Computing Tools Project Proposal

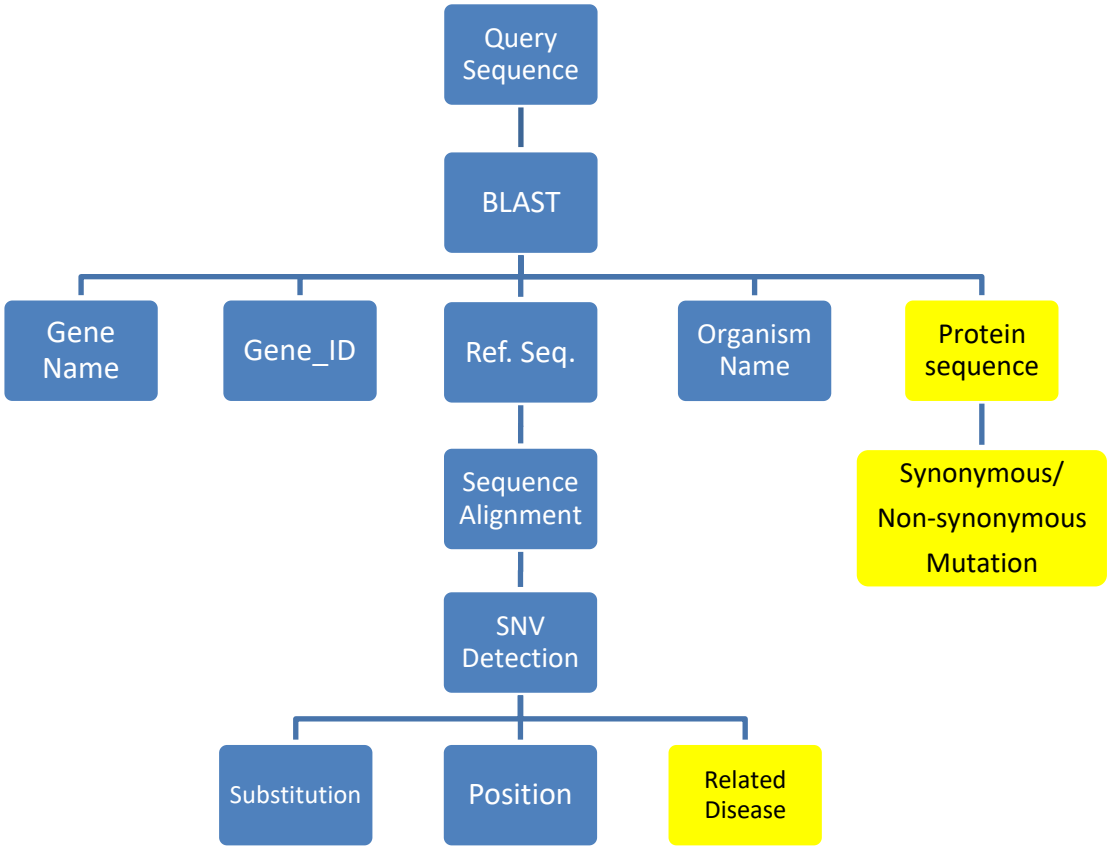
Aim : To predict SNV from query sequence(with one nucleotide mutation) and preparation of related database form the output results.

Introduction :

Prediction of SNV (Single Nucleotide Variation) in given query sequence containing single mutation and comparison of this with reference sequence obtained from BLAST (Basic Local Alignment Search Tool). Sequence alignment of reference and query sequence will lead in detection of conserved region with one mutated nucleotide/amino acid. Finding whether the mutation is synonymous or non-synonymous. This mutation will ultimately give us the information of disease related to that particular mutation.

The database will include **Organism Name, Gene Name, Gene ID, Substitution Type and Position, Mutation Type, Related Disease.**

Workflow:



Modules to be Used :

Biopython, Numpy, Scipy, MySql, SqLite3, Matplotlib, Some User Defined Modules if required, etc.

Steps Involved :

1. Query sequence (with one nucleotide change) **.fasta file** is taken as user input to python script.
2. Blast is performed on the query sequence by using **Biopython module.**
3. Blast results in **.xml file** format are stored for further implementations.
4. .xml file is parsed to get desired results i.e. closely related **gene name**, **gene_id**, **organism name** and sequence as **reference sequence**.
5. Reference sequence and Query sequence is aligned to get the substitution of nucleotide.
6. **Position** of the substitution is also determined here.
7. Respective translated protein sequences can also be extracted using BLAST.
8. By comparing protein sequences it can be predicted whether the mutation is synonymous or non-synonymous
9. If the gene is related to human by comparing results on OMIM(Online Mendelian Inheritance in Man) database to know related disease due to that mutation.
10. Database of all the results are made using **DBMS(Database Management System)**.

Conclusion :

Integrating data from publicly available databases and using various modules in python will give overview about SNV and related disease.

Note : Yellow highlighted part will be done if time allows.