Screening biological products for antimicrobial resistant genes using the HIVE AMR pipeline

Sean Smith, Ph.D.
Office of Biostatistics and Pharmacovigilance
Center for Biologics Evaluation and Research

Disclaimer: The information in this presentation represents the opinions of the speaker and does not necessarily represent FDA’s position or policy.
Introduction

Definition

Bacterial antimicrobial resistance (AMR) occurs when changes in bacteria increase resistance of the bacteria to drug treatments.
Introduction, cont.

- **World-wide impact**
  - AMR estimates for year 2019
    - 4.95 million AMR-associated deaths\(^1\)
    - 1.27 million AMR-attributable deaths\(^1\)
  - Predicted estimates for 2050
    - 10 million AMR-associated deaths annually (more than currently die from cancer)\(^2\)

---
Objectives

• FDA Review
  – Biologic therapeutics (bacteria, phages) reviewed at FDA
    • Screen for safety issues, antibiotic resistance genes

• Application of technology / scientific computing
  – Developed antimicrobial resistance pipeline through the High-performance Integrated Virtual Environment (HIVE)
  – Accepts high-throughput sequencing (HTS) reads as input and provides a detailed report on potential antibiotic resistant (ABR) genes present in the biologic product
### AMR Pipeline

Designed, optimized for different high-throughput sequencing technologies:

- **Short reads** (dark blue)
  - Illumina, Ion Torrent
  - Implemented in HIVE

- **Long reads**
  - PacBio, Oxford Nanopore

- **Hybrid**
  - Both

---

**Second Generation**  
Massively Parallel Sequencing

- Sequencing by synthesis
- Amplified templates are generated during sequencing, reducing the requirements for starting material
- High accuracy
- Short read lengths
  - e.g., MiSeq (Illumina), Ion Torrent (Thermo Fisher Scientific)

**Third Generation**  
Single-molecule Sequencing

- Single-molecule templates
- Low accuracy
- Long read lengths
  - e.g., Single-Molecule Real-Time (SMRT) — Sequencing (Pacific Biosciences), MinION (Oxford Nanopore Technologies)

---

Ronholm, 2016, Clin Microbiol Rev
AMR Pipeline

Quality control and trimming of reads

https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
AMR Pipeline

Assemble reads into contigs

1. Fragment DNA and sequence
2. Find overlaps between reads
3. Assemble overlaps into contigs

AMR Pipeline

Align contigs to CARD database

Contig: ATGTCACAT...
AMR Gene: TCCATTAGTGTCA...
Python script (developed internally) to filter/merge/format results for output to user
HIVE Platform

- A cloud-based environment that comprises both a storage library of data and a powerful computing capacity.

- Can consume, digest, analyze, manage, and share all this data.
**AMR Short Reads Pipeline**

This pipeline consists of 14 chained processes. The most notable programs within this pipeline are SPAM, BLAST, BLASTX, RGI BLAST, RGI DIAMOND, and Hexagon. Output from these programs are then processed by a Python script to return the desired outcome.

The inputs for this pipeline are the reads and the CARD database, in both java and fasta formats. For the database parameters, please go to the Offsite CARD Database and download the "CARD DTA" of the desired version. Under the Files, upload the "nucleotide_fasta_protein_hormology_modelfasta" and "proteins_fasta_protein_hormology_modelfasta" and "CARD DTA" respectively.

Only short-read sequences for which Spades is recommended are Illumina and SFF format. NGS data with Hexamer length less than 37 is not supported.

**Output**

**AMR Merge**

Available Files to Download

- **enterococcus_faecalis_shortread**
  - **Step 14**

**Parameters**

- **Input Reads**
  - Read 1
  - Read 2

**Results**

- **Available Files to Download**
  - _all files_
  - armr-merge-output.txt
  - armr-merge-output.txt.compressed.txt

