sequences

OTUs

Diversity metrics
**Approach**

Traditional

Sample

Isolate DNA

rRNA PCR

Sequence

Divide into groups

OTUs

Trees

Diversity metrics

Phylogenetic analysis

This paper

Entire ssu-rRNA

1000-1500bp

Variable region 6

50-75bp

454 Pyrosequencing

hundreds of sequences

hundreds of thousands of sequences

**Sequencing**

**Phylogenetetic analysis**

**Taxonomic bins**
Competing methods

- Cultivation
- Traditional environmental rRNA sequencing
- rRNA DGGE
- rRNA tRFLP
- Metagenomics
Competing methods

- Cultivation
- Traditional environmental rRNA sequencing
- rRNA DGGE
- rRNA tRFLP
- Metagenomics
Competing methods

• Cultivation
• Traditional environmental rRNA sequencing
  • rRNA DGGE
• rRNA tRFLP
• Metagenomics
Competing methods

- Cultivation
- Traditional environmental rRNA sequencing
- rRNA DGGE
- rRNA tRFLP
- Metagenomics
Competing methods

- Cultivation
- Traditional environmental rRNA sequencing
- rRNA DGGE
- rRNA tRFLP
- Metagenomics
Supplementary Figure 1. Submersible pictures of sampling sites (a) Marker 52, FS396 and (b) Bag City, FS312, with intake of fluid sampling probe held in the ROV arm shown. Images courtesy of NOAA/PMEL Vents Program.
Supplementary Figure 1. Submersible pictures of sampling sites (a) Marker 52 and (b) Bag City, FS312, with intake of fluid sampling probe held in the ROV arm shown. Images courtesy of NOAA/PMEL Vents Program.
**Figure 8**

**DNA Library Preparation**
- 4.5 HOURS
  - Anneal ssDNA to an excess of DNA Capture Beads

**emPCR**
- 8 HOURS
  - Emulsify beads and PCR reagents in water-in-oil microreactors
  - Clonal amplification occurs inside microreactors

**Sequencing**
- 7.5 HOURS
  - Break microreactors enrich for DNA-positive beads

**gDNA → ssDNA Library**
**Figure 8**

DNA Library Preparation → emPCR → Sequencing

- **4.5 HOURS**
  - Anneal ssDNA to an excess of DNA Capture Beads
- **8 HOURS**
  - Emulsify beads and PCR reagents in water-in-oil microreactors
- **7.5 HOURS**
  - Clonal amplification occurs inside microreactors
  - Break microreactors enrich for DNA-positive beads

```
gDNA → ssDNA Library
```

**Figure 9**

DNA Library Preparation → emPCR → Sequencing

- **4.5 HOURS**
  - Amplified ssDNA library beads
- **8 HOURS**
  - Well diameter: average of 44μm
  - 400,000 reads obtained in parallel
  - A single cloned amplified ssDNA bead is deposited per well
- **7.5 HOURS**
  - Quality filtered bases

```
Amplified ssDNA library beads → Quality filtered bases
```
Secondary Structure: small subunit ribosomal RNA

Citation and related information available at http://www.rna.icmb.utexas.edu

Escherichia coli
(J01695)
1. Bacteria 2. Proteobacteria 3. gamma subdivision
4. Enterobacteriaceae and related symbionts
5. Enterobacteriaceae 6. Escherichia
Nov 1999

Symbols Used In This Diagram:
- c - Canonical base pair (A-U, G-C)
- u - G-U base pair
- A - G-A base pair
- u - Non-canonical base pair

Every 10th nucleotide is marked with a tick mark.
Escherichia coli
(J01695)
1. Bacteria 2. Proteobacteria 3. gamma subdivision
4. Enterobacteriaceae and related symbionts
5. Enterobacteriaceae 6. Escherichia
Nov 1999

Symbols Used In This Diagram:
- c - Canonical base pair (A-U, G-C)
- G - G-U base pair
- A - G-A base pair
- U - Non-canonical base pair

Every 10th nucleotide is marked with a tick mark.
Escherichia coli (J01695)
Nov 1999

Symbols Used In This Diagram:
- - Canonical base pair (A-U, G-C)
G-U - G-U base pair
G-A - G-A base pair
U-U - Non-canonical base pair

