**Project Proposal**

Many plasmid genomes are manipulated to express the phenotypes of genes inserted by researchers. In order to grow colonies and test these types of phenotypic changes, researches cut sections of the plasmid in order to insert new genes or remove genes. They are able to do this with the use of restriction enzymes, which “digest” the DNA by cutting it at sites which hold specific nucleotide patterns.The objective of this project proposal is to provide a web interface which serves as a restriction enzyme annotation and mapping tool as it pertains to genomes queried by users. The genomes uploaded by users will have their overall properties and their genes annotated, as well as the sites of possible digestion based on the restriction enzymes in the database. The user will have the ability to select which enzymes they wish to test a digest on and the results that are returned to the user will be a list of DNA fragments and their properties, including any genes which are cut as well as complete genes within the fragment. This list will serve as a “map” to the user.

The HTML page will utilize a CSS file for visual effects, as well as Javascript widgets to improve the user interface. The user input will consist of an accession number which will be used to grab FASTA and GenBank files from NCBI or they will have the option to upload files using JQuery widgets. Once these files are uploaded, a CGI script will parse both files and pull all of the annotated data out, compiling the data nicely into the database schema, which includes organism information, gene information, and the location of digestion sites for all restriction enzymes held in the database. The idea is to maintain and build a library of plasmid genomes which can be easily accessed after they are uploaded. The users will have the ability to add genomes or select from a list of currently held genomes. After this is complete and the genome is either added or selected from the database, the user will be redirected to a different HTML in order to select which restriction enzymes they wish to query against this plasmid. Based on user input, a query will run against the database for the enzymes selected. The locations of the sites will be used to calculate fragment properties, and the data from the annotated genes will also be used to calculate the gene content within each fragment, as well as which genes would be lost due to digestion. This dataset will be returned to the user in a viewer-friendly way which is easy to understand.

The schema for the database will include four tables. The first table will include organism information such as an organism ID (accession number), the genus, species, plasmid vector name, length of plasmid, description of plasmid (if any), and the number of genes within the plasmid. The gene table will hold a geneID, an organism ID, gene name, length of gene, start site, stop site, strand, and description (if any). The restriction enzyme table will hold the enzyme ID, the name of the enzyme, and the nucleotide restriction pattern. Finally, the last table will be the site table which will hold a site ID, organism ID, enzyme ID, and location. This table will be populated once the organism and gene tables are updated. Once these tables are successfully updated with all the annotated information, the user will be able to run queries against the database for different digestion scenarios. The database will pull up site information for each selected enzyme and divide the genome up into fragments. These fragments will be queried against the gene table to determine the gene content and gene digestion. Once the fragment properties have been successfully stored in a fragment object, these objects will be uploaded to a results page for the user to see.