

Hamza Khan ¹, Ivy Dong, Jeremy Jeroham Hartmann, Naman Paul ¹, Penghzi Duan

School of Computing Science, Simon Fraser University, Burnaby, BC, Canada

¹ The University of British Columbia, Vancouver, BC, Canada

Abstract

Implementation of a web application that utilizes high-throughput genome sequencing data from The National Center of Biotechnology Information (NCBI). This would enable researchers, doctors and other health-care authorities to analyze the patient's sample data to determine the infectious agent providing valuable information to clinicians for administering the most effective treatment.

Introduction

Infections acquired in hospitals and other healthcare facilities are termed as nosocomial infections. Canada spends about \$1 billion annually for treating these infections. According to the Canadian Committee on Antibiotic Resistance, about 30% of these are preventable. Over 99,000 deaths are reported in United States of America alone out of the 1.7 million cases.



Figure 1: Possible surfaces of contamination in a hospital ward

In addition to many solutions pre-existing to combat these infections; our system lays out a prototype which can be used as a detector tool and as an additional monitoring layer for reporting and predictive diagnosis.

The system is a suite of wide range of applications includes the following:

1. Patient Information Management
2. BLAST (Basic Local Alignment Search Tool)
3. PHYLIP (Phylogenetic Tree Builder)
4. Tree Viewer (Phylogenetic Tree analyzer)
5. Data Reporting

Scenario

An infectious outbreak in a hospital never reported before. Doctors attempt to control the spread, however the infectious agent (say, a bacterium) seems to be a new resistant strain; with no treatment catalogued. With our system, the patient sequence could be matched against the database and based on comparative genomics results, a custom treatment could be created while considering the efficacy of pre-existing treatments for a close relative, from which the new resistant strain might have evolved.

Methodology

Procedural Flow



All the various applications have been knit into the python framework known as django (version 1.8.3). This facilitates secured access into the system with valid credentials.