Every 10th nucleotide is marked with a tick mark.
sampled on three occasions. Bag City, one of the sites closest to the most active vents, had the highest abundance. These were the largest colonies sampled on any occasion; Bag City, with its highest abundance, and most likely the largest colonies sampled on any occasion. Marker 52 has a lower abundance, but the second-highest diversity of which was neither Bag nor Marker 52, and Marker 52 was more diverse than any other site. Archaeal diversity was lower at Marker 52 than at Bag City. The diversity of archaeal sequences at Marker 52 was lower than at Bag City. The diversity of archaeal sequences at Marker 52 was lower than at Bag City. The diversity of archaeal sequences at Marker 52 was lower than at Bag City. The diversity of archaeal sequences at Marker 52 was lower than at Bag City. The diversity of archaeal sequences at Marker 52 was lower than at Bag City.

Table 1. Chemical and SSU rRNA tag characteristics of the two sites.

<table>
<thead>
<tr>
<th></th>
<th>FS312</th>
<th>FS396</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vent name</td>
<td>Bag City</td>
<td>Marker 52</td>
</tr>
<tr>
<td>Sample year</td>
<td>2003</td>
<td>2004</td>
</tr>
<tr>
<td>Volume filtered (ml)</td>
<td>1003</td>
<td>2000</td>
</tr>
<tr>
<td>Cells ml⁻¹ (range)</td>
<td>(1.21 × 10⁵)</td>
<td>(1.57 × 10⁵)</td>
</tr>
<tr>
<td>DNA recovered (µg)</td>
<td>0.9</td>
<td>2.4</td>
</tr>
<tr>
<td>Total number of archaeal V6 tag sequences†</td>
<td>200,199</td>
<td>16,428</td>
</tr>
<tr>
<td>Total number of bacterial V6 tag sequences†</td>
<td>442,058</td>
<td>247,662</td>
</tr>
<tr>
<td>Total number of ε-proteobacterial V6 tag sequences†</td>
<td>122,823</td>
<td>147,515</td>
</tr>
<tr>
<td>Depth (m)</td>
<td>1,537</td>
<td>1,529</td>
</tr>
<tr>
<td>Latitude and longitude</td>
<td>45.92°N, 129.99°W</td>
<td>45.94°N, 129.99°W</td>
</tr>
<tr>
<td>Average temperature (°C)</td>
<td>31.2</td>
<td>24.4</td>
</tr>
<tr>
<td>Maximum temperature (°C)</td>
<td>31.4</td>
<td>24.9</td>
</tr>
<tr>
<td>H₂S/ΔT (µmol kg⁻¹ °C⁻¹)</td>
<td>7.2</td>
<td>18.9</td>
</tr>
<tr>
<td>pH</td>
<td>6.26</td>
<td>5.08</td>
</tr>
<tr>
<td>Mg (mmol/kg)</td>
<td>48.3</td>
<td>50.8</td>
</tr>
<tr>
<td>Alkalinity (meq/liter)</td>
<td>2.4</td>
<td>3.7</td>
</tr>
<tr>
<td>Mn (µmol/kg)</td>
<td>19.8</td>
<td>4.8</td>
</tr>
<tr>
<td>Fe (µmol/kg)</td>
<td>0.8</td>
<td>7.9</td>
</tr>
<tr>
<td>Silica (mmol/kg)</td>
<td>1.46</td>
<td>1.07</td>
</tr>
</tbody>
</table>

* Cultured at 70° or 90°C in 0.3% yeast extract and peptone with elemental sulfur; Ar headspace. † Trimmed reads that passed quality control [as described in (11)].
Table 2. Sequencing information and diversity estimates for all bacteria and archaea.

<table>
<thead>
<tr>
<th></th>
<th>Bacteria</th>
<th>Archaea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of V6 tag sequences*</td>
<td>689,720</td>
<td>216,627</td>
</tr>
<tr>
<td>Total unique V6 tag sequences</td>
<td>30,108</td>
<td>5,979</td>
</tr>
<tr>
<td>Total OTUs at 3% difference (phylotypes)</td>
<td>18,537</td>
<td>1,931</td>
</tr>
<tr>
<td>Chao1 estimator of richness at 3% difference (95% CI)</td>
<td>36,869 (36,108 to 37,663)</td>
<td>2,754 (2,594 to 2,952)</td>
</tr>
<tr>
<td>ACE estimator of richness at 3% difference (95% CI)</td>
<td>37,038 (36,613 to 37,473)</td>
<td>2,678 (2,616 to 2,745)</td>
</tr>
<tr>
<td>Bray-Curtis similarity index at 3% difference†</td>
<td>0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>Jaccard similarity index at 3% difference†</td>
<td>0.12</td>
<td>0.08</td>
</tr>
</tbody>
</table>

