ADVANCES IN METAGENOMICS FOR CHARACTERIZING MICROBES IN THE ENVIRONMENT

Jason Dobranic, Ph.D.
VP Microbiology & Life Sciences
EMSL Analytical, Inc.
Swimming in contaminated beach waters
- 120 million cases of gastrointestinal (GI) disease
- 50 million cases of acute respiratory diseases world-wide

Shifts in rainfall and temperatures driven by climate change
- exaggerate microbial contamination issues
- increase risk of water-borne disease in coastal and inland regions
WATER QUALITY INDICATORS: CULTURE-BASED BIAS

- EPA established indicators
  - Rely on culturing fecal indicator bacteria to predict risk
  - Underestimate the densities and diversity of microbes in environmental samples
- Non-culturable microbes
  - May be more important as etiological agents
  - Better indicators of fecal contamination
  - Useful in source tracking
Evaluation of the 2012 RWQC

- other possible fecal indicator such as *C. perfringens*, *Bacteroides*, and human enteric viruses
- or new genomic approaches (e.g., metagenomics) should also be evaluated in developing new RWQC.

2017, 5 YR review

- majority of experts (>90%) favor a PCR-based methodology for MST
- focused only on qPCR methods
- identification of top human-associated fecal markers
  - HF183/BacR287
  - HumM2

Molecular-based methods overcome culture limitations

- Gut microbiota are now being considered as alternate indicators for human and animal fecal contamination
- Characterize and map the different microbial community populations associated with human sewage and various animal feces
- Monitor water quality using a single, comprehensive suite of microorganisms by analyzing patterns of relative abundance
GLOBAL FUNGAL SPECIES RICHNESS

- ~100,000 species described
- ~600,000 from conservative estimates
- ~1 billion from optimistic estimates
- Newer molecular methods are increasing our knowledge of microbial diversity
Direct genetic analysis of genomes contained in environmental samples
Targeted vs. shotgun approach
Most data sets related to prokaryotic microbial communities
Increased awareness and studies on fungi

- Culture-free method (only small amount culturable in environmental samples)
- Entire microbial community analyzed in a sample
- 16s rRNA gene used for bacteria (universal marker)
- 18s rRNA gene or ITS regions for fungi
- Bioinformatics is key (computational tools chosen for sequence analysis)
Sample Processing Pipeline:

1. Sampling (larger volumes increases chances of retrieving rare groups)
2. Filtration
3. Cell lysis/DNA extraction
4. Gene amplification
5. Sequencing library construction
6. Next-gen sequencing (MiSeq)
7. Bioinformatic data analysis
8. Results interpretation

Adapted from: http://hulab.ucf.edu/research/projects/metagenomics/introduction.html
CONSIDERATIONS

- Efficiency of extraction is very important
- Amount and type of non-biological material in a sample may interfere with extraction, quantification, and amplification
- PCR inhibitors must be removed in clean-up step
  - e.g., cations, humics, pollutants can interfere
Characterize the predominant microbes in the gut of humans and predominant microbes in the gut of various other animals. Community “fingerprint” can be established to determine if the source of contamination is from human or other animal feces.

Humans are made up of around 23,000 genes but the human body also contains 3,000,000 non-human genes.
More studies to determine if metagenomics methods are feasible for general use and if the generated data can be used to improve future RWQC.

As an example, McLellan et al., determined that members within Clostridales and Lachnospiraceae were among the most dominant bacterial groups in Milwaukee’s wastewaters, and these bacteria were found in an estuary and bathing beaches of Lake Michigan following storm events.
APPLICATIONS

- Biodiversity studies & community composition
- Study metabolic genes from microbial communities
  - Biased towards previously known genes
- Discovering novel organisms and genes in the environment
- Bioremediation studies
GLOBAL OCEAN SAMPLING EXPEDITION (GOS)

