

Predictive Web Server

Team II: Sarthak Sharma, Shrey Mathur, Ragy Haddad, Ayush Semwal, Yusuph Mavura, Jacob Boswell, Pooja Khurana, Brian Merritt, Xinyu Wang, Yinquan Lu

Preliminary Results

The web server is up and running

[Home](#)

[Github](#)

[Wiki](#)

[Contact](#)

BIOPREDICT

Georgia Institute of Technology Klebsiella Predictive Webserver

Genome Assembly

De Novo Assembly using Skesa

[About this pipeline](#)

**Job
confirmation:**

Accession: **SRR345678**
E-mail: **sarthaksharma@gatech.edu**
ETA: 15 minutes

SUCCESS!!... e-mail with assembly will be sent to sarthaksharma@gatech.edu

Single SRA Assembly Confirmation

Navigation bar with icons for search, notifications, settings, and help, along with the user name "Sharma, Sarthak" and a profile picture.

Biopredict Job Done! [Assembled Reads]



ragy202@gmail.com

Tue 4/17/2018 9:55 AM

To: Sharma, Sarthak ^

Reply | v

To help protect your privacy, some content in this message has been blocked. To re-enable the blocked features, click here.

To always show content from this sender, click here.

Job confirmation SRR345678

Download Assembly:

<http://predict2018b.biosci.gatech.edu/assemble/user/SRR345678-sarthaksharma@gatech.edu-2018-04-1709-55-22-04:00-assembly.skesa.fa>

for support please contact rhaddad7@gatech.edu

Design Principles

Function over Form

- Minimal
 - Mobile Friendly
 - Short Load Time
 - Contrasting Colors
-

How does it look now?

Current Status

1. Assembly

- a. Input - SRA accession or FASTQ reads file
- b. Output - Assembled genome in FASTA file

2. Email confirmation

- a. An email is sent out to the user when the job is finished
- b. The email also contains the link to download the result

3. Downloads

- a. Allows the user to download results from the group analyses
- b. Currently, allows download of assemblies (FASTA) and annotation (GFF) files

Home Page - Basic Layout

Home

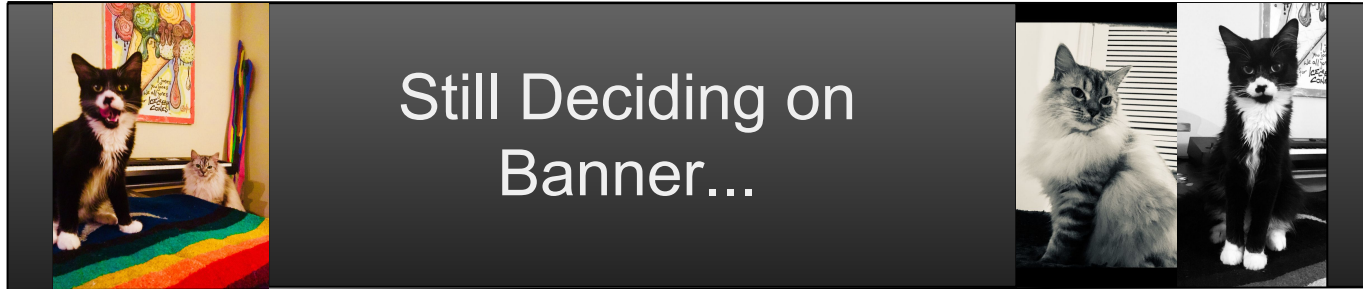
Github

Wiki

Contact

BIOPREDICT

Georgia Institute of Technology Klebsiella Predictive Webserver



ASSEMBLE

DOWNLOADS

PREDICT

Assembly Page - Batch mode

[Home](#)[Github](#)[Wiki](#)[Contact](#)

BIOPREDICT

Georgia Institute of Technology Klebsiella Predictive Webserver

Genome Assembly

De Novo Assembly using Skesa

[About this pipeline](#)

Input type: Select read types: *Single SRR Accession, a batch of SRR Accessions, or local files from your machine*

Enter a list of SRR accessions to assemble

Input format: *SRRXXX435,SRRXXX221,SRRXXX223...* (comma separated IDs)

Enter your email to be notified when your assembly is complete. ~8-10 min

Downloads Page

[Home](#)[Github](#)[Wiki](#)[Contact](#)

BIOPREDICT

Georgia Institute of Technology Klebsiella Predictive Webserver

Downloads

Klebsiella pneumoniae data dump

Accession	Assembly	Annotation
SRR3467249	FASTA	GFF
SRR3982229	FASTA	GFF
SRR3982230	FASTA	GFF
SRR3982253	FASTA	GFF
SRR3982316	FASTA	GFF
SRR3982346	FASTA	GFF
SRR3987112	FASTA	GFF
SRR3987115	FASTA	GFF
SRR3987118	FASTA	GFF
SRR3987119	FASTA	GFF
SRR3987120	FASTA	GFF
SRR3987121	FASTA	GFF
SRR3987122	FASTA	GFF

Suggestions

1. Predict Species and Strain from the given FASTQ reads
2. Display genes which have implication in antibiotic resistance
3. Show important virulence factors
4. Summary stats about the assemblies

How will it look tomorrow?