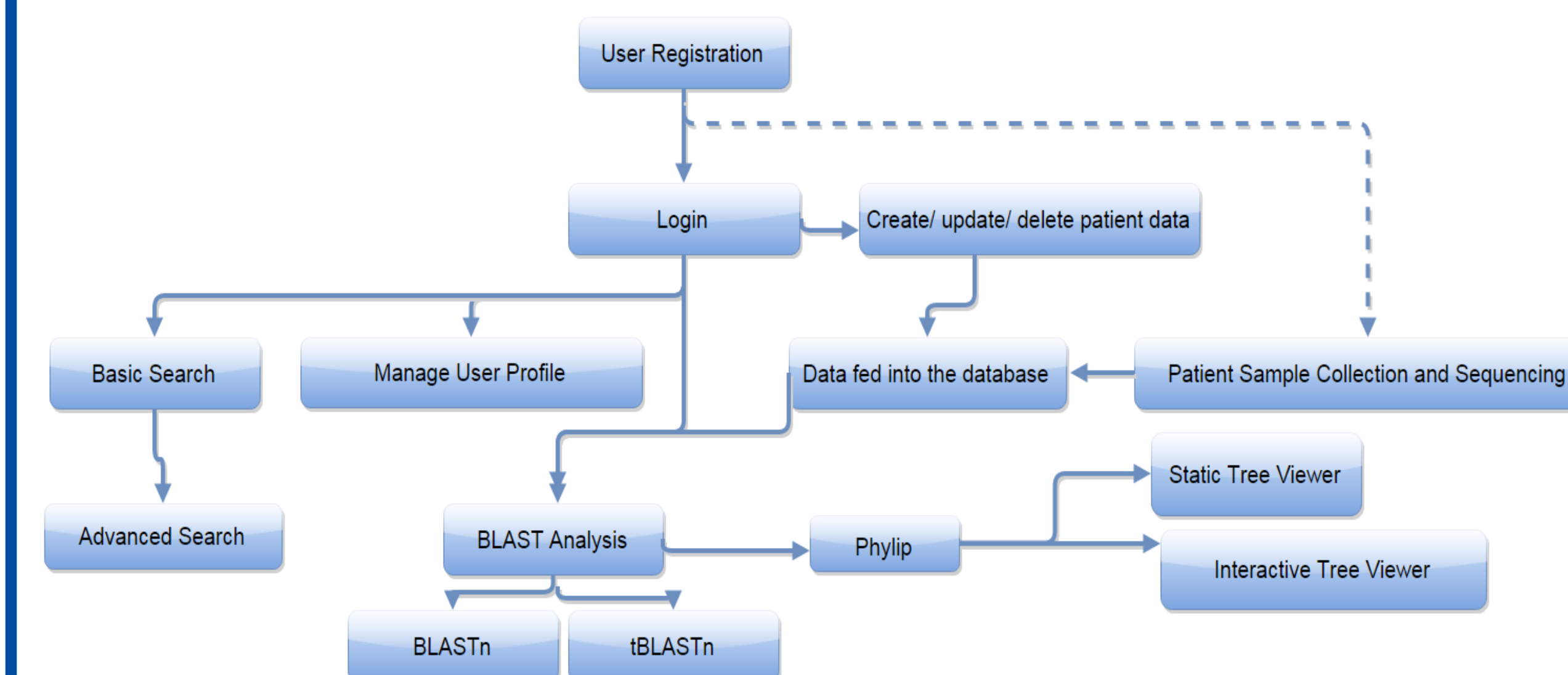


Figure 2: A general workflow of the implementation of NISS

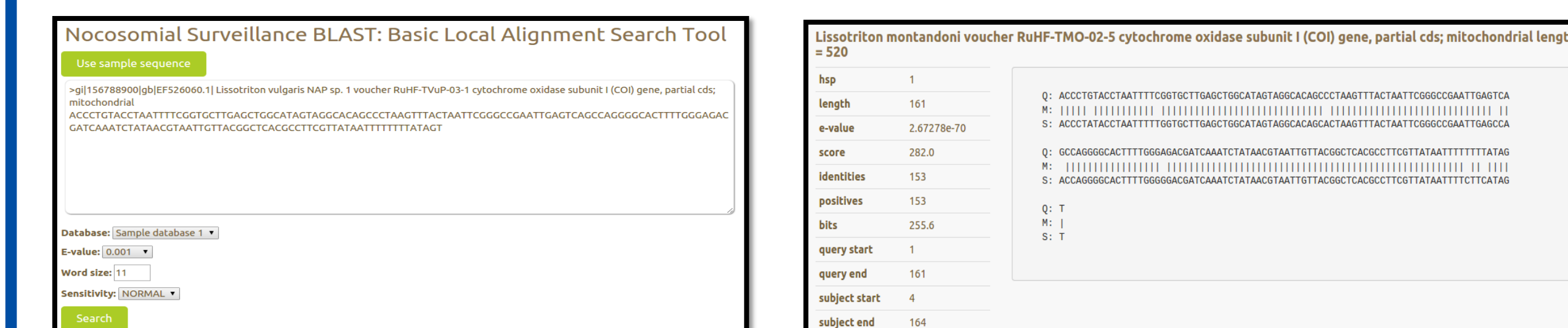


Figure 3: BLAST search and result pages

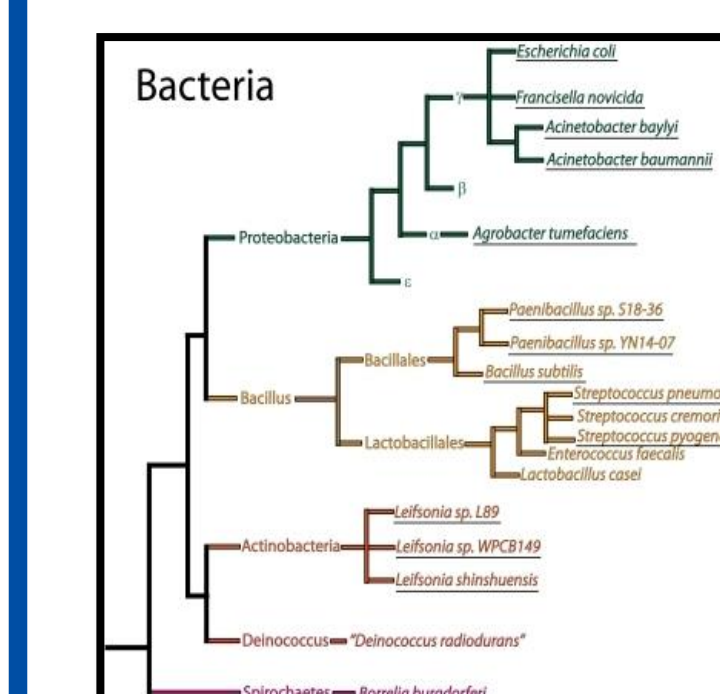


Figure 4: Dendrogram

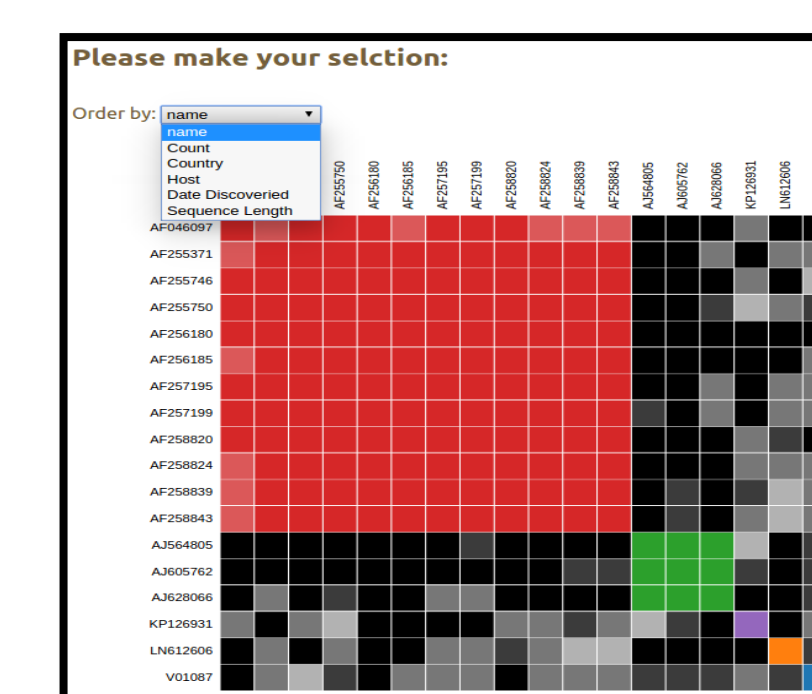


Figure 5: Matrix Visualization

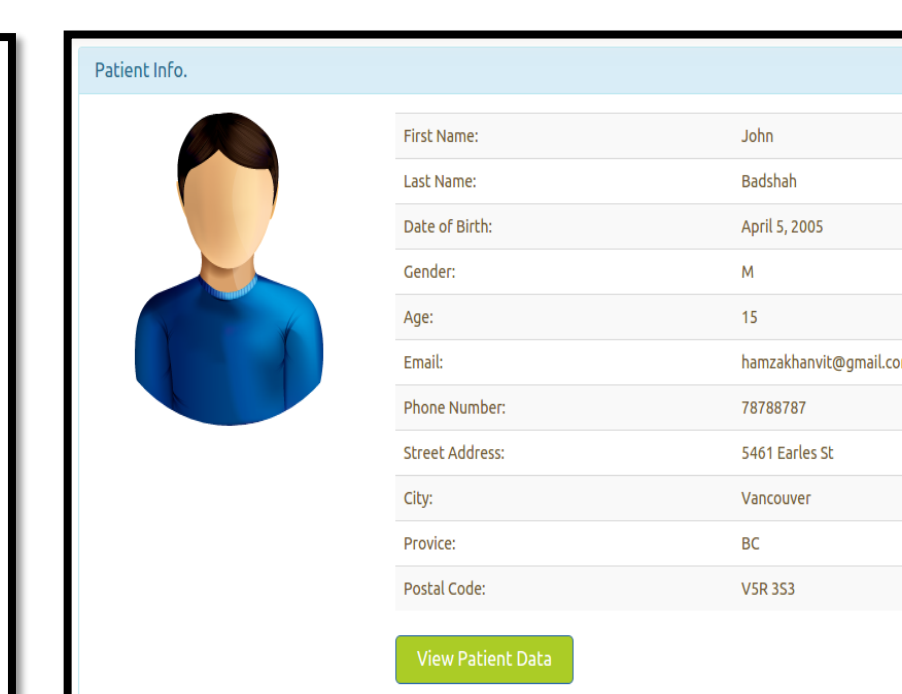


Figure 6: Patient Profile

Technologies used



Views

- As you get on the main page of our system, users such as doctors, scientists or researchers could register themselves and login to access all the functionalities.
- The user can then create, update or delete patient data.
- Once logged in, a basic search could be performed to mine the NISS database.
- Currently the database has 1,500 nucleotide and 1,500 protein entries from NCBI's influenza genome. NISS has the capability to download "N" number of entries through Django.
- Advance search options include GenBank accession number, virus type, genome or protein segments, strain name, subtype, country of isolation, sequence completion status and year of isolation.
- The BLAST feature equips the researchers, with a powerful platform to analyze the similarity of the patient's sequence against NCBI database. The top 10 results which are the most similar to the input sequence are then taken forward.
- The PHYLIP applications performs Multiple Sequence Alignment of the BLAST results and generates phylogenetic trees in a newick format. These trees are visualized using javascript and d3.js.

Features

- Secured access to patient information, only for doctors/ researchers.
- REST API to access NISS database.