- Started in 2003 by Craig Venter
- Goal to assess the genetic diversity of marine microbial communities around the Earth
- Starting in Halifax, Canada, samples were collected at sites along the U.S. east coast, Gulf of Mexico, Galapagos Islands, central and south Pacific Oceans, Australia, Indian Ocean, South Africa, and across the Atlantic back to the U.S.
- A total of 41 different samples were taken from a variety of aquatic habitats
- Yielding 6.4 million contiguous sequences
- Goal of understanding microbial diversity and function
Upstream preserved and non-urbanized area vs. a polluted urbanized area with discharged sewage

Non-urbanized area was overrepresented by genera of ubiquitous microbes that act in the maintenance of environments

Urbanized metagenome was rich in genera pathogenic to humans

Antibiotic resistance, metal-resistance, and stress response-related genes were disseminated in the urbanized environment
Fig. 1 – Phylogenetic profile of the freshwater urbanized and non-urbanized metagenome using the 16S rDNA sequences and all the shotgun reads.
Fig. 3 – Metabolic pathways of methane, nitrogen and sulfur from the freshwater metagenomes. The numbers label in black indicate the number of sequences from urbanized metagenome affiliated with the KEGG function and the number label in red indicate the non-urbanized metagenome. Data obtained through MEGAN.
- one of the largest tributaries of the Chesapeake Bay
- home to more than 25 million Virginians
- >1500 point sources permitted to discharge pollutants
- nonpoint source pollutants from urban, agricultural, wildlife, and transportation runoff
- heavily accessed for recreation
Current recreational water monitoring practices (e.g., *E. coli* and coliform testing)
- only as coarse indicators of potential contamination
- provide little information on the diversity, source, ecology, or evolution of organisms that cause WBDOs
## PUTATIVE ROLES OF THE MOST ABUNDANT BACTERIAL AND EUKARYOTIC OTUs

### Table 11: Putative roles of the most abundant bacterial and eukaryotic OTUs

<table>
<thead>
<tr>
<th>Putative Role</th>
<th>JREM1 Rifles_WGS</th>
<th>JREM1 Rifles_16S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% OTUs</td>
<td>reads</td>
</tr>
<tr>
<td><strong>Bacteria</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common free-living</td>
<td>52</td>
<td>1072474</td>
</tr>
<tr>
<td>Pollution degraders</td>
<td>22</td>
<td>484716</td>
</tr>
<tr>
<td>Sludge, industrial, and medical waste</td>
<td>12</td>
<td>133200</td>
</tr>
<tr>
<td>Pathogens (human, crops, and fish)</td>
<td>14</td>
<td>297920</td>
</tr>
<tr>
<td><strong>Eukaryote</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common free-living</td>
<td>52</td>
<td>3468</td>
</tr>
<tr>
<td>Terrestrial/agriculture/aquaculture</td>
<td>33</td>
<td>1123</td>
</tr>
<tr>
<td>Pathogens (human, crops, and fish)</td>
<td>14</td>
<td>261</td>
</tr>
</tbody>
</table>

*not examined*
STUDY CONCLUSIONS

- Microbial community closely mirrors the upper Mississippi River
- A river microbial response exists to anthropogenic pollution
- Limited sampling, generalizations cannot be made regarding spatio-temporal distributions
- Further studies should allow health agencies to better identify organism specific health risks and to enhance waterborne disease prevention efforts
DEVELOPMENT, CALIBRATION, AND VALIDATION OF A SIMPLE TOOL FOR GUIDING MOLD INSPECTION AND REMEDIATION IN U.S. HOMES

- Dr. Peccia (Yale University), Dr. Shaughnessy (U. of Tulsa), Jason Dobranic (EMSL)
- Mold assessment in homes using NextGen Sequencing
- Water damaged vs. non-water damaged
- Nationwide HUD funded research
CONCLUSIONS

- Instrumentation and technical knowledge achieved
- Improvements needed on bioinformatics tools (ITS databases), fungal genomics, & overall knowledge of fungi in the environment
- Momentum is there to use metagenomic data for microbial characterization of ambient water
- Metagenomic approach will be common place in the future
Jason Dobranic, Ph.D.
Vice-President of Microbiology & Life Sciences
EMSL Analytical, Inc.
jdobranic@emsl.com
www.emsl.com