Under Construction

[Home](#)[Github](#)[Wiki](#)[Contact](#)

BIO-PREDICT

Georgia Institute of Technology Klebsiella Predictive Webserver

Predict

Bacterial Strain and Species prediction
Antibiotic Resistance Genes prediction
Virulence Factors prediction

[About this pipeline](#)

Input type: *Select Input Type: fasta or fastq*

No file chosen

Report:

Frequency	Strain
100.00000%	Klebsiella_oxytoca_strain_CAV1374

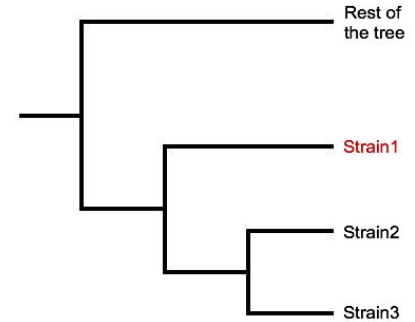
Under Construction

1. Prediction of Species and Strain using StrainSeeker
2. Summary stats about the assemblies
3. Assembling in “Batch Mode” (Under testing)
4. Display entities which play a role in Antibiotic Resistance
5. Display Virulence Factors

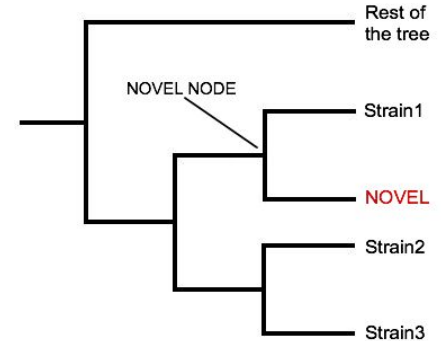
StrainSeeker

- Built-in Database
OR
Custom Database
 - Newick Guide Tree, .fna files for all OTUs
- Input: fastq or fasta sample
- Strains detected by k-mer pool similarity

