Tonya Ward, PhD
Postdoctoral Associate
Biotechnology Institute
University of Minnesota
Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies

Daryl M Gohl\textsuperscript{1}, Pajau Vangay\textsuperscript{2}, John Garbe\textsuperscript{3}, Allison MacLean\textsuperscript{1}, Adam Hauge\textsuperscript{1,9}, Aaron Becker\textsuperscript{1}, Trevor J Gould\textsuperscript{4}, Jonathan B Clayton\textsuperscript{5}, Timothy J Johnson\textsuperscript{5}, Ryan Hunter\textsuperscript{6}, Dan Knights\textsuperscript{7,8} & Kenneth B Beckman\textsuperscript{1}
Knights Lab

- Computer Science & Engineering
- Biotechnology Institute
- Computational microbiology lab
- Develop methods and tools that bring precision medicine to the microbiome
- Find patterns in microbial communities

http://www.knightslab.org/
The Microbiome

- The microorganisms in a particular environment
- Bacteria, fungi, viruses
- External environment, host-associated
The Microbiome

• Bacteria
  • ~39 trillion bacterial cells per person
  • ~30 trillion human cells

Microbiome Sequencing

Amplicon Sequencing

Amplify 16S rDNA
20,000 sequences/sample

Sample1
>ATGCATGCTAGCTA...
>GTAGCTAGCTATGA...
>CGATCGATTATTAC...
>CGATCGTAGGCTAG...
>CGATCGTATCGGAG...

Shotgun Metagenomic Sequencing
40,000,000 sequences/sample
Microbiome Characterization

16S Amplicon Sequencing

- Who’s there
  - QIIME
  - mothur
  - NINJA, ...

Shotgun Metagenomic Sequencing

- Who’s there
  - MG-RAST
  - MEGAN
  - MetaphlAn, ...

Reference Database of all sequenced bacteria

Sample1

>ATGCATGCTAGCTA...
>GTAGCTAGCTATGA...
.CGATCGATTATTAC...
>CGATCGTAGGCTAG...
>CGATCGTATCGGAG...

Sample1

>Bacteria_A
>Bacteria_B
>Bacteria_C
>Bacteria_A
>Bacteria_A
Microbiome Characterization

16S Amplicon Sequencing

• Who’s there
  • QIIME
  • mothur
  • NINJA, ...

• Predicted functionality
  • PICRUSt, ...

Shotgun Metagenomic Sequencing

• Who’s there
  • MG-RAST
  • MEGAN
  • MetaphlAn, ...

• Functionality
  • HUMAnN
  • MG-RAST, ...
The Treatment-Naive Microbiome in New-Onset Crohn’s Disease

Specific, but informative?

- K03074 preprotein translocase subunit SecF
- K03072 preprotein translocase subunit SecD
- K01916 NAD+ synthase [EC:6.3.1.5]
- K01879 glycyl-tRNA synthetase beta chain [EC:6.1.1.14]
- K09824 hypothetical protein
- K06193 phosphonoacetate hydrolase [EC:3.11.1.2]
- K02843 heptosyltransferase II [EC:2.4.-.-]
- K03274 ADP-L-glycero-D-manno-heptose 6-epimerase [EC:5.1.3.20]
- K03272 D-beta-D-heptose 7-phosphate kinase [EC:2.7.1.- 2.7.7.-]
KEGG Pathways

- Manually drawn pathway maps
  - 6 general categories:
    - Metabolism
    - Environmental Information Processing
    - Organismal Systems
    - Genetic Information Processing
    - Cellular Processes
    - Human Diseases

**KEGG Orthology (KO) of preprotein translocation subunit SecF**

- Genetic Information Processing
  - Folding, sorting and degradation
    - Protein export
      - secF; preprotein translocase subunit SecF

- KEGG harbors 18,669 orthology groups

How can we pull the most meaningful information out of our data?
BugBase

• Analysis tool for:
  • 16S Amplicon
  • Shotgun Metagenomic Sequencing

• Goals:
  • Predict organism-level microbiome phenotypes
  • Host-relevant traits
  • User-friendly interface
  • Publication quality outputs
Welcome to BugBase.

About Bugbase
BugBase is a tool for measuring high-level phenotypes in your microbiome.

Parse Data
Upload your microbiome data and run it through our pipeline.

Downloads
Download BugBase and see sample analyses.