**File Details**

- **armr-merge-output.txt**: 23.47 KB
- **armr-merge-output.txt.compressed.txt**: 9.42 KB
<table>
<thead>
<tr>
<th>ARO</th>
<th>Gene</th>
<th>Protein Product</th>
<th>Protein Accession</th>
<th>Drug Class</th>
<th>Resistance Mechanism</th>
<th>Tool</th>
<th>Length</th>
<th>Pident</th>
<th>Bitscore</th>
<th>Number of Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>3000839</td>
<td>arlS</td>
<td>conserved hypothetical protein</td>
<td>ABD30512.1</td>
<td>acridine dye; disinfecting agents and intercalating dyes; fluoroquinolone antibiotic</td>
<td>antibiotic efflux</td>
<td>Blastx</td>
<td>451</td>
<td>100</td>
<td>812</td>
<td>5</td>
</tr>
<tr>
<td>3000026</td>
<td>mepA</td>
<td>MepA</td>
<td>AAU95768.1</td>
<td>glycolcycline; tetracycline antibiotic</td>
<td>antibiotic efflux</td>
<td>Blastx</td>
<td>451</td>
<td>100</td>
<td>885</td>
<td>5</td>
</tr>
<tr>
<td>3004661</td>
<td>fosB</td>
<td>fosfomycin resistance protein</td>
<td>EHS19134.1</td>
<td>fosfomycin</td>
<td>antibiotic inactivation</td>
<td>Hexagon</td>
<td>420</td>
<td>97.38</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>3005097</td>
<td>blaZ</td>
<td>beta lactamase protein BlaZ</td>
<td>CBZ41939</td>
<td>penam</td>
<td>antibiotic inactivation</td>
<td>Blast</td>
<td>405</td>
<td>99.01</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>3000617</td>
<td>mecA</td>
<td>MecA</td>
<td>AGC51118.1</td>
<td>carbapenem; cephalosporin; cephamycin n; monobactam; penam</td>
<td>antibiotic target replacement</td>
<td>Hexagon</td>
<td>2007</td>
<td>99.85</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>3000250</td>
<td>ErmC</td>
<td>macrolide-lincosamide-streptogramin B-resistance protein</td>
<td>AAA98296.1</td>
<td>lincomycin antibiotic; macrolide antibiotic; streptogramin antibiotic</td>
<td>antibiotic target alteration</td>
<td>Hexagon</td>
<td>735</td>
<td>99.86</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>3004572</td>
<td>LmrS</td>
<td>drug resistance transporter, EmrB/QacA subfamily</td>
<td>AAW38464.1</td>
<td>aminoglycoside antibiotic; diaminopyrimidine antibiotic; macrolide antibiotic; oxazolidinone antibiotic; phenicol antibiotic</td>
<td>antibiotic efflux</td>
<td>Blast</td>
<td>1443</td>
<td>99.86</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>3004667</td>
<td>norA</td>
<td>norA</td>
<td>BAA14147.1</td>
<td>fluoroquinolone antibiotic</td>
<td>antibiotic efflux</td>
<td>Hexagon</td>
<td>1167</td>
<td>99.91</td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>
# Output (Enterococcus faecalis, Blautia obeum)

## Enterococcus faecalis

<table>
<thead>
<tr>
<th>ARO</th>
<th>Gene</th>
<th>protein Product</th>
<th>Protein Accession</th>
<th>Drug Class</th>
<th>Resistance Mechanism</th>
<th>Tool</th>
<th>Length</th>
<th>Pident</th>
<th>Bitscore</th>
<th>Number of Tools Identifying Potential AMR Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>3002875</td>
<td>dfrE</td>
<td></td>
<td>EOD99669.1</td>
<td>diaminopyrimidine antibiotic</td>
<td>antibiotic target replacement</td>
<td>Blastx</td>
<td>164</td>
<td>98.171</td>
<td>335</td>
<td>5</td>
</tr>
<tr>
<td>3003948</td>
<td>efrA</td>
<td></td>
<td>WP_104671188.1</td>
<td>fluoroquinolone antibiotic;macrolide antibiotic;rifamycin antibiotic</td>
<td>antibiotic efflux</td>
<td>RGI-Blast</td>
<td>99.83</td>
<td>1168.3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>3003551</td>
<td>emeA</td>
<td>multidrug efflux pump</td>
<td>BAC11911.1</td>
<td>acridine dye;disinfecting agents and intercalating dyes</td>
<td>antibiotic efflux</td>
<td>Blast</td>
<td>736</td>
<td>99.46</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>3000300</td>
<td>isaA</td>
<td>Lsa</td>
<td>AAO43110.1</td>
<td>lincomamide antibiotic;macrolide antibiotic;oxazolidinone antibiotic;phenicol antibiotic;pleuromutilin antibiotic;streptogramin antibiotic;tetracycline antibiotic</td>
<td>antibiotic target protection</td>
<td>Blastx</td>
<td>498</td>
<td>98.193</td>
<td>954</td>
<td>5</td>
</tr>
<tr>
<td>3000186</td>
<td>tetM</td>
<td>Tet(M)</td>
<td>BAB82500.1</td>
<td>tetracycline antibiotic</td>
<td>antibiotic target protection</td>
<td>Hexagon</td>
<td>1920</td>
<td>96.77</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

## Blautia obeum

<table>
<thead>
<tr>
<th>ARO</th>
<th>Gene</th>
<th>protein Product</th>
<th>Protein Accession</th>
<th>Drug Class</th>
<th>Resistance Mechanism</th>
<th>Tool</th>
<th>Length</th>
<th>Pident</th>
<th>Bitscore</th>
<th>Number of Tools Identifying Potential AMR Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

- Antibiotic resistance is a serious global health concern
- **AMR pipeline** developed and optimized, hosted on the HIVE platform, and available to registered users
- Pipeline identified antibiotic resistant genes in agreement with current knowledge of the bacterial species analyzed
- Pipeline is ready to be utilized by FDA reviewers
  - Applies technology and scientific computing to provide **fast, efficient, consistent,** and **unbiased** results to validate that products are free from antibacterial resistant genes
  - User friendly, bioinformatics knowledge not needed
Availability

- sciHIVE
  - scihive.fda.gov

- regHIVE
Acknowledgements

Paul Carlson Group (FDA)
Deepika Prasad

HIVE (FDA)
Grace Alterovitz
Arya Eskandarian
Tigran Ghazanchyan
Anton Golikov
Hannah Howell
Konstantinos Karagiannis
Luis Santana-Quintero
Sean Smith
Thank you!

Disclaimer: The information in this presentation represents the opinions of the speaker and does not necessarily represent FDA’s position or policy.