*Trimmed reads that passed quality control [as described in (11)]. †Similarity between communities at sites FS312 and FS396 on a scale of 0 to 1 (where 1 represents identical communities).
Comparing each unique sequence to our V6 reference database revealed well-characterized taxa, as well as many unknown microbial phylogenotypes. The 10 most abundant sequences occurred more than 10,000 times and were exact matches to sequences in our database, indicating that our sampling was representative. Of the 36,725 unique sequences found at the two sites, 36,180 were represented by fewer than 100 tags; of these, 13,385 were >10% different and ~4000 were >20% different from known SSU rRNA genes. Many rare, divergent taxa account for most of the observed novel microbial diversity.

![Image](image_url)

**Fig. 1.** Taxonomic breakdown of bacterial V6 tags from each vent. Pie charts show the Phylum_Class_Order distribution for taxonomically assigned tags that occurred more than 1000 times; the remaining tag sequences are grouped into "Other." The taxonomic distribution of ε-proteobacterial genera is shown in normalized histograms for each site, with further breakdown of the dominant ε-proteobacteria in additional histograms, with each color in the histograms representing a unique tag sequence. For FS312, the Arcobacter are expanded to the left with a histogram showing those tag sequences that occurred ≥10 times, followed by a histogram showing the diversity of tags that occurred 10 to 1800 times. For FS396, the Sulfurovum are expanded to the right, with a histogram showing those tag sequences that occurred ≥10 times, followed by a histogram showing the diversity of tags that occurred 10 to 8400 times. Nonparametric estimates suggested more than 900 phylotypes each of Arcobacter at FS312 and Sulfurovum at FS396.
The Epsilon proteobacteria consist of only a few genera, mainly the curved to spirilloid *Wolinella*, *Helicobacter*, and *Campylobacter*. Most of the known species inhabit the digestive tract of animals and humans and serve as symbionts (*Wolinella* in cows) or pathogens (*Helicobacter* in the stomach, *Campylobacter* in the duodenum). There have also been numerous environmental sequences of epsilons recovered from hydrothermal vent and cold seep habitats. (Wikipedia)
Fig. 2. Rarefaction curves for total bacterial and archaeal communities at the two sampling sites FS312 and FS396 at 3% and 6% difference levels.

In addition, the importance of microbial diversity reconstructions from the two vents here would likely be largely chimeric of sequences from closely related which may mask important biological. Methods such as the massively parallel sequencing approach used here, combine a multitude of other quantitative and tools now available to microbial ecologists serve as necessary accompaniments to genomic gene surveys as we strive to the impact of diversity on ecosystem and long-term stability (24).

References and Notes
5. H. R. Johnson, R. W. Foubert, J. Cooper, R.
Supplementary Figure 2. Rarefaction curves at 3% difference for bacterial phylotypes within (a) FS312 alpha-proteobacteria, (b) FS396 alpha-proteobacteria, (c) FS312 beta-proteobacteria, (d) FS396 beta-proteobacteria, (e) FS312 gamma-proteobacteria, (f) FS396 gamma-proteobacteria, (g) FS312 delta-proteobacteria, (h) FS396 delta-proteobacteria, (i) FS312 epsilon-proteobacteria, (j) FS396 epsilon-proteobacteria, (k) FS312 Clostridia, (l) FS396 Clostridia, (m) FS312 Flavobacteria, (n) FS396 Flavobacteria, (o) FS312 Arcobacter spp., and (p) FS396 Sulfurovum spp.
Similarity of 454 sequence tags from FS396 to the V6RefDB database. “All tag distribution” plots the number of tag sequences for all samples versus the percentage difference from the best-matching sequence in V6RefDB. “Percent unique reads” from all samples shows the percentage difference between each distinct tag sequence and its best match in V6RefDB. “Percent total tags” plots the cumulative percentage of reads in all samples at or below a given percentage difference from best matches in V6RefDB.
1. Proof of principle for the method ✓
2. How big is microbial diversity? Huge!
3. What is the community composition of these environments? Lots of proteobacteria, especially epsilons
4. How similar are the microbial communities in these two nearby similar environments? Very different at all phylogenetic levels
Not only are there billions and billions of Bacteria, there are billions and billions of kinds of Bacteria.