Visit the documentation to get started!
<table>
<thead>
<tr>
<th>Microbiome Traits of Interest</th>
<th>Organism-level Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Gram Staining</td>
<td>• Pre-existing Tools:</td>
</tr>
<tr>
<td>• Biofilm Formation</td>
<td>• Gene content for the metagenome</td>
</tr>
<tr>
<td>• Pathogenicity</td>
<td>• “Bag of genes” approach</td>
</tr>
<tr>
<td>• Oxygen Tolerance</td>
<td>• Our approach</td>
</tr>
<tr>
<td>• Stress Tolerance</td>
<td>• Gene content per microbe</td>
</tr>
</tbody>
</table>
Biofilm Formation:

Gene A  Gene B  Gene C  Gene D

Bag of Genes Approach

33% of genes are related to biofilm formation or 2 copies of pathway is present

Organism-Level Approach

33% of the microbiome can form a biofilm

Other Genes:

Genes

Microbiome

Gene A

Gene B

Gene C

Gene D
Biofilm Formation:

Gene A  Gene B  Gene C  Gene D

Microbiome

Bag of Genes Approach

Organism-Level Approach

Other Genes:

Bag of Genes Approach

67% of genes are related to biofilm formation
or
4 copies of pathway are present

Organism-Level Approach

33% of the microbiome can form a biofilm

x 0.33

✓  x 0.33

✘  x 0.33
Biofilm Formation:

Gene A  Gene B  Gene C  Gene D

Threshold:
Percentage of genes in the category needed to possess the trait

Threshold (% of category covered)

Relative Abundance

Mouth
Stool

Threshold

Relative Abundance

0.0  0.2  0.4  0.6  0.8  1.0

0  0.25  0.5  0.75  1

Threshold
Human Microbiome Project

Human Microbiome Project

Microbiome phenotypes vary by body site
Human Gut Microbiome Viewed Across Age and Geography

Human Gut Microbiome Viewed Across Age and Geography

Human Gut Microbiome Viewed Across Age and Geography
Human Vaginal Microbiome and Bacterial Vaginosis

Ravel J, et al. 2011. PNAS.
Human Vaginal Microbiome and Bacterial Vaginosis

Ravel J, et al. 2011. PNAS.
Human Vaginal Microbiome and Bacterial Vaginosis

Table 2. Wet Mount - Nugent’s Scoring

<table>
<thead>
<tr>
<th>Normal</th>
<th>Abnormal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nugent Score = 0</td>
<td>Nugent Score = 10</td>
</tr>
</tbody>
</table>

- Lactobacilli common
- Few to no other organisms
- Polymicrobial preparations
- Clue cells
- Coccobacilli, Gram variable
- Lactobacilli few or absent

Ravel J, et al. 2011. PNAS.
Yellowstone Hot Springs

Crohn’s Disease

Microbiome phenotypes vary by diagnosis
Additional Features

• BugBase can be used for organism-level predictions of *any* pathway
Increased sensitivity with organism-level predictions

Starch Metabolism
Increased sensitivity with organism-level predictions

Starch Metabolism
Increased sensitivity with organism-level predictions

470 pathways

E

Stool vs Tongue

Sub– vs Supra–gingival Plaque

mean p-value

Number of Samples

5 10 50 100 250

5 10 50 100 250

Bag of genes  Organism level
BugBase

- Analysis tool for amplicon and shotgun metagenomic datasets
- Organism-level approach for prediction of microbiome traits
- Focuses on high level traits, as well as all known KEGG pathways or custom pathways
- Supplements pre-existing tools and pipelines
- User-friendly web application

https://bugbase.cs.umn.edu
Thank You

• Collaborators
  • Jake Larson, UMN
  • Ben Hillmann, UMN
  • Jeremy Meulemans, UMN
  • Dimitri Sidiropoulus, UMN
  • Joshua Lynch, UMN
  • Ran Blekhman, UMN
  • Ryan Finch, St Cloud State University
  • Rob Knight, University of California at San Diego
  • Greg Caporaso, Northern Arizona University
  • John Spear, Colorado School of Mines

• Knights Lab
  • Dimitri Sidiropoulus
  • Ben Hillmann
  • Emmanuel Montassier
  • Gabe Al Ghalith
  • PJ Vangay
  • Abby Cole
  • Robin Cutler
Validation

Cross-validation of BugBase phenotype predictions using a ten-fold jackknife.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Dataset Size (OTUs)</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gram Positive</td>
<td>2,580</td>
<td>0.9686</td>
</tr>
<tr>
<td>Gram Negative</td>
<td>2,580</td>
<td>0.9694</td>
</tr>
<tr>
<td>Aerobic</td>
<td>2,587</td>
<td>0.8627</td>
</tr>
<tr>
<td>Anaerobic</td>
<td>2,587</td>
<td>0.9234</td>
</tr>
<tr>
<td>Facultative Anaerobic</td>
<td>2,587</td>
<td>0.8828</td>
</tr>
<tr>
<td>Contains Mobile Elements</td>
<td>1,335</td>
<td>0.951</td>
</tr>
<tr>
<td>Pathogenic</td>
<td>1,598</td>
<td>0.9767</td>
</tr>
<tr>
<td>Oxidative Stress Tolerant</td>
<td>1,039</td>
<td>0.9664</td>
</tr>
</tbody>
</table>
PICRUSt

Reference tree and Gene content table

Prune unknown gene-content tips

Infer ancestral gene content

Infer contemporary gene content

Gene Content
- Known
- Unknown

Persistent gut microbiota immaturity in malnourished Bangladeshi children

Persistent gut microbiota immaturity in malnourished Bangladeshi children

Persistent gut microbiota immaturity in malnourished Bangladeshi children