Kristen Steenbergen

March 30, 2021

410.712 Advanced Computer Concepts

Final Project Proposal

My proposal for the final project is a web application that will display the significant mutations of the most common hCoV-19 variants, provide annotation and other genomic information using the hCoV-19 reference genome. This will be considered a 'data-mining tool' that displays less-accessible data.

The entire hCoV-19 genome is 30Kb in length and is readily available from NCBI Nucleotide Database. I also have obtained access to GISAID which has a detailed annotation of the hCoV-19 Reference Sequence. Significant mutations and SNPs can be found at several open-source websites, but I will be <https://cov-lineages.org/global_report.html> to provide my source data for the significant variant mutations.

For SQL database utilization, I intend to parse the GenBank file for the reference genome and input that data into a 'reference genome' table indicating the location of genes, gene names, gene length, etc. A second table titled 'variant mutations' would include the variant name, location of first identification, mutation, mutation type (insert/del/sub), synonymous/nonsynonymous, genomic location, and gene name. The key that would connect the two tables would most likely be the gene (an alternative option would be the loci).

A CGI script would run the query, match the variant to the genomic information from the reference genome, and return an object whose attributes contain the various pieces of information needed. The query results would be displayed in a table format using the Jinja template. The table would display defining mutations, where they are found in the genome, what gene is affected, etc. Depending on how the project progresses, I may instead display the data as a comparison between two of the variants instead of highlighting one variant at a time.

Jquery UI autocomplete will be used for querying the variant names. Also, the data display will use the JQuery UI Accordion widget to display the table categories or the Color Animation to expand the displayed results. Since I intend to use a Jinja template, I will not be using AJAX, so JQuery will be the extent of the JavaScript at this point.

Due to the ever-changing nature of the COVID-19 pandemic, this page will have to be time-stamped with when the data was last updated, and I will limit the variants to the current top 4 using PANGO lineage nomenclature (B.1.1.7, B.1.351, P.1, B.1.1.28). If there is a significant change in the variant situation and I have the resources to respond to that, I will update it accordingly.

I will prepare two HTML pages, one for the query and the template HTML displaying the results. CSS would be applied to improve aesthetics.