Standards for Microbiome and Metagenomic Measurements

Scott Jackson
Leader, Complex Microbial Systems Group
Biosystems and Biomaterials Division
MML

MML Microbial Metrology Mission Statement

“To develop advanced measurements that will permit the exploitation of microbes to promote human health, precision medicine and advanced manufacturing”
WHAT/WHO IS NIST?

- Non-regulatory federal agency within the U.S. Department of Commerce.
- Located in Gaithersburg, Maryland and Boulder, Colorado
- ~5000 Scientists and Engineers (4 Nobel Laureates)
- Founded in 1901
- ~$1B Annual Budget
- NIST's Mission: “To promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology”
The FDA has also been active in addressing other regulatory issues surrounding personalized medicine. Along with authorizing the Illumina technology for marketing, the FDA recognized the need for reference materials and methods that would permit performance assessment. As a result, the FDA collaborated with the National Institute for Standards and Technology (NIST) to develop reference materials consisting of whole human genome DNA, together with the best possible sequence interpretation of such genomes.

The FDA based its decision to grant marketing authorization for the Illumina instrument platform and reagents on their demonstrated accuracy across numerous genomic segments, spanning 19 human chromosomes. Precision and reproducibility across instruments, users, days, and reagent lots were also demonstrated.
HUMAN GENOME IN A BOTTLE CONSORTIUM (GIAB)
HOSTED BY US NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

Goal: Provide infrastructure to assess confidence in human variant calls

Samples are commercially available.

High-accuracy reference data for these samples.

Tools to facilitate their use.

• With the Global Alliance Data Working Group Benchmarking Team

U.S. Introduces New DNA Standard for Ensuring Accuracy of Genetic Tests

WASHINGTON — The federal government opened the door to a new era of genetic medicine on Thursday by introducing a standard way to ensure the accuracy of DNA tests used to make treatments for individual patients.

Scientists have identified hundreds of genetic mutations that appear to increase the risk of diseases, including cancer. At times, doctors and patients are confused about the likelihood of these mutations. The new effort aims to ensure that different laboratories can use samples of the same blood or tissue.

The National Institute of Standards and Technology said Thursday that it had developed "reference materials" that could be used by laboratories to determine whether their machines and software are properly analyzing a person's genetic makeup, or genome.

The institute disseminates such reference materials for thousands of products including steel, cement and personal biology. These materials are used for quality purposes — to calibrate instruments, to make sure
FDA

Industry Standards

Industry

NIST

MATERIAL MEASUREMENT LABORATORY
The Human Microbiome

- The human microbiome plays a major role in human health and disease and represents an understudied diagnostic and therapeutic target.
- Many different species of bacteria are found at different body sites – and perform different roles.
- Many many diseases and health conditions have been linked to dysbiosis: including cancer, mental illness, diabetes, obesity, autoimmune disorders, etc.

“Precision medicine is about moving beyond this one-size-fits-all kind of approach to medicine and instead approaching disease prevention and treatment by taking into account individual differences in people’s genes, their microbiomes, their environments, and their lifestyles,”

Jo Handelsman, White House OSTP
Bias in Microbiome Measurements

**Sample Handling:**
- Isolation
- Storage
- Shipping

**DNA Extraction Efficiency:**
- Different Extraction Efficiencies from Different Cell Types.
- Different DNA Extraction Methods (bead beating vs. enzymatic vs. chaotropic, etc.)
- Extracellular Matrix “Goop”

**16S PCR:**
- 16s rDNA Amplicons
- 16s rDNA PCR: Primer selection and Number of PCR cycles

**NGS Library:**
- Transposon
- Mechanical Shearing
- Nuclease
- Adapter Ligation
- PCR-Bias

**NGS:**
- Next-Gen Sequencing: Different Platforms Exist. Each With Inherent Biases

**Bioinformatics:**
- Bioinformatic Interpretation of the NGS Data: Different Bioinformaticists Analyze and Interpret the Data Differently
ZymoBIOMICS™ Microbial Community

<table>
<thead>
<tr>
<th>Species</th>
<th>mOTU counts</th>
<th>mOTU Abun. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Bacillus subtilis</em></td>
<td>9048</td>
<td>11.86</td>
</tr>
</tbody>
</table>

**GRAM-NEGATIVE**

- *Pseudomonas aeruginosa*: 4484, 5.88

**GRAM-POSITIVE**

- *Propionibacterium acnes*: 1, 0.0013

---

Expected
Measured (Operator 1)
Measured (Operator 2)

**Organism**

PA, EC, SE, LF, EF, SA, LM, BS

Relative Abundance
MICROBIOME MEASUREMENTS ARE HARD

"I asked two different companies to analyze my gut microbiome. American Gut (left) gave nearly opposite results to those from uBiome (right) with respect to the major phyla of bacteria in a duplicate sample."

