

Help Solve Disease Outbreaks: Design and Implementation of a GUI for a Bioinformatics Pipeline for use in Infectious Disease Research and Investigations

Sponsor Information:

Darrin Lemmer, Associate Bioinformatician
Translational Genomics Research Institute (TGen)
3051 W. Shamrell Blvd, Ste 106, Flagstaff AZ
dlemmer@tgen.org

Project Description:

TGen North is focused on advancing public health through the study of bacterial and fungal pathogens, primarily by looking at their genomic sequences. To assist with these studies, we have developed NASP, a command-line based bioinformatics pipeline written in Python which accepts as input a collection of genomic sequencing data files, chains together some external tools to align them and determine single nucleotide polymorphisms (SNPs) or places where one or more of the input samples have sequence differences, and then computes a matrix detailing the positions where SNPs were found and what those SNPs were for each sample. Doing this allows us to compare many different samples at once to determine if they are part of an outbreak and to find the source. For example, this technique was used to trace a cholera outbreak that occurred in Haiti after the 2010 earthquake to a group of UN peacekeepers from Nepal. We can also compare multiple strains of an organism to determine how the different strains evolved and what makes some more virulent than others, and ultimately create clinical assays to detect what strain a patient has in order to treat them more effectively.

We require a graphical interface to make running NASP and examining the results easier to use for scientists who are not as familiar with command-line interfaces. The UI should assist with selecting input files (which may already reside on the server, or may need to be uploaded to it), allow the user to enter any other parameters that the pipeline needs to run, monitor the progress of the run, and display the output files when the run completes. As the run make take hours or days to complete. The user should be able to close and reopen the UI at any time and reconnect to any running or finished projects that they have started. The UI may be web-based or a client/server application that the user runs on a Mac-based computer – support for Windows and Linux is desired but optional – or both. NASP itself will be run on a large Linux cluster.

Knowledge, skills and expertise required for this project:

- Graphical user interface design
- Python programming (proficiency in other languages may also be helpful)
- Knowledge of user interface toolkits/frameworks (GTK+, Cocoa, Eclipse SWT, Django, etc.)
- XML
- MacOS X and Linux
- Software Configuration Management with Github
- Unit Testing
- Good communication skills
- Bioinformatics knowledge/experience will be helpful but not required

Equipment Requirements:

No special equipment is required for this project. We will provide VPN access to our compute cluster and web server.

Deliverables:

Generate a design document with screen shots/mockups and implementation details, developed after gathering requirements from multiple users, which must be reviewed and approved by the user community at TGen North.

Implement the approved design, using a version control system, good code documentation, and unit testing for verification, and present the completed project to the user community.