- Option to populate the local database with as many entries as required.
- User friendly GUI as opposed to the command line tools that are usually used in bioinformatics analysis.
- Maintain patient and user records with proper authentication.

Results

The Nosocomial Infections Surveillance System is a competent prototype of an additional screening process, which can be adopted and would be highly comprehensive in taming these infections. The use of a python framework such as django, ensures restricted access limited to users with appropriate authentication. This eliminates the factor of information leakage, since the database contains sensitive patient data. Applications such as BLAST and PHYLIP, aid the comparative genomics outlook that are the key players in predicting the infection-causing agent.

Future Directions

- The functionality of this system-prototype could be extended for
- Screen for identifying point mutations
- Sequence annotation for new submissions
- Single Nuclear Polymorphism (SNP) detection
- Application of Genetic Algorithms, for evolutionary studies
- Metadata analysis
- Faster analyses and reporting

References

- [1] Fournier et al. Modern clinical microbiology: new challenges and solutions. Nature Reviews Microbiology 11, 574–585(2013).
- [2] Sherry et al. Outbreak Investigation Using HighThroughput Genome Sequencing within a Diagnostic Microbiology Laboratory, J ClinMicrobiol. 51(5):1396–1401, (2013)
- [3] Underwood et al. Public Health Value of NextGeneration DNA Sequencing of Enterohemorrhagic Escherichia coli Isolates from an Outbreak. J. Clin. Microbiol.51: 232-237, (2013).