Or as Knight puts it, "All sorts of unlikely things are possible, and finding out which one is true is difficult."

"But DNA extraction is not the only thing that could go wrong. It seems that every step of the process — from how you collect the sample through the computer programs used to analyze the DNA data — is a potential culprit."

"Another blogger, who is a bioinformatician, got different results than American Gut reported to him when he used his own software to analyze their raw data."

Rashmi Singha, an epidemiologist at the National Cancer Institute stated: "The lack of reproducibility between studies was frustrating. To me it seemed like cowboy country. It needed to have some kind of order."

"The point is, scientists are trying to find ways to standardize microbiome studies so that they can directly compare results. They don’t yet have the answers, but they will take the first steps toward figuring it out at a workshop this fall."
MICROBIOME DIAGNOSTICS

CLIA-certified and CAP-accredited clinical laboratory

Helping You Help Patients

Using advanced DNA sequencing technology, SmartGut™ accurately identifies dozens of important species and genera so you can diagnose your patient’s condition with greater precision.

Learn more
LIVE MICROBIAL THERAPEUTICS

“It’s 1982 Again”

-Scott Jackson (almost daily)
A novel Class of BioTherapeutics: Live Microbial Therapeutics (LMT)

MANIPULATING THE MICROBIOME
In the past five years, dozens of companies have emerged to develop treatments based on, or targeting, our microbiota

<table>
<thead>
<tr>
<th>THERAPEUTIC APPROACH</th>
<th>NOTABLE FUNDING/PARTNERSHIPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avid Biotics</td>
<td>Engineered proteins to target select strains of microbiota</td>
</tr>
<tr>
<td>Enterome Biosciences</td>
<td>Microbiome-derived small molecules</td>
</tr>
<tr>
<td>Rebiotix</td>
<td>Live human-derived microbes to treat Clostridium difficile infection</td>
</tr>
<tr>
<td>Vedanta Biosciences</td>
<td>Rationally designed cocktail of microbiome-derived bacteria</td>
</tr>
<tr>
<td>Second Genome</td>
<td>Small molecules to disrupt microbiome-host interaction</td>
</tr>
<tr>
<td>Seres Therapeutics</td>
<td>Orally delivered bacterial spores to restore symbiosis</td>
</tr>
<tr>
<td>Symberix</td>
<td>Small molecules to target bacterial enzymes</td>
</tr>
<tr>
<td>Synlogic</td>
<td>Engineering microbiome-derived bacteria to perform a therapeutic task</td>
</tr>
</tbody>
</table>

SBIR = Small Business Innovation Research. SOURCE: Company information

1. SERES Therapeutics
2. Whole Biome
3. AO Biome
4. Naked Biome
5. MicroBiome Therapeutics
6. Assembly Biosciences
7. Second Genome
8. Janssen
9. ViThera Pharmaceuticals
10. Vedanta
11. Rise Therapeutics
12. Enterome
13. Ritter
14. Synthetic Biologics
15. Abbot
16. Pfizer
17. Enterome
18. Avid
19. Synlogic
20. Symberix
21. Rebiotics
Question: If handed a pill containing a mixture of different bacterial species and asked to demonstrate the “identity” or the “purity”, what measurement would you perform?
THE NIST MONOCLONAL ANTIBODY (NISTMAB)

“NIST's monoclonal antibody reference material is a standard benchmark for all. With it, manufacturers and regulators can better assess methods for analyzing and assuring the quality of complex biological drugs.”

