Use and Evaluation of New Genomic Techniques To Understand Microbes In Waterways

Michael J. Sadowsky, Cliairessa Brown and Chanlan Chun

University of Minnesota
Department of Soil, Water and Climate; and BioTechnology Institute
Fecal Pollution

- Indicator organisms imply fecal pollution
- The fecal coliform bacteria, enterococci, and *E. coli* are primary indicators of fecal contamination
- Conventional microbiological methods cannot determine sources of fecal pollutants
Progress

• Problems Associated with Traditional Coliform Testing
• We now have a much better idea of what is there and where it is coming from but...
The problem of throughput!

Large Genetic Diversity in Indicator Bacteria require analysis of a large number of samples!
Overall Hypotheses

• Fecal indicator bacteria are most likely coming from human sources.

and

• Fecal indicator bacteria signal the possible presence of human pathogens.
• Human fecal bacteria are more harmful than those from other animals.
• But we have not really tested this directly (no undergrads or prisoners).
Can Metagenomics (more specifically high throughput 16S-based taxon analyses) save the day?
What is Metagenomics?

- The study of the totality of genetic material (genomes or their fragments) recovered directly from environmental samples.

- Many types of Metagenomic Analyses
  a. Diversity (16S rDNA)
  b. Microbial Community Analyses
  c. Functional Gene Discovery Analyses
Why use Metagenomic Analyses

• The majority of microorganisms in environmental and animal samples (estimated to be less than 1%) remain uncultured or non-culturable.

• *E. coli* and enterococci grow in the environment.

• *E. coli* does not show up in most metagenomic analyses it is only 1 of a 1000 species of gut microbes.
General Approaches

1. Directly Pyrosequence
   - Isolation of DNA from environmental sample
   - Manipulation of DNA
   - Ligation of fragments with vectors
   - BAC vector
   - High molecular weight DNA
   - Amplify 16S rRNA -sequence

2. Metagenomic library
   - Cultured clones
DNA Sequencing

• High-Throughput, Large Scale, and Inexpensive DNA Sequencing Technology
Illumina Sequencing Platforms

• HiSeq
  – 2 x 250 bp read length; 170-200 million output
  – ~8-10 days ~ 80-120 samples
• MiSeq
  – 2 x 250bp read length; 4.5-7.1 million total output - ~24 hours
Produces 100s of millions of DNA sequences
Microbes and The Tree of Life

Our studies are only targeting 1 domain of life – the bacteria
Minnesota Mississippi Metagenome Project

M3P

Lock and Dam Number 7, Wikimedia.org
The Itasca Headwaters are relatively pristine – but....
.. as is passes through the U.S. heartland, anthropogenic activity alters is structure and function.
Confluence of other U.S. Rivers lead to further inputs into the Mississippi
Alterations in the Gulf of Mexico and Delta
Summary of diversity indices among different sample volumes. $S_{obs}$ is the number of OTUs/

<table>
<thead>
<tr>
<th>Volume Filtered (L)</th>
<th>$S_{obs}$</th>
<th>Shannon</th>
<th>NP_Shannon</th>
<th>Simpson</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.33</td>
<td>9218</td>
<td>4.9</td>
<td>4.9</td>
<td>0.03</td>
</tr>
<tr>
<td>1.00</td>
<td>8990</td>
<td>4.8</td>
<td>4.8</td>
<td>0.04</td>
</tr>
<tr>
<td>2.00</td>
<td>8490</td>
<td>4.7</td>
<td>4.7</td>
<td>0.05</td>
</tr>
<tr>
<td>6.00</td>
<td>8432</td>
<td>4.6</td>
<td>4.6</td>
<td>0.05</td>
</tr>
<tr>
<td>Mean</td>
<td>8783</td>
<td>4.7</td>
<td>4.8</td>
<td>0.04</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>383</td>
<td>0.1</td>
<td>0.1</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Rarefaction curves

Number of Sequences

Number of OTUs (96%)
Most Abundant Phyla

![Graph showing the abundance of different phyla across various sampling sites. The x-axis represents the sampling sites, and the y-axis represents the percentage of sequence reads.]
Genera of Classified Bacteria Recovered at Each Site
What are OTUs?

