Team 1: Web server

Devishi Kesar, Shuheng Gan, Winnie Zheng, Priya Narayanan, Aaron Pfennig

Genome Assembly
De-Novo assembly of genomic sequences

Gene Prediction
Identify encoding regions in the DNA

Functional Annotation
Obtaining functional information of the genome

Comparative Genomics
Comparison of genomes features of different organisms

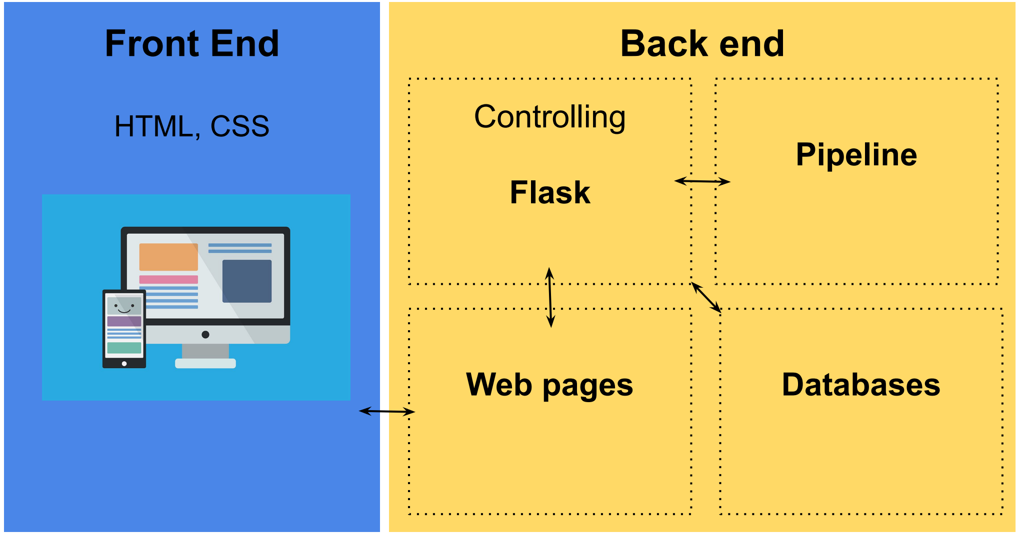
Web server
Platform providing analysis of virulence factors in *E.coli*

Aims

- Provide a comprehensive, automated platform to analyze *E.coli* isolates in order to predict virulence factors
- Functionalities of the web server:
  - Identify virulence factors/microbial resistance in provided isolates
  - Allow data upload at each step of outline pipeline
  - Visualize findings in an comprehensible way
- Design
  - Intuitive usage
  - Provide only essential options

Technical notes

- Rely on Python framework **Flask**
    - Varieties of choice when developing a web application (tools, libraries and mechanics)
    - Categorized as a micro or lightweight framework with minimal or no dependencies on external libraries
    - Easy to get started with
    - Well documented
    - Popular choice as easy to integrate with other tools
    - Used by companies like Netflix and reddit in their tech stack



Front End

HTML, CSS

Back end

Controlling

Flask

Pipeline

Web pages

Databases

Genome Assembly

- Performs *de-novo* assembly with FastQ files as input
- Runs following tools by default:
  - fastp : read pre processing
  - Unicycler : Genome assembly
- Options:
  - Perform read preprocessing
  - Kmer-size
  - Spades as alternative assembly method
- The input FastQ files must be paired end reads
- Outputs FASTA file

Gene Prediction

- Gene finding in assembled isolates or provided FASTA fileTakes FastQ files as input
- Runs following tools by default:
    - CDS: Prodigal
    - tRNA: Aragorn
    - rRNA: barrnap
- Options:
    - GeneMarkS-2 as alternative tool for CDS predictions
    - tRNAscan-SE as alternative tool for tRNA predictions
    - RNAmmer as alternative tool for rRNA predictions

- Outputs *.gff file, *_cds.fna file, *_protein.faa file and *_rna.fna file

Functional Annotation

- Obtain functional information about predicted genes
- Input: FASTA file
- Cluster Tool: usearch
  - Output: centroid.fasta
- Homology Tools:
  - General annotation: InterProScan, EggNOGmapper
  - Antibiotic resistance gene: DeepARG
- Abinitio Tools:
  - Signal Peptides: SignalP 5.0
  - Transmembrane Proteins: TMHMM
  - CRISPR Sites: CRT
- Output: *.tsv file

Comparative Genomics (tentative)

- Compare genomic features with other organisms
- Input: FASTA file
- Tools used:
  - MUMmer
  - stringMLST
  - kSNP

Demo

Next steps

- Include final pipeline
    - Add missing options to the web server
- Improve design
- Visualizations as final output