The standard is an antibody protein—consisting of more than 20,000 atoms—analyzed so thoroughly that the material can be used by organizations around the globe to verify and improve their analytical methods for quality control.
Neutron Measurements for Predicting & Understanding Protein Stability

PI: Marcus Cicerone

Protein Stability During Storage

- 1/3 of therapeutic proteins are freeze-dried, but formulation for freeze-drying is empirical with 60% success rate
- Neutron scattering discovers new metric, fast $\beta$ relaxation, correlating with long-term protein stability
- Bench-top optical method developed at NIST to measure $\beta$ relaxation

Degradation Tracks $<u^2>^{-1}$

Cicerone & Douglas, Soft. Matter 8 2983 (12)
THE NATIONAL INSTITUTE FOR INNOVATION IN MANUFACTURING BIOPHARMACEUTICALS (NIIMBL)

- [http://www.niimbl.us/](http://www.niimbl.us/)

- A Public-Private partnership dedicated to biopharmaceutical manufacturing innovation

- 140+ Partners: industry, academics, non-profits, and government (state and federal)

- The NIIMBL mission is to accelerate biopharmaceutical manufacturing innovation, support the development of standards that enable more efficient and rapid manufacturing capabilities, and educate and train a world-leading biopharmaceutical manufacturing workforce, fundamentally advancing U.S. competitiveness in this industry.

- Investment of $70M over the next 5 years by the National Institute of Standards and Technology (NIST)
PATHOGEN DETECTION – CLINICAL DIAGNOSTICS FOR INFECTIOUS DISEASE

Metagenomics

Microbiome Measurements

Pathogen Detection

Jason Kralj
An “Unbiased” Approach for Pathogen Detection: Shotgun Metagenomics via Next-Gen Sequencing

Clinical Sample

DNA extraction

Total DNA (human + pathogen)

Fragmentation

Shotgun Library

DNA sequencing

Metagenomic Sequence Data

Bioinformatically Search for Pathogen-Specific Signatures
NIST is currently developing microbial reference materials... will be a "valuable tool for use.." in evaluating new diagnostic devices.

"FDA intends to regulate Infectious Disease NGS Dx devices as systems, including all of the components necessary to generate a result."

‘NIST is currently developing microbial reference materials.... will be a “valuable tool for use..” in evaluating new diagnostic devices. ‘
### Table: Abundance of Pathogens

<table>
<thead>
<tr>
<th>Source</th>
<th>Abundance*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human DNA</td>
<td>$10^1$</td>
</tr>
<tr>
<td>Pathogen #1</td>
<td>$10^{-1}$</td>
</tr>
<tr>
<td>Pathogen #2</td>
<td>$10^{-2}$</td>
</tr>
<tr>
<td>Pathogen #3</td>
<td>$10^{-3}$</td>
</tr>
<tr>
<td>Pathogen #4</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td>Pathogen #5</td>
<td>$10^{-5}$</td>
</tr>
<tr>
<td>Pathogen #6</td>
<td>$10^{-6}$</td>
</tr>
</tbody>
</table>

*Abundance* is genome copy number relative to human reference DNA

### Graph: Relative Genome Abundance

- **Y-axis:** Relative Genome Abundance
- **X-axis:** Pathogens
- **Legend:**
  - Expected Copy Number
  - qPCR Copy Number
  - NGS Copy Number

### Notes:
- **Targeted (PCR) Based Detection and NGS-Based Metagenomic Detection**
- **MATERIAL MEASUREMENT LABORATORY**
RM Characterization

Each tube will contain a single genome and will be characterized for the following metrics:

- Genome Assembly – Long Read Sequencing
- Base Level Purity (i.e. Rare Variants)
  - Deep Illumina Sequencing
- Genomic Contaminants
  - Environmental/Reagent/Platform Contaminants
- Quantity (concentration in ng/µl)
  - Infer genomes per µl
- DNA Stability
AVAILABLE SUMMER 2017

Strains 1-4 - 1000 tubes of each

“We’re going to need a bigger boat”

Jason Krajl
WHAT WILL THE NIST MIXED MICROBIAL DNA REFERENCE MATERIAL LOOK LIKE?

2017
Panel – 26 Tubes

Expansion (2018)

Long Synthetic DNA Constructs Representing Clinically-Relevant Viral Pathogens.
Possibly RNA Virus
ORGANISM-LEVEL PURITY ANALYSIS OF “PURE” SALMONELLA GENOMIC DNA

<table>
<thead>
<tr>
<th>Name</th>
<th>Strain</th>
<th>Biosample</th>
<th>Size</th>
<th>%GC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MG001</td>
<td>Salmonella enterica LT2</td>
<td>SAMN02854572</td>
<td>4.8 Mb</td>
<td>52</td>
</tr>
</tbody>
</table>