Case 1

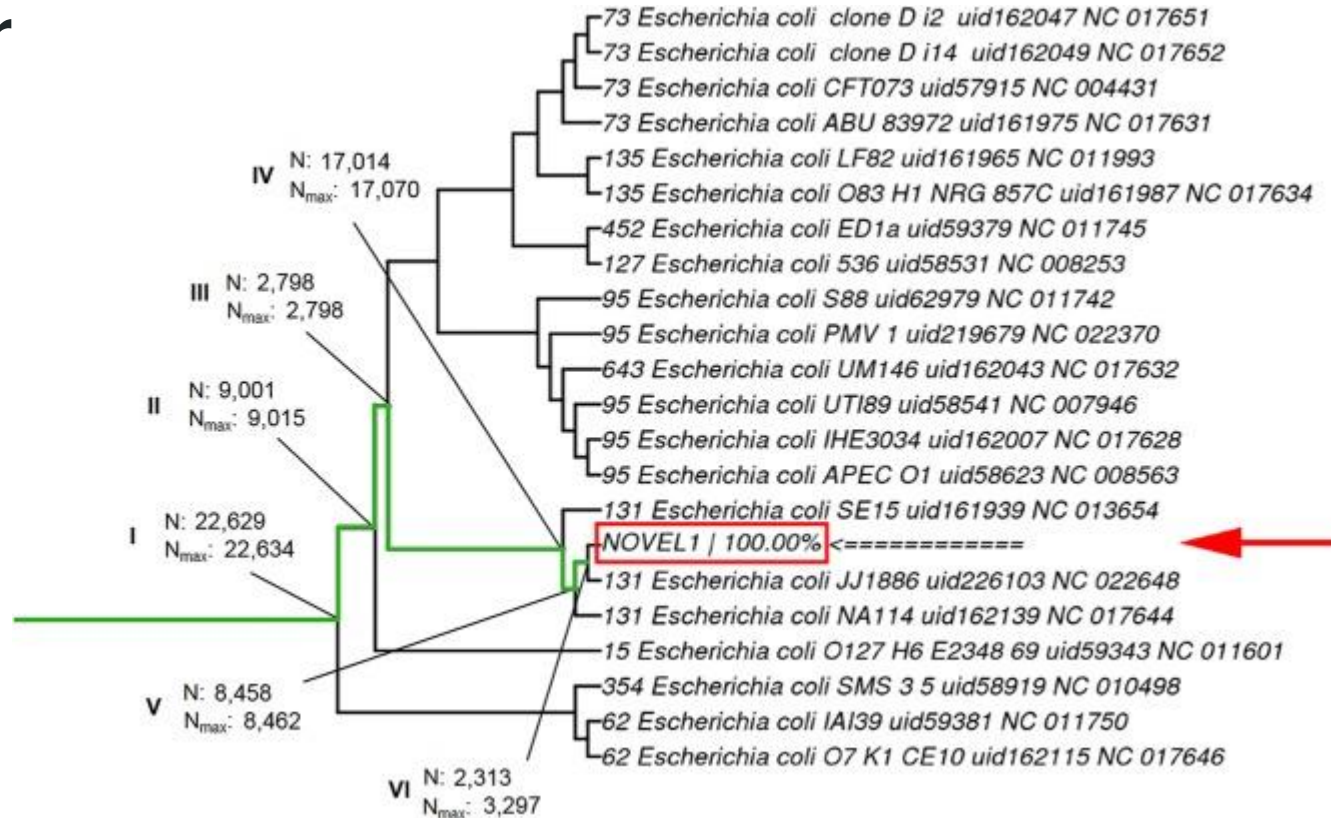


Case 2



StrainSeeker

-output tree

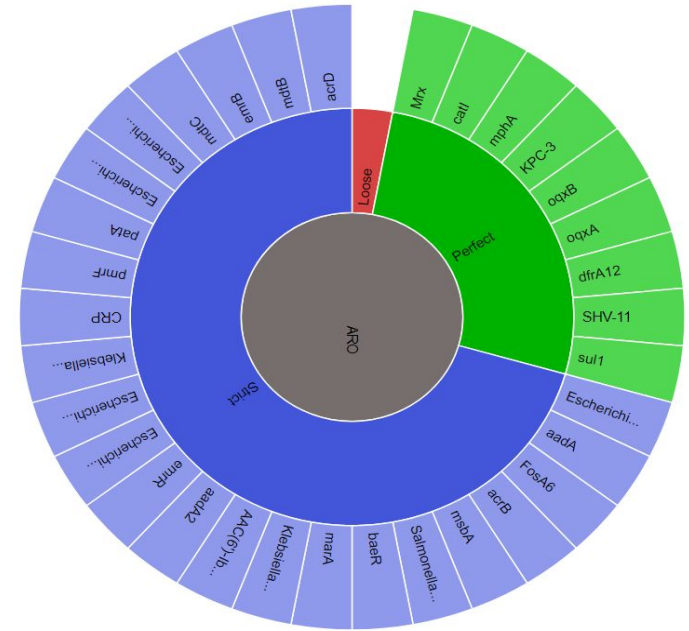


CARD

- The Comprehensive Antibiotic Resistance Database
- BLAST/DIAMOND against database of resistance genes, their products and associated phenotypes.
- Use of rgi command-line tool to generate JSON containing all antibiotic hits along with other key characteristics for each query
 - Drug class
 - Hit score & percent identity (query to hit)
 - Resistance mechanism
 - ARO name
 - AMR gene family

CARD

- D3 visualization tool (interactive)
- Perfect, Strict, Loose
- “Clicking” provides additional information for AMR gene



Best_Hit_ARO	AMR Gene Family	Drug Class	Model_type	Percentage Length of Reference Sequence	Resistance Mechanism
FosA6	fosfomycin thiol transferase	fosfomycin	protein homolog model	100.00	antibiotic inactivation
PmrF	pmr phosphoethanolamine transferase	peptide antibiotic	protein homolog model	100.00	antibiotic target alteration
baeR	resistance-nodulation-cell division (RND) antibiotic efflux pump	aminocoumarin antibiotic; aminoglycoside antibiotic	protein homolog model	100.00	antibiotic efflux

Virulence Factors

- blastn against virulence factor database (VFDB)
- Remove redundant matches, output list of unique VF homologues
 - -Accession numbers, blast scores, positions in query

How will it look next week?

Coming Soon ...

The following features are yet to be implemented -

1. Pyani for Average Nucleotide Identity (heatmap)
2. Mashtree (dendrogram)
3. Pretty Visualizations (if time permits)

Also, a cool name for the server!