**410.712.82 - Advanced Practical Computer Concepts for Bioinformatics**

**Johns Hopkins University, Krieger School of Arts and Sciences**

**Advanced Academic Programs, Biotechnology Studies**

Name: Brian Poon

UNIX Username: bpoon1

**Final Project – Proposal**

**Background and Purpose**

The main challenge in starting a new project is determining what the project will be based on- what the goal of the project is. In considering the intention of this final project with this final proposal, this particular conundrum was persistent with project ideas resulting in one of two conclusions that is paired with two dynamic reasonings. The first is that the idea has been done before, which spans from a difficulty of simple to moderately complex. The second is that the idea has not been done before or does not have an adequate tool for. However, in gauging the difficulty to accomplish the idea and taking into account the resources available, the idea is usually too complex to accomplish in the time frame.

During the brainstorming for this project, a thought had come to mind after viewing the variety of bioinformatics tools available online. Many of these tools are specific in nature or stem from large, established host with enormous database resources. However, if one was to use these tools and host-databases, it would require inputting the same sequences or starting information in each of them. Hence, the thought came to mind for this idea: a Swiss army knife for bioinformatics. A tool where it brings together the tools from multiple resources for a high-level overview of a particular DNA, RNA, or protein sequence. These high-level overviews can helpful during initial introduction of a sequence, such as that of a student learning about an assigned sequence, or for someone engaged in research to broadly look for leads on multiple sources. The input to the tool will be a sequence of DNA, RNA, or Protein. The output will first include a basic analysis summary of the sequence and display the results from other bioinformatics tools as if the sequence was searched with those tools, such as NCBI search, Ensembl search, or BLAST. Lastly, the results will be recorded in the backend and the frontend page will have an option for downloading the information.

**Technical Details**

Starting Page: The starting page will consist of HTML and CSS. The most useful aspect of any tool is the options that it comes with. The starting page will come with the option to input the sequence into a search box. An option to hopefully incorporate is the option to upload sequences as FASTA files. Additionally, some options to consider for the results page would be using checkboxes to identify what resources wants to be used. Once the submission button is clicked, it will lead to the results page.

Results Page: The results page will be comprised of HTML and CSS. However, it will use Computer Gateway Interface (CGI) written using Python to perform its tasks, then publish the results on the HTML results page. The results page will consist of the following information:

General Summary: The *General Summary* portion will include basic analysis of the sequence. It can include the *type* of sequence (DNA, RNA, or Protein), the breakdown of residues or nucleotides, and the length of the sequence. There can be special outputs shown based on the type of the sequence. Sequences containing DNA can have G/C content ratios.

**Design and Development**