Analysis done via Pathoscope

Credit: Nate Olson
A PROTOTYPE MIXTURE

<table>
<thead>
<tr>
<th>Name</th>
<th>Strain</th>
<th>Biosample</th>
<th>Size</th>
<th>%GC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MG001</td>
<td>Salmonella enterica LT2</td>
<td>SAMN02854572</td>
<td>4.8 Mb</td>
<td>52</td>
</tr>
<tr>
<td>MG002</td>
<td>Staphylococcus aureus</td>
<td>SAMN02854573</td>
<td>2.8 Mb</td>
<td>33</td>
</tr>
<tr>
<td>MG003</td>
<td>Pseudomonas aeruginosa</td>
<td>SAMN02854574</td>
<td>6.3 Mb</td>
<td>67</td>
</tr>
<tr>
<td>MG004</td>
<td>Clostridium sporogenes</td>
<td>SAMN02854575</td>
<td>4.1 Mb</td>
<td>28</td>
</tr>
<tr>
<td>HG001</td>
<td>Human</td>
<td>N/A</td>
<td>6000 Mb</td>
<td>50</td>
</tr>
</tbody>
</table>

Mixture #1

**Eqigenomic Mixture**

- Relative Genomes
- Nanograms of DNA

Mixture #2

**Log-Dilution Mixture**

- Relative Genomes
- Nanograms of DNA
RESULTS: EQUIGENOMIC MIXTURE

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>15273662</td>
<td>14949327</td>
<td>34801</td>
<td>14041</td>
<td>8501</td>
<td>5457</td>
<td>3778</td>
<td>326</td>
<td>152</td>
<td>132</td>
<td>12</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>14</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. enterica</td>
<td>1000</td>
<td>100</td>
<td>33699</td>
<td>10473</td>
<td>7920</td>
<td>3578</td>
<td>734</td>
<td>501</td>
<td>269</td>
<td>6</td>
<td>26</td>
<td>34</td>
<td>33</td>
<td>64</td>
<td>64</td>
<td>1</td>
<td>22</td>
<td>2</td>
<td>19</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>8</td>
<td>12</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Homo sapiens</th>
<th>Human herpesvirus 4</th>
<th>Salmonella enterica</th>
<th>Pseudomonas aeruginosa</th>
<th>Staphylococcus aureus</th>
<th>Clostridium botulinum</th>
<th>Halorubrum halophilum</th>
<th>Streptococcus pneumoniae</th>
<th>Penicillium verrucosum</th>
<th>Weissella helenica</th>
<th>Mrakia biliolopis</th>
<th>Diaporthe longicolla</th>
<th>Leuconostoc pseudomesenteroides</th>
<th>Staphylococcus epidermidis</th>
<th>Pseudomonas stutzeri</th>
<th>Aspergillus ochraceoroseus</th>
<th>Pseudomonas fluorescens</th>
<th>Epichloe bromicola</th>
<th>Ophiostoma piceae</th>
<th>Escherichia coli</th>
<th>Basidioascus undulatus</th>
<th>Diplodia pinea</th>
<th>Bipolaris victoriae</th>
<th>Syncyphalastrium racemosum</th>
<th>Grosmmannia clavigera</th>
<th>Pseudomonas putida</th>
<th>Thielaviopsis punctulata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>15273662</td>
<td>14949327</td>
<td>34801</td>
<td>14041</td>
<td>8501</td>
<td>5457</td>
<td>3778</td>
<td>326</td>
<td>152</td>
<td>132</td>
<td>12</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>14</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>S. enterica</td>
<td>1000</td>
<td>100</td>
<td>33699</td>
<td>10473</td>
<td>7920</td>
<td>3578</td>
<td>734</td>
<td>501</td>
<td>269</td>
<td>6</td>
<td>26</td>
<td>34</td>
<td>33</td>
<td>64</td>
<td>64</td>
<td>1</td>
<td>22</td>
<td>2</td>
<td>19</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>1</td>
</tr>
</tbody>
</table>
RESULTS: LOG-DILUTION MIXTURE

![Graph showing dilution species and genus](image)

**SPECIES**

<table>
<thead>
<tr>
<th>Tool 1</th>
<th>Kraken</th>
<th>MetaPHAn</th>
<th>Cosmos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>S. enterica</td>
<td>S. aureus</td>
<td>P. aeruginosa</td>
</tr>
</tbody>
</table>

