Predictive Web Server

Team 1
Jacob Feldman, Zainab Arshad, Patrick Howard, Hanying Pan, Shrinkhla Sharma
Genome Assembly
- De novo assembly of 50 isolates

Gene Prediction
- Predicted the genes

Functional Annotation
- Annotated the function of these genes

Comparative Genomics
- Identify the initial strain/outbreak isolates

Predictive Web Server
- Provide a platform for present the findings of the team
Objectives

- Create a predictive webserver that allows for automated breakout analyses
- The functionality of the webserver to include
  - distance from the closest isolate in the database
  - determine whether the uploaded isolate is an outbreak or sporadic strain
  - visualization of the distance between test isolate and database isolate as a phylogenetic tree and/or heatmap
  - virulence factor and antimicrobial resistance profiling of your isolates
- Design Goals
  - Mobile friendly
  - Easy to use
  - Minimal
Webserver Structure
Why PHP?

● 80% of the web is powered by PHP e.g. Wikipedia, Yahoo, Slack...
● Many useful libraries
● Easy to integrate with a database
● Extensive online documentation
● Numerous forums and blogs for guidance and troubleshooting for developers
Why Laravel?

- Laravel is a free, open-source PHP web framework
- Intended for the development of web applications following the model–view–controller (MVC) architecture
- Well-structured documentation
- Offers many utilities that aid in application deployment and maintenance
- Allow rapid application development
- Current stable version (Laravel 5.8) was released on February 26, 2019
**Genome Assembly**

**Input**
- Fetch fastq files using SRA accession number or browsing from client’s device.
- Check for the validity of the input fastq file
- Choice of trimming
- Choice of assembly tool
- Kmer size - optional field
- User’s email and output filename

**Output**
- Downloadable links of contigs.fasta, scaffolds.fasta and quast report to user’s email

**Visualization of quast report**
Gene Prediction

**Input**

- Contigs file produced from genome assembly
- Choice of Tools
- User’s email and output filename
- Choice of continuation to annotation and comparative genomics

**Output**

Downloadable links of General Feature Formal (.gff) files, FASTA Amino Acid (.faa) and Nucleic Acid (.fna) files
Functional Annotation

Input:

- FASTA Amino Acid (.faa) and Nucleic Acid (.fna) files generated from gene prediction

Output:

- General Feature Formal (.gff) files for each inputted genome
- Visualization of the number of annotations per genome

Links to downloadable .gff files

Visualized output results
Comparative Genomics

Input:

- General Feature Format (gff) file
- Annotated FASTA file for nucleotide & protein sequences (faa & fna)

Output:

- Label uploaded isolate as either sporadic or outbreak strain
- Phylogenetic tree and/or heatmap to visualize distance between uploaded isolate & database isolates
  - Include virulence factor & antimicrobial resistance information within the figure(s)
  - Distance of uploaded isolate from closest database isolate
Future Direction

- Allow user to begin and end pipeline at any of the 4 steps, uploading the correct input file format & receiving the appropriate output results
- Automate pipeline to progress through all 4 steps from start to end
  - Involves linking output of one step to input of the next
- Efficiently store files, remove unnecessary files
- Incorporate ability for user to pick which tool(s) they would like to use in each step
- Improve user interface & platform design
- Merge back end with front end
Demo