Operational taxonimic units – sequence data Independent of Taxonomy
Unique OTUs
Genera of Classified Bacteria Identified in Source Fecal Samples
Not sure this fits here. Maybe after the mothur pipeline.
Christopher Staley, 10/12/2012
NMDS of Community Composition by Site

NMDS Axis 1

NMDS Axis 2

- Headwaters
- Southern Agricultural Sites
- Northern Agricultural Sites
- Urban Sites
- Hastings
- Twin Cities
- Minnesota River
- Confluence
- La Crescent
- Clearwater
- St. Cloud
- St. Croix River
- Rochester
- Itasca
Staley, et al. Figure 6. Consensus inferred Bayesian network relating bacterial orders (blue circles), physicochemical parameters (green diamonds), chemicals (nutrients and ions; yellow hexagons), E. coli (pink circle), and land use (orange squares). Solid black lines indicate positive associations while dashed red lines indicate negative associations and arrows are directed from parent to child. Edge length is arbitrary. Asterisks (*) designate orders that are unclassified or designated *incertae sedis.*
Understanding Metagenomes is of Primary Importance

Next Generation Fecal Taxon Libraries - FTL

- Contains all the taxonomic units and OTUs in pooled fecal samples from known animal sources.

- Gives information about fecal and environmental bacteria in samples.

- Can be used for Microbial Source Tracking
Overall Goals

Match DNA Sequences in Data Sets created from feces of known animals to those recovered in rivers samples.

1. Shared OTUs – Taxonomy Independent
2. Shared Taxonomic Units - Genera
What are Shared OTUs?

OTUs from fecal and environmental DNA samples
Independent of Taxonomy
OTU based network analysis of sample A, sample B, AND sample C

Network image of shared OTUs

Shared OTUs
New MST method with Next Generation Sequencing technique

*Environ. Sci. Technol. 2010, 44, 7777–7782*

**Use of Barcoded Pyrosequencing and Shared OTUs To Determine Sources of Fecal Bacteria in Watersheds**

TATSUYA UNNO, †JEONGHWAN JANG, †DUKKI HAN, †JOON HA KIM, ‡MICHAEL J. SADOWSKY, §OK-SUN KIM, ‡‡JONGSIK CHUN, ‡§AND HOG-GIL HUR*, †‡, ‡§

Department of Environmental Science and Engineering and International Environmental Research Center, Gwangju Institute of Science and Technology, Gwangju 500-712, Republic of Korea, Department of Soil, Water, and Climate and BioTechnology Institute, University of Minnesota, St. Paul, Minnesota 55108, and School of Biological Sciences, Seoul National University, N570, 56-1 Daehak-dong, Kwanak-gu, Seoul 151-742, Republic of Korea
Result: Number of shared OTUs

- Urban area
- Open Area
- Agricultural area

- Pig contamination
- Human
- Mixed
- Dairy cattle
- Beef cattle
- Wild goose
- Duck
- Chicken

Legend:
- Pig
- Mixed
- Dairy cattle
- Beef cattle
- Wild goose
- Duck
- Chicken
- Human
Development of **automated** MST system

Applying batch program to automate pyrosequencing analysis

Integrated Online System for a Pyrosequencing-Based Microbial Source Tracking Method that Targets Bacteroidetes 16S rDNA.

Research Questions

1. Are bacterial communities in environmental and fecal samples similar?
2. Are the bacterial composition of different animals distinguishable from each other?
3. Can the fecal microbiomes of different animals be differentiated using a NGS MST method?
Sample Collection → Sample Processing → DNA Isolation → Next Generation Sequencing → Data Analysis

Southworth Marsh
storm drain

St. Louis River

[Map]

[Image of storm drain]

[Image of Southworth Marsh]

[Image of St. Louis River with geese]
Brewery Creek storm drain

Storm drains are known microbial contaminant sources

Had some of the highest levels of *E.coli* across summer
Conclusions

• Bacteria in animal fecal samples and environmental samples are different but some cluster together.
• Community structure is different between animals.
• Differences in the less abundant bacterial members in feces may allow differentiation of animals.
The most sensitive source tracking tool of all is
Acknowledgments

College of Biological Sciences
Federal Stimulus Funds
LCCMR

Sea Grant

Claireessa Brown and Chanlan Chun
Brian Gibbens
Chris Staley
Jane Phillips
Trevor Gould
Tatsu Unno
Bruce Jarvis
James B. Cotner
Thank you for the Opportunity to Speak with you.

Thanks for your attention!