**GENUS**

<table>
<thead>
<tr>
<th>Homo sapiens</th>
<th>R1</th>
<th>R2</th>
<th>GM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human sapiens</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>R1</th>
<th>R2</th>
<th>GM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspergillus ochraceus</td>
<td>12</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>Micrococcus luteus</td>
<td>15</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Staphylococcus aureus</td>
<td>14</td>
<td>17</td>
<td>10</td>
</tr>
<tr>
<td>Bacillus megaterium</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>R1</th>
<th>R2</th>
<th>GM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo sapiens</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

**Note:** The above table shows a summary of species and their dilution levels in a log-dilution mixture. The species and their corresponding R values indicate their presence and concentration levels in the mixture.
• In Fall 2015, there were ~50 metagenomic data analysis tools

• Spring of 2017, there are ~70 metagenomic data analysis tools

• Reference Database is NOT Standardized: Tool Specific

• The problem is getting worse, not better

• Compare Different Software Tools: “Profiling the Profilers”
  • Problem is there is no single metric to assign “correct” or “best”
  • Tools typically trade-off strengths and weaknesses
An evaluation of the accuracy and speed of metagenome analysis tools

Stinus Lindgreen1,3,4, Karen L. Adair3,4 & Paul P. Gardner1,4

Scientific Reports 6:19233 | DOI: 10.1038/srep19233

Tools typically trade-off strengths and weaknesses

<table>
<thead>
<tr>
<th>Method</th>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>FN</th>
<th>SEN</th>
<th>SPEC</th>
<th>PPV</th>
<th>NPV</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARK</td>
<td>23571720</td>
<td>170750</td>
<td>4718915</td>
<td>0</td>
<td>1.0000</td>
<td>0.8012</td>
<td>0.9527</td>
<td>1.0000</td>
<td>0.8736</td>
</tr>
<tr>
<td>EBI</td>
<td>13879</td>
<td>9939</td>
<td>5782564</td>
<td>2364153</td>
<td>0.0006</td>
<td>0.9983</td>
<td>0.5826</td>
<td>0.1964</td>
<td>-0.0157</td>
</tr>
<tr>
<td>Genometa</td>
<td>11733272</td>
<td>99524</td>
<td>5782564</td>
<td>1184075</td>
<td>0.4968</td>
<td>0.9831</td>
<td>0.9917</td>
<td>0.3280</td>
<td>0.3926</td>
</tr>
<tr>
<td>GOTTCHA</td>
<td>12756512</td>
<td>0</td>
<td>5782564</td>
<td>10921450</td>
<td>0.5588</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.3462</td>
<td>0.4327</td>
</tr>
<tr>
<td>Kzralen</td>
<td>21305228</td>
<td>86</td>
<td>5782545</td>
<td>2372576</td>
<td>0.8994</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.7691</td>
<td>0.7991</td>
</tr>
<tr>
<td>LMAT</td>
<td>11516886</td>
<td>1592274</td>
<td>4295866</td>
<td>8035328</td>
<td>0.6435</td>
<td>0.7396</td>
<td>0.9050</td>
<td>0.3382</td>
<td>0.3025</td>
</tr>
<tr>
<td>MEGAN</td>
<td>12068515</td>
<td>63500</td>
<td>5782564</td>
<td>10749597</td>
<td>0.5452</td>
<td>0.9891</td>
<td>0.9951</td>
<td>0.3499</td>
<td>0.4365</td>
</tr>
<tr>
<td>MetaPhlan</td>
<td>15073488</td>
<td>0</td>
<td>5782564</td>
<td>22170624</td>
<td>0.0036</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.2069</td>
<td>0.1150</td>
</tr>
<tr>
<td>MetaPhyler</td>
<td>153386</td>
<td>713</td>
<td>5781915</td>
<td>23544072</td>
<td>0.0057</td>
<td>0.9999</td>
<td>0.9947</td>
<td>0.1972</td>
<td>0.0327</td>
</tr>
<tr>
<td>MG–FAST</td>
<td>16254882</td>
<td>44269</td>
<td>5782562</td>
<td>7078762</td>
<td>0.7012</td>
<td>0.9924</td>
<td>0.9973</td>
<td>0.4486</td>
<td>0.5605</td>
</tr>
<tr>
<td>mOTU</td>
<td>47846</td>
<td>0</td>
<td>5782564</td>
<td>23630126</td>
<td>0.0020</td>
<td>1.0000</td>
<td>1.0000</td>
<td>0.1966</td>
<td>0.0200</td>
</tr>
<tr>
<td>OneCodex</td>
<td>21808925</td>
<td>320</td>
<td>5782541</td>
<td>1808749</td>
<td>0.9210</td>
<td>0.9999</td>
<td>1.0000</td>
<td>0.7558</td>
<td>0.8345</td>
</tr>
<tr>
<td>qiIME</td>
<td>12914</td>
<td>37</td>
<td>5782564</td>
<td>23662021</td>
<td>0.0005</td>
<td>1.0000</td>
<td>0.9972</td>
<td>0.1964</td>
<td>0.0102</td>
</tr>
<tr>
<td>Taxator-tk</td>
<td>16010500</td>
<td>1898276</td>
<td>5782562</td>
<td>10169197</td>
<td>0.5335</td>
<td>0.7537</td>
<td>0.8593</td>
<td>0.3625</td>
<td>0.2537</td>
</tr>
</tbody>
</table>

Table 2. Phylum level performance metrics for the individual methods. Average numbers for the simulated data sets are given. The metrics are true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) as well as sensitivity (SEN), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV) and Matthew’s Correlation Coefficient (MCC).

B) Genus level

Performance

Tools typically trade-off strengths and weaknesses
METAGENOMIC PATHOGEN DETECTION: CONCLUSIONS

• Simple Mixtures Aren’t So Simple

• We Need “Ground Truth” to Assess the Analytical Sensitivity and Specificity of NGS Metagenomics

• What is “Real”? What is artifact?

• Metagenomic Analysis Software:
  • Produce \textit{in silico} Data Set that is “Ground Truth”
  • Host Analysis “Challenge”
  • Implement A Standard Reference Database?
THE METAGENOMICS MVP CHALLENGE v1.0

Presented by

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

When using metagenomic methods to assess the content of a complex microbial sample, there are many steps in the measurement process where bias might be introduced; including, but not limited to: sample collection and storage, DNA extraction method, library construction and/or PCR, NGS platform, raw data filtering and data analysis/interpretation.

Investigating and understanding the source and extent of measurement bias is often accomplished through the development and utilization of a shared reference material ("ground truth") followed by repeated measurements of the material in multiple independent laboratories.

Welcome to The NIST-ABRF-CosmosID Metagenomic MVP Challenge v1.0

http://www.cosmosid.com/nist-challenge/
NIST-FDA-DHS Standards for Pathogen Detection for Biosurveillance and Clinical Applications Workshop

NIST-DHS Workshop: Standards for Pathogen Detection for Biosurveillance and Clinical Applications

August 14-15, 2017 at NIST, Gaithersburg, Md.

#NISTPathogen
"Uncover, identify, and find solutions to the challenges and hurdles to the commercial success and clinical adoption of microbiome-based services and products"

Regulatory Engagement

"Engage the FDA and other relevant regulatory agencies to establish dialogue and find ways to collaborate and guide regulatory science and development pertaining to the microbiome"

Founding Members

AO Biome
Larry Weiss, MD, Chief Medical Office

Abbott Nutrition
JoMay Chow, Ph.D., Senior Research Scientist

CosmosID
Manoj Dadlani, CEO

Diversigen
Joseph F. Petrosino, Ph.D., Founder and Chief Scientific Officer

Mayo Clinic
Purna Kashyap, MBBS, Assistant Professor at Mayo Clinic
Kelly Krajnik, Business Development Manager

Second Genome
Mohan S. Iyer, Chief Business Officer

Whole Biome
Colleen Cutcliffe, Ph.D., CEO and Co-Founder
The International Metagenomics and Microbiome Standards Alliance (IMMSA)

WebEx on 2nd Thursday of Every Month from 2pm till 4pm EST
NIST-NRC POSTDOCTORAL POSITIONS AVAILABLE

National Research Council (NRC)
- US Citizens ONLY
- Within 5 years of PhD
- 2 years @ $67,588/y

Inquire for details:
Scott Jackson  scott.jackson@nist.gov
Jason Kralj  jason.kralj@nist.gov

#NISTPathogen
AKNOWLEDGEMENTS

Sam Forry
Jason Kralj
Nate Olson

CosmosID
Nur Hasan
Manoj Dadlani

FDA
Heike Sichtig
Brittany Goldberg

MRI Global
Jonathan Jacobs

#NISTpathogen
#NISTmicrobiome