

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

Background And Strategy

Content

1. Introduction
2. Goals
3. Methods and Technology
4. Visualization
5. Questions

What is a Web Server?

- A server that hosts web pages as requested.
- It stores, processes data at the same time.
- It takes in requests in the form of HTTP.
 - **H**yper **T**ext **T**ransfer **P**rotocol
 - Client Server Protocol

What is a Predictive Web Server?

- Quite simply, a web server that predicts something.
- Eg., Web Server to predict RNA secondary structure, or Cellular and Viral Internal Ribosome Entry Site, etc.
- In our case, the web server would ideally predict a phenotype based on the genetic information it is given.

Goal

The Web Server should provide the following:

- An easy-to-use tool to help distinguish between *Klebsiella* phenotypes, by implementing the work of the comparative genomics group.
- A robust and easy-to-use web-based de-novo assembly tool.
- A feature to visualize and download the results of the 258 genomes.

Servers Out There

Data intensive biology for everyone

Galaxy is an open, web-based platform for accessible, reproducible, and transparent computational biomedical research.

- **Accessible:** Users without programming experience can easily specify parameters and run tools and workflows.
- **Reproducible:** Galaxy captures information so that any user can repeat and understand a complete computational analysis.
- **Transparent:** Users share and publish analyses via the web and create Pages, interactive, web-based documents that describe a complete analysis.

Welcome to the Galaxy Community Hub, where you'll find community curated documentation of all things Galaxy.






Galaxy

- Galaxy is an open source, web-based platform for data intensive biomedical research.
- One of the most popular web-servers in Bioinformatics Research.
- Does not do any prediction.
- A scientific workflow system that provides a GUI for researchers with no background in Bioinformatics Programming.
- Developed in Python and JavaScript.

DUET

DUET: a server for predicting effects of mutations on protein stability via an integrated computational approach

Douglas E. V. Pires, David B. Ascher, Tom L. Blundell

Nucleic Acids Research, v. 42 (W1), p. W314-W319, 2014.   

Abstract

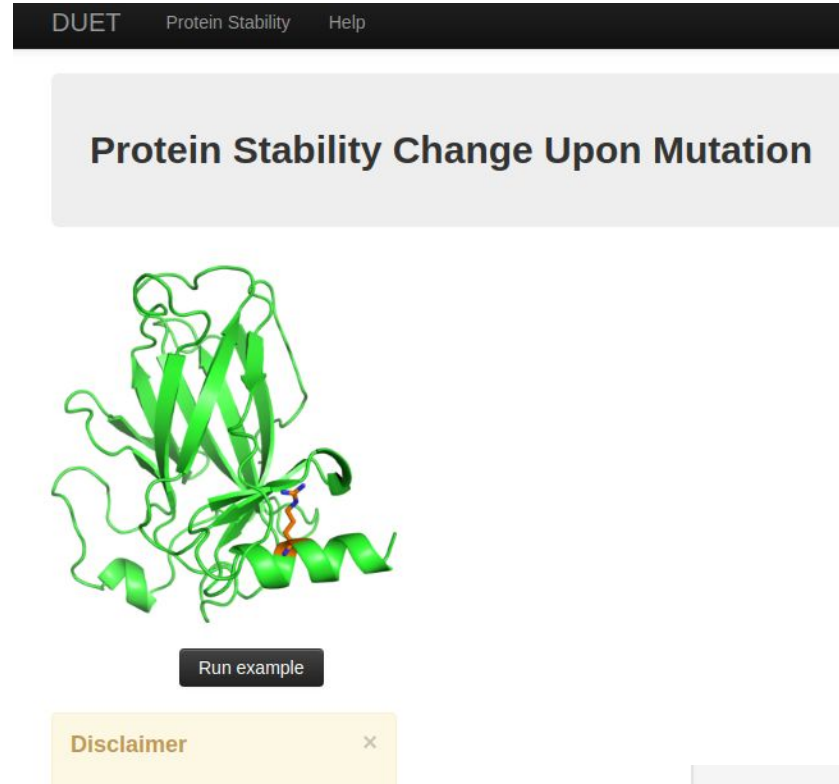
Cancer genome and other sequencing initiatives are generating extensive data on non-synonymous single nucleotide polymorphisms (nsSNPs) in the human and other genomes. In order to understand their impacts on the structure and function of the proteome, as well as to guide protein engineering, robust *in silico* methodologies are required to study and predict the effects of nsSNPs on protein stability. Despite the diversity of available computational methods in the literature, none has proven robust and dependable on its own under all scenarios where mutation analysis is required.

Here we present DUET, a web server for an integrated computational approach for studying missense mutations in proteins. DUET consolidates two complementary approaches (mCSM and SDM) in a consensus prediction, obtained by combining the results of the separate methods in an optimised predictor using Support Vector Machines (SVM). We demonstrate that the proposed method improves overall accuracy of the predictions in comparison with either method individually and performs as well as or better than similar methods.



