The Application of Next Generation Sequencing in Metagenomics

IEH Laboratories & Consulting Group
Sam Myoda, Ph.D.
Executive Vice President, IEH Laboratories & Consulting Group
sam@iehinc.com
mobile: (831) 261-007
Call from a Regulator: Your Product Has Been Implicated by Whole Genome Sequencing (WGS) to an Ongoing Outbreak

This is the typical impression of WGS, however: WGS is so much more, e.g., shelf life extension, spoilage analysis, transient vs. resident determination, organism identification, population profiles, etc., etc.
Next Generation Sequencing (NGS)

NGS refers to the latest technology to determine the order of DNA base pairs (sequences) in an organism. There are two major applications of NGS:

1) Whole Genome Sequencing (WGS): looks at the entire genome of a single organism
2) Metagenomics: looks at a small fragment of DNA of every organism in a population to determine the community profile
All organisms have DNA, it is the blueprint that is responsible for all attributes of the organism. All the DNA in an organism that creates this blueprint, e.g. genes, etc. is called the organism’s **Genome**.

DNA is a long string of paired bases in a double helix formation. There are four nucleotide bases: adenine (A), cytosine (C), guanine (G) and thymine (T).

A always pairs with T, C always pairs with G.

Humans have ~3,200,000,000 base pairs

*Salmonella* has ~5,000,000 base pairs.

**WGS** is mapping the order of all the base pairs in a single organism.
Metagenomics

The difference between metagenomics and WGS is that in metagenomics, only a small portion of the genome is sequenced. However, it is done on a mixed population so multiple organisms are sequenced simultaneously. The result is a profile of the entire population of organisms. In WGS, the entire genome of only 1 organism is sequenced.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Target gene region</th>
<th>Primer set</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteria</td>
<td>16S</td>
<td>341F-805R</td>
<td>Greengenes</td>
</tr>
<tr>
<td>Archaea</td>
<td>16S</td>
<td>340F-806R</td>
<td>Greengenes</td>
</tr>
<tr>
<td>Fungi</td>
<td>ITS2</td>
<td>BITS-B58S3</td>
<td>UNITE</td>
</tr>
</tbody>
</table>
Identification of Population Profiles using Metagenomics

Genetic Identification:
Direct Metagenomic Testing via Next Generation Sequencing

Microbial ID:
- Bacillus sp.
- Lactobacillus sp.
- Enterobacteraceae sp.
- Yeast
- Mold
Metagenomics: Spoilage Analysis

Traditional spoilage analysis is done by using various tests and medias to detect potential spoilage organism. This is a very time consuming and expensive approach that may or may not be successful in identifying the spoilage organism(s). Using metagenomics, the entire population of organisms, both from “good” or “fresh” product may be compared with spoiled product to identify the culprit.
Spoiled, “Bloated” Yogurt Cup

- In addition to yogurt starter cultures, twenty bacterial genera were found (4 sample sets, 48 samples)
- Mucor was dominant fungal isolate

<table>
<thead>
<tr>
<th>Taxonomic ID</th>
<th>Set 1</th>
<th>Set 2</th>
<th>Set 3</th>
<th>Set 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mucor circinelloides</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mucor circinelloides</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mucor sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncultured Ascomycota sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mucoraceae sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mucoraceae sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mucor sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncult. Saccharomyces sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
"good" product

No need to culture and differentiate all potential spoilage organisms. *Clostridium algidicarnis* directly identified as the causative spoilage organism using metagenomics.
Spoiled Veggie Puree

“good” product

spoiled product

In this case, fungi, *Cryptococcus sp.*, not bacteria caused spoilage
### Spoiled “Blown” Juice Bottles

<table>
<thead>
<tr>
<th>IEH work order #</th>
<th>Spoiled</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1108360-01 1109360-01 1109360-01 1109360-01 1109360-01 11109360-01 11109360-01 11109360-01 11109360-024</td>
<td>1108360-01 11109360-01 11109360-01 11109360-01 11109360-01 11109360-01 11109360-01 11109360-01 11109360-024</td>
</tr>
<tr>
<td><strong>Species</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calothrix parietina</td>
<td>45% 46% 15% 27% 4%</td>
<td>89% 94% 95% 99% 95%</td>
</tr>
<tr>
<td>Leuconostoc citreum</td>
<td>46% 45% 70% 58% 7%</td>
<td>9% 4% 0% 0% 0%</td>
</tr>
<tr>
<td>Fructobacillus pseudofaciunus</td>
<td>5% 5% 7% 6% 8%</td>
<td>1% 0% 0% 0% 0%</td>
</tr>
<tr>
<td>Sulfurimonas denitrificans</td>
<td>0% 0% 0% 0% 0%</td>
<td>0% 0% 0% 0% 0%</td>
</tr>
<tr>
<td>Leuconostoc palmae</td>
<td>2% 2% 4% 3% 5%</td>
<td>0% 0% 0% 0% 0%</td>
</tr>
<tr>
<td>Leuconostoc gascomitatum</td>
<td>1% 1% 1% 1% 3%</td>
<td>0% 0% 0% 0% 0%</td>
</tr>
</tbody>
</table>

### Diagram

The diagram illustrates the relationship between the control and spoiled juice bottles, showing the production of gas in spoiled bottles.
Metagenomics: Shelf Life Extension

Traditional shelf life studies are done using total counts and “indicator” organisms. Using metagenomics, the optimal population of microbial flora may be identified and processing parameters developed to achieve that optimum. These parameters may include but are not limited to:

- Selection of cultivars, breeds, amendments, nutrients, etc.
- Temperature, e.g., during harvesting, storage, cooling, processing, transporting, etc.
- Antimicrobials used during processing, e.g., chorine, quat, PAA, lactic acid, etc.
- Packaging, e.g., modified atmosphere, film permeability, size and configurations, color, etc.
Metagenomics
Shelf Life Extension Flow Chart

Metagenomic analysis of raw material to determine starting point population

Metagenomic analysis of Product with “Best” Shelf Life to determine optimal end point population

Note: this step is used as a basis for estimating what processes to consider in step 2, it is often skipped

Multiple process runs using different options under consideration

Metagenomic analysis of all runs at different time points throughout the desired shelf life

Select optimal parameters
Metagenomic Analysis to Evaluate Heat Treatment Efficacy on Population Profile

Brand D – Raw Material

- Calothrix: 65.9%
- Rickettsia: 0.9%
- Unclassified: 2.3%
- Nostoc: 0.3%
- Dolichospermum: 0.3%
- Thiomonas: 0.3%
- Microcoleus: 0.2%
- Bacillus: 0.1%
- Heliorestis: 0.1%
- Candidatus Blochmannia: 0.1%
- Erwinia: 0.1%
- Symploca: 0.1%
- Other: 0.3%

Brand D – Roasted Exit

- Calothrix: 30.9%
- Methylobacterium: 18.0%
- Bacillus: 10.4%
- Erwinia: 9.8%
- Rickettsia: 6.7%
- Janthinobacterium: 4.7%
- Unclassified: 2.9%
- Candidatus Solibacter: 2.8%
- Pseudomonas: 1.6%
- Paenibacillus: 1.1%
- Alicyclobacillus: 1.1%
- Acinetobacter: 0.9%
- Corynebacterium: 0.9%
- Enhydrobacter: 0.9%
- Thermus: 0.6%
- Parascardovia: 0.5%
- Klebsiella: 0.6%
- Other: 5.5%
Metagenomics: Identification of Previously Unknown Organisms

(Clostridium tepidum sp. nov.)

Metagenomics can be combined with WGS to identify unknown organisms. First, the population profile within a sample is identified. Second, a pure culture of the potential novel organism is isolated. WGS of the culture is then used identifies the organism. Example: A new species of *Clostridium*, *Clostridium tepidum sp. nov.* isolated from non-dairy protein shakes in bloated bottles was identified using metagenomics and WGS. It is related to *C. sporogenes* and is more thermophilic and less halotolerant than *C. botulinum*. 
Metagenomics: Identification of Previously Unknown Organisms (*Clostridium pearianum* sp. nov.)

- **Preliminary identification**
  - Identified the pear spoilage organism as *Clostridium spp.* from 16S community metagenomics.

- **Isolation & Characterization**
  - Whole genome sequencing (WGS) of the bacterial isolate revealed a new species of *Clostridium* named *C. pearianum* sp. nov., a close relative of *C. pasteurianum*.

- **Development of a detection kit**
  - Developed a PCR based detection system using two distinct genetic markers obtained from WGS.

Figure 2. The strain was isolated on PYG-CaCO3 agar plate. This plate has been incubated at 37°C for 5 days in anaerobic condition. The clear zone was appeared around umbostrate colonies that indicated this strain is producing acids in anaerobic condition.
Microbial Contamination In the Oil and Gas Industry

Problem Type  Microorganisms implicated
- Souring / H2S production        Sulfate Reducing Bacteria (SRB)
- Anaerobic corrosion                 Sulfate Reducing Bacteria (SRB)
- Clogging                                      Iron-oxidizing bacteria
 remorseforming bacteria (various genera incl. Pseudomonas sp.)
 Molds
- Oil degradation                         Aerobic and anaerobic hydrocarbon degrading bacteria
(Pseudomonas sp., methylotrophs, SRB)

Microbial Impact on Infrastructure & Products:

- Sulfate-reducing bacteria (SRB) pose a major threat to infrastructure and products in the oil- and gas industry.
- SRB cause souring (accumulation of hydrogen sulfide) and anaerobic corrosion of metal.

Many small and medium-size oil and gas operations have very sporadic or no control over microorganisms.

Microbial Souring:

- VFA, Alcohols, Sugars, Hydrocarbons, Crude Oil, Hydrogen
- CO₂ + VFA
- H₂S

Anaerobic Corrosion:
Microbial Contamination In the Oil and Gas Industry (2)

Detailed Genetic Identification Of Microbial Contaminants:

Rapid identification of potential microbial hazards in oil wells, production water, refinery systems and storage facilities using Next-Generation Sequencing Technology.

- Rapid High-Throughput Species Identification:

- Determination of Metabolic Diversity:

Accelerated Microbial Corrosion Potential:

- Rapid identification of corrosion potential before major damage occurs.
- Effective routine monitoring tool to keep risk under control.

Total Microbial Load:
- Microbial load gives information about the degree of microbial contamination and potential risk for damages.

Biocide Efficacy Test:
- Lab scale evaluation of biocides in the context of operation chemistry.
- Fast identification and monitoring of biocide efficacy.

Isolation and Characterization of Sulfate-reducing Bacteria:
Metagenomic Analysis to Investigation Population Profile Correlation to a Specific Target (aflatoxin in peanuts)

Brand D (~20 ppb aflatoxin)
Thank You

sam@iehinc.com
831.261.0076