

A Graph Representation of the Genomes Collected in a Viral Outbreak

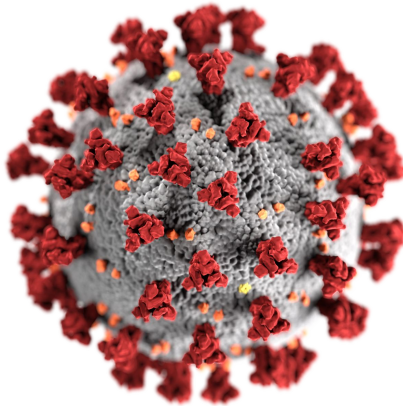
Michele Ambrose, Jake Fisher, Jared Hubert, Drew Pompeii, Caroline Turner

Dr. Edward Fox

CS 4624 Multimedia, Hypertext, and Information Access

Virginia Tech, Blacksburg VA, 24061, 4/29/21

Outline:



Project Overview

User Walkthrough

Back-End Work