DUET

- Predicts change in stability of a protein upon the introduction of a single mutation
- Uses a combination of two distinct approaches:
 - SDM - a statistical potential energy function
 - mCSM - a predictive model using graph-based signatures



The screenshot displays the DUET web application interface. At the top, a dark navigation bar contains the text "DUET", "Protein Stability", and "Help". Below this, a light gray header box features the title "Protein Stability Change Upon Mutation". The central part of the interface shows a 3D ribbon model of a protein structure in green, with a specific region highlighted in red and blue. A dark button labeled "Run example" is positioned below the protein model. At the bottom, a yellow banner contains the text "Disclaimer" and a close icon (an 'x').

Methods And Technology:

- Implementation.
 - Explore
 - Downloads
 - Tools.
- Back-end.
- Front-end.

User Functionality

Explore

Project summary

Github

Pipelines

Downloads

Data Dump

Visualize data

Tools

Assembly

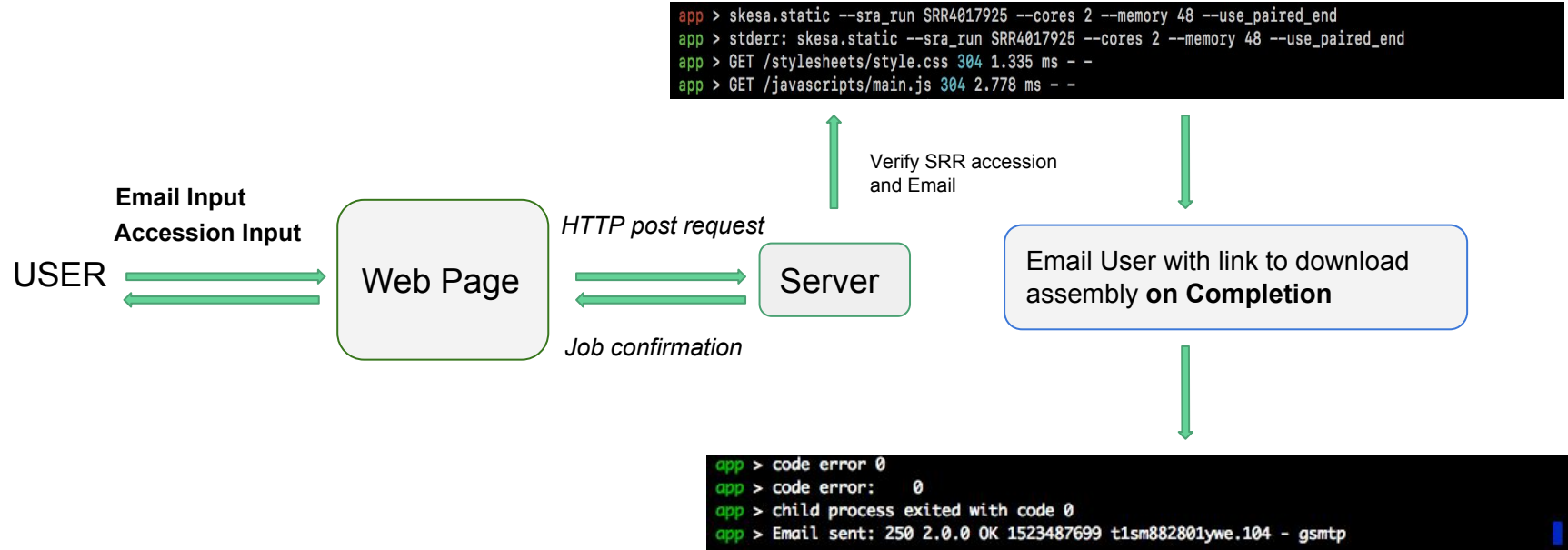
Comparative

Assembly

Three input modes:

1. Single SRR read (input a single SRR accession).
2. Batch mode (input a list a SRR accessions).
3. Upload fastq reads from your local machine.

Assembly workflow



Assembly UI

BIOPREDICT

Georgia Institute of Technology *Klebsiella Predictive Webserver*

Genome Assembly

De Novo Assembly using Skesa

About this pipeline

Input Type: *Select read types: Local Reads correspond to reads stored on your local machine.*

SRA Accession: *SRA option handles file uploads for you and saves you time and storage*

Execute

Reset

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

Downloads

- The downloads section will provide the data created by the previous groups.
- The downloads table is automatically generated based on SRR IDs stored in MySQL database.
- The downloads section also allows the users to visualize stats that are based on the data.

Basic Layout (Downloads)

BIOPREDICT

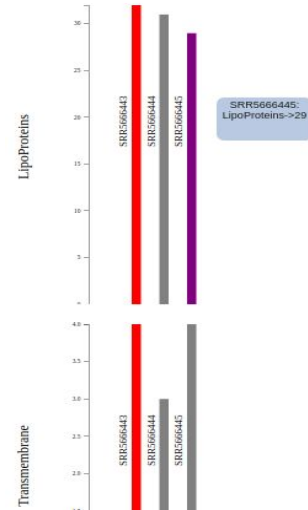
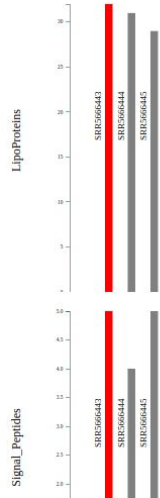
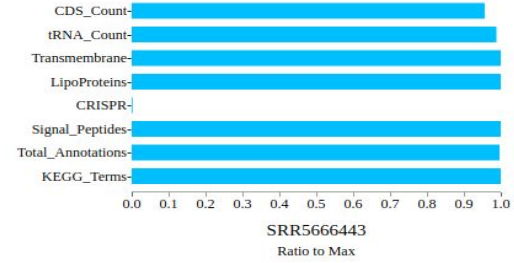
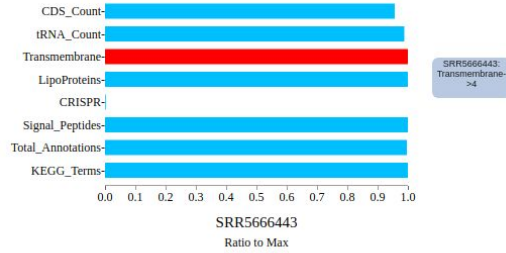
Georgia Institute of Technology *Klebsiella* Predictive Webserver

Downloads

Klebsiella pneumoniae data dump

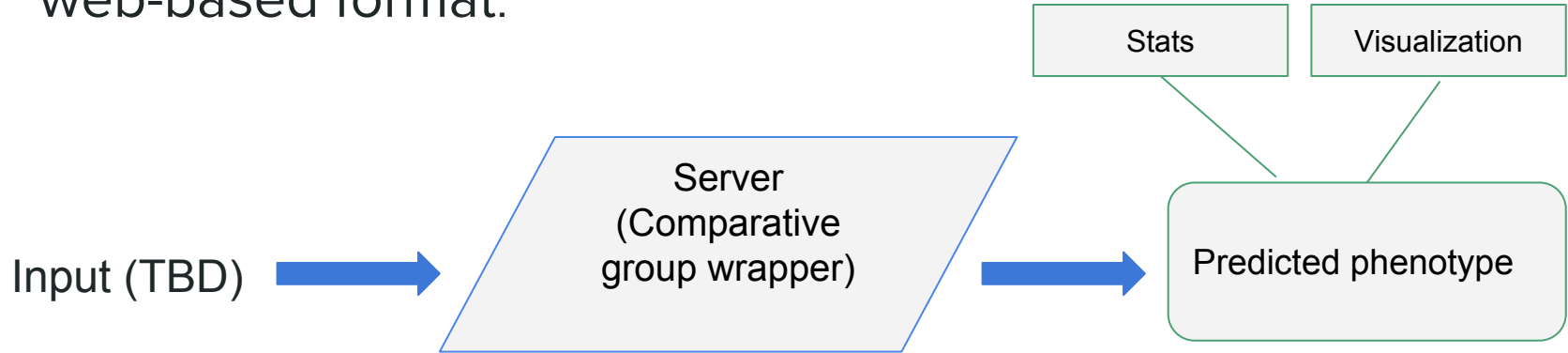
Accession	Assembly	Annotation
SRR3467249	FASTA	GFF
SRR3982229	FASTA	GFF
SRR3982230	FASTA	GFF
SRR3982253	FASTA	GFF
SRR3982316	FASTA	GFF
SRR3982346	FASTA	GFF
SRR3987112	FASTA	GFF
SRR3987115	FASTA	GFF
SRR3987118	FASTA	GFF
SRR3987119	FASTA	GFF
SRR3987120	FASTA	GFF
SRR3987121	FASTA	GFF
SRR3987122	FASTA	GFF
SRR3987123	FASTA	GFF
SRR3987124	FASTA	GFF
SRR3987125	FASTA	GFF
SRR3987126	FASTA	GFF
SRR3987127	FASTA	GFF

Visualization



Comparative

- Currently we are developing methods to include results from Comparative Group and make them recreatable in a web-based format.



BACK - END DEVELOPMENT

- NodeJS
- MySQL: Designing a relational database to track our SRR IDs, Genes and annotations.
- Linux Cron jobs for server maintenance

FRONT - END DEVELOPMENT

- CSS/HTML/Javascript.
- Our front-end approach:
 - Simple
 - Compact
 - Easy to use.
- D3.js (for visualization).

Next steps

Plans

- Incorporate comparative functionality.
- Implement the explore section.
- Implement batch mode for assembly.
- Create our MySQL database.
- Finish D3 visualization.

Questions?
