Metagenomics and Its Application in Microbial Consortia

Spring, 2007

Guangyi Wang, Ph.D.
POST103B
guangyi@hawaii.edu

http://www.soest.hawaii.edu/marinefungi/OCN403webpage.htm
Overview of Last Lecture

- Gene silence using antisense mRNA
  - Single gene antisense
  - Genome-wide antisense
- Proteomics
  - Introduction
  - Challenges and application
  - 2-D gel
  - Mass spectrometry (MS)
    - Major components
  - MS-based proteomics
    - Major steps
    - Limitations
    - Advantage and drawback for 2-D gel and MS-based proteomics
  - Application in marine environment (continued today).
Application in Marine Environment

Marine proteomics: generation of sequence tags for dissolved proteins in seawater using tandem mass spectrometry

Matthew J. Powell\textsuperscript{b}, Jennifer N. Sutton\textsuperscript{b}, Carlos E. Del Castillo\textsuperscript{a}, Aaron T. Timperman\textsuperscript{b,}\textsuperscript{*}

\textsuperscript{a}National Aeronautics and Space Administration, CODE MA10, Stennis Space Center, MS 39529, USA
\textsuperscript{b}Department of Chemistry, West Virginia University, Morgantown WV, 26506, USA

Received 7 April 2004; received in revised form 29 September 2004; accepted 29 September 2004
Available online 10 December 2004

Abstract

Dissolved proteins in seawater samples from the Gulf of Mexico were concentrated using tangential flow ultrafiltration and methanol/chloroform/water precipitation. Following concentration and purification, two different separation methods were employed. In one method, intact proteins were separated by SDS-PAGE and digested enzymatically in-gel. In the second method, the peptides resulting from a solution proteolytic digest of the whole protein pellet mixture were separated by capillary HPLC. In both methods, the final chromatographic separation was coupled on-line with a mass spectrometer using an electrospray interface, and peptide CID spectra were collected using tandem mass spectrometry (MS). De novo sequencing of the peptide tandem mass spectra generated short amino acid sequences (peptide tags) that were used to search databases for protein class and source information. Trends of conserved sequences for two specific classes of proteins were observed: membrane/envelope proteins and enzymes. Similarity searching of peptide tags produced identification of conserved sequences from several protein homologues originating from many different species, including: long chain fatty acyl CoA synthetase, anthranilate synthase, ribulose bisphosphate carboxylase, and luminal binding protein. These results provide new insight into the sources and production mechanisms for dissolved organic matter (DOM), as there is direct evidence for dissolved proteins other than the bacterial outer membrane proteins reported by Tanoue et al. Furthermore, the data presented herein support the idea that physical protection and selective preservation are not mutually exclusive survival mechanisms, but rather these two models are dependent upon one another for explaining the survival of refractory dissolved proteins in seawater.

Marine Chemistry 95 (2005) 183–198
Analysis scheme for the concentration, recovery, separation, identification, and characterization of refractory dissolved proteins in seawater.
Schematic of the lab-built ultrafiltration system. The dashed arrows indicate the direction of water movement. The major steps in the ultrafiltration process are labeled A to E. The UF system components are labeled and described in the inserted table.
Fig. 3. A silver-stained SDS–PAGE gel from the separation of two fractions of dissolved proteins from a 400-m depth Gulf of Mexico seawater sample. The first two lanes contain molecular weight markers (~500 and 250 ng/band protein). The dilution series represent the retentate fractions precipitated by the methanol/chloroform/water (left) and the trichloroacetic acid (right) techniques. Each fraction contains 1/4 of a 60-L seawater sample recovered by using tangential flow ultrafiltration with two 10-kDa membranes. The protein bands labeled A, B, and C were excised, enzymatically digested with trypsin, separated by reverse phase HPLC, and analyzed by tandem mass spectrometry.

Fig. 4. A sequenced CID spectrum from a tryptic peptide (parent m/z 767.37; z=+2) acquired in the MS/MS analysis of Band C from the SDS–PAGE gel in Fig. 3. The b-ion series is labeled and shown on the lower peptide ladder and the y-ion series is similarly shown in the upper peptide ladder. Additional fragment ions are shown in the table inset in the upper right corner. The b-ions carry the charge on the amino terminus; the y-ions carry the charge on the carboxy terminus.
### Table 3
Similarity search results for the 2D-HPLC-MS in-solution tryptic digest analyses

<table>
<thead>
<tr>
<th>Sample</th>
<th>Total no. of peptides searched</th>
<th>Percentage of tag results yielding a 50% correlation</th>
<th>Peptidic CID not sequenced by Denovo X</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Inconclusive/mixed results</td>
<td>Membrane/envelope protein</td>
</tr>
<tr>
<td>Solution digest+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2D-HPLC-MS/MS analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loading effluent</td>
<td>23</td>
<td>34.8</td>
<td>41.3</td>
</tr>
<tr>
<td>20 mM AmAc</td>
<td>13</td>
<td>61.5</td>
<td>15.4</td>
</tr>
<tr>
<td>40 mM AmAc</td>
<td>14</td>
<td>50.0</td>
<td>25.0</td>
</tr>
<tr>
<td>50 mM AmAc</td>
<td>5</td>
<td>60.0</td>
<td>20.0</td>
</tr>
<tr>
<td>60 mM AmAc</td>
<td>3</td>
<td>33.3</td>
<td>0</td>
</tr>
<tr>
<td>80 mM AmAc</td>
<td>4</td>
<td>50.0</td>
<td>50.0</td>
</tr>
<tr>
<td>100 mM AmAc</td>
<td>11</td>
<td>18.2</td>
<td>27.3</td>
</tr>
<tr>
<td>150 mM AmAc</td>
<td>4</td>
<td>25</td>
<td>37.5</td>
</tr>
<tr>
<td>200 mM AmAc</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>500 mM AmAc</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*The 2D-HPLC-MS/MS samples listed are the strong cation exchange fractions used to load the C18 column prior to HPLC-MS/MS analysis.*

*The positive similarity search correlation required 50% of the returned proteins for a peptide tag for that protein-class category.*

### Table 4
Summary of protein similarity search results

<table>
<thead>
<tr>
<th>Protein</th>
<th>Protein class</th>
<th>Peptide tag</th>
<th>No. of species matched in NR database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long chain fatty acyl synthetase</td>
<td>Membrane-associated enzyme</td>
<td>WTAGHV</td>
<td>20+</td>
</tr>
<tr>
<td>Anthranilate synthase</td>
<td>Enzyme</td>
<td>GDLFQVV</td>
<td>16</td>
</tr>
<tr>
<td>Ribulose biphosphosphate carboxylase</td>
<td>Enzyme</td>
<td>LPLAYLK</td>
<td>30+</td>
</tr>
<tr>
<td>Luminal binding protein</td>
<td>Membrane-associated protein</td>
<td>EELNNDLF</td>
<td>20+</td>
</tr>
</tbody>
</table>

*The number different species of organisms that have homologous proteins matching the conserved peptide tag sequence.*
Metagenomics

1. Metagenome (environmental genome) - the genomes of the total microbiota found in nature.

2. Metagenomics
1) META-analysis - the process of statistically combining separate analyses.
2) Genomics - the comprehensive analysis of an organism’s genetic material.
3) Definition - the study of the collective genomes of microorganisms (as opposed to clonal cultures). The technique is to clone DNA in large fragments directly from the microorganism's environment (such as seawater) into a culturable host and conduct a sequence-based and functional genomic analysis on it.
3) Goal - isolate new chemical signals, new secondary metabolites that might have utility to humans, and the reconstruction of an entire genome of an uncultured organism.

3. Cloning Vectors Used in Environmental Genomic Studies

<table>
<thead>
<tr>
<th>Vector</th>
<th>Insert Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cosmids</td>
<td>~ 40 Kb</td>
</tr>
<tr>
<td>BAC</td>
<td>~ 200 Kb</td>
</tr>
<tr>
<td>YAC</td>
<td>&gt; 200 Kb</td>
</tr>
</tbody>
</table>
Cloning Vectors: Cosmids

Nonessential region

Cohesive ends

Digestion with restriction enzymes

Ligation with foreign DNA

Foreign

Hybrid DNA

Packaging with phage

Infective phage particle
Metagenomics vs Community Genomics

1. Dark green region at the bottom - genes primarily for the replication and segregation of the F plasmid.
2. The light green region - the genes involved in conjugative transfer.
3. The oriT sequence - the origin of transfer during conjugation.
4. The arrow indicates the direction of transfer.
5. The regions shown in yellow on F are transposable elements where integration into identical elements on the bacterial chromosome can occur and lead to the formation of different Hfr strains.
1. BACs are derivatives of the F plasmid of Escherichia coli (6.7 kb).
2. At the top of the map is a cloning region that contains several unique restriction enzyme sites.
3. The cat gene confers resistance to the antibiotic chloramphenicol.
4. The other genes shown are involved in plasmid replication.
5. BACs contain only a small fraction of the entire F plasmid.
Construction and analysis of metagenomic libraries.

Function-driven analysis
1) Identification of clones expressing a desired trait;
2) Sequence & biochemical analysis of the clones;
3) Quick identification of useful activities;
4) All genes required for the function in one clone & expression in host cell.
5) Good assay methods & tedious.

AEM 66:2541-2547
68:4301-4306
124:9968-9969
Sequence-driven analysis: use of conserved DNA sequences to design hybridization probes or PCR primers to screen metagenomic libraries for clones that contain sequences of interest.

Environ. Microbiol 2:516-529
Science 289:1902-1906
Nature 411:786-789

Sustained frustration: the low frequency of clones of a desired nature.
Enrichment for specialized DNA from environmental samples

(a) BrdU-enriched metagenomic library construction
- Environmental sample
- Substrate (PCB)
- Enrichment for bacteria that metabolize substrate (PCB)
- DNA extraction
- Immunocapture of BrdU-labeled DNA
- Library construction using BrdU-labeled DNA

(b) Stable isotope probe enriched metagenomic library construction
- Environmental sample
- Substrate (13C-PCB)
- Enrichment for bacteria that metabolize 13C-labeled substrate (13C-PCB)
- DNA extraction
- Ultracentrifugation to separate 13C-labeled DNA from other DNA
- Library construction using 13C-labeled DNA

(c) Metagenomic clone library enrichment
- Metagenomic library
- Enrichment for clones that metabolize substrate (PCB) independently or cooperatively
- Plate enriched library on solid medium (LB-Amp) and characterize unique clones

bromodeoxyuridine
<table>
<thead>
<tr>
<th>Year</th>
<th>Site</th>
<th>Depth (meters)</th>
<th>Community type</th>
<th>Library type</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991</td>
<td>Subtropical Pacific</td>
<td>0</td>
<td>Open ocean picoplankton</td>
<td>λ</td>
<td>32</td>
</tr>
<tr>
<td>1996</td>
<td>Oregon coast</td>
<td>200</td>
<td>Coastal picoplankton</td>
<td>Fosmid</td>
<td>33, 41</td>
</tr>
<tr>
<td>1996</td>
<td>Santa Barbara, California</td>
<td>10</td>
<td>Sponge symbionts</td>
<td>Fosmid</td>
<td>40, 49</td>
</tr>
<tr>
<td>1999</td>
<td>Delaware Bay, Delaware</td>
<td>0</td>
<td>Estuarine picoplankton</td>
<td>λ</td>
<td>53</td>
</tr>
<tr>
<td>2000</td>
<td>Monterey Bay, California</td>
<td>0</td>
<td>Coastal picoplankton</td>
<td>BAC</td>
<td>34, 35, 65</td>
</tr>
<tr>
<td>2002</td>
<td>Antarctic peninsula</td>
<td>0</td>
<td>Coastal picoplankton</td>
<td>Fosmid</td>
<td>101</td>
</tr>
<tr>
<td>2002</td>
<td>Mission Bay, California</td>
<td>0</td>
<td>Coastal planktonic virus</td>
<td>Shotgun*</td>
<td>36, 37</td>
</tr>
<tr>
<td>2003</td>
<td>Subtropical Pacific</td>
<td>0</td>
<td>Open ocean picoplankton</td>
<td>BAC</td>
<td>38</td>
</tr>
<tr>
<td>2004</td>
<td>Antarctic Polar front</td>
<td>500</td>
<td>Pico-plankton</td>
<td>Cosmid</td>
<td>43, 45</td>
</tr>
<tr>
<td>2004</td>
<td>Sargasso Sea</td>
<td>0</td>
<td>Open ocean picoplankton</td>
<td>Shotgun</td>
<td>77</td>
</tr>
<tr>
<td>2004</td>
<td>Mediterranean Sea</td>
<td>0</td>
<td>Open ocean picoplankton</td>
<td>Shotgun</td>
<td>39, 72</td>
</tr>
<tr>
<td>2004</td>
<td>Eel River Basin, California</td>
<td>550</td>
<td>Deep-sea sediment</td>
<td>Shotgun and</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>microorganisms</td>
<td>fosmid</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>Pacific Ocean</td>
<td>1,674</td>
<td>Deep-sea whale fall</td>
<td>Shotgun</td>
<td>55</td>
</tr>
<tr>
<td>2005</td>
<td>Antarctic peninsula</td>
<td>560</td>
<td>Deep-sea whale fall</td>
<td>Shotgun</td>
<td>55</td>
</tr>
</tbody>
</table>

*Modified linker ligation shotgun libraries. BAC, bacterial artificial chromosome.
Application

1. Read the following paper (required!!!)

Metagenomics: Application of Genomics to Uncultured Microorganisms

Jo Handelsman*
Department of Plant Pathology, University of Wisconsin–Madison, Madison, Wisconsin

INTRODUCTION ..................................................................................................................669
HISTORY OF THE CULTURE DIVIDE ...........................................................................670
Early Microbiology and the Microscope ...........................................................................670
Modern Microbiology—a Pure Culture Is Not Enough ......................................................670
THE PARADIGM SHIFT ...................................................................................................671
rRNA Analysis and Culturing ............................................................................................671
METAGENOMICS—CULTURE-INDEPENDENT INSIGHT ...................................................673
APPROACHES TO METAGENOMIC ANALYSIS ...............................................................673
  Sequence-Based Analysis ...............................................................................................675
  Functional Metagenomics ...............................................................................................675
  Heterologous expression .................................................................................................675
  Identifying active clones—screens, selections, and functional anchors .........................675
ECOLOGICAL INFERENCES FROM METAGENOMICS ..................................................676
  Symbiosis ......................................................................................................................676
    Buchnera-aphid symbiosis ............................................................................................676
    Protobacterium-tube worm symbiosis .........................................................................676
  Competition and Communication ..................................................................................677
  What can metagenomics tell us about microbial competition and communication? ........677
  Role of small molecules ...............................................................................................677
  Sequence-based screening for small molecules .............................................................677
  Antibiotics as signal molecules ....................................................................................678
Biogeochemical Cycles .....................................................................................................680
  Acid mine drainage ........................................................................................................680
  Sargasso Sea ..................................................................................................................680
  Population Genetics and Microheterogeneity ................................................................680
CONCLUSIONS AND FUTURE DIRECTIONS ...................................................................682
REFERENCES ..................................................................................................................682

2. More discussion is coming
Summary

• Proteomics
  – Application in marine environment.

• Metagenomics and its application
  – Terminology and goal
  – Commons vectors for metagenomics
  – Type of metagenomics and their goals
  – Application
    • Review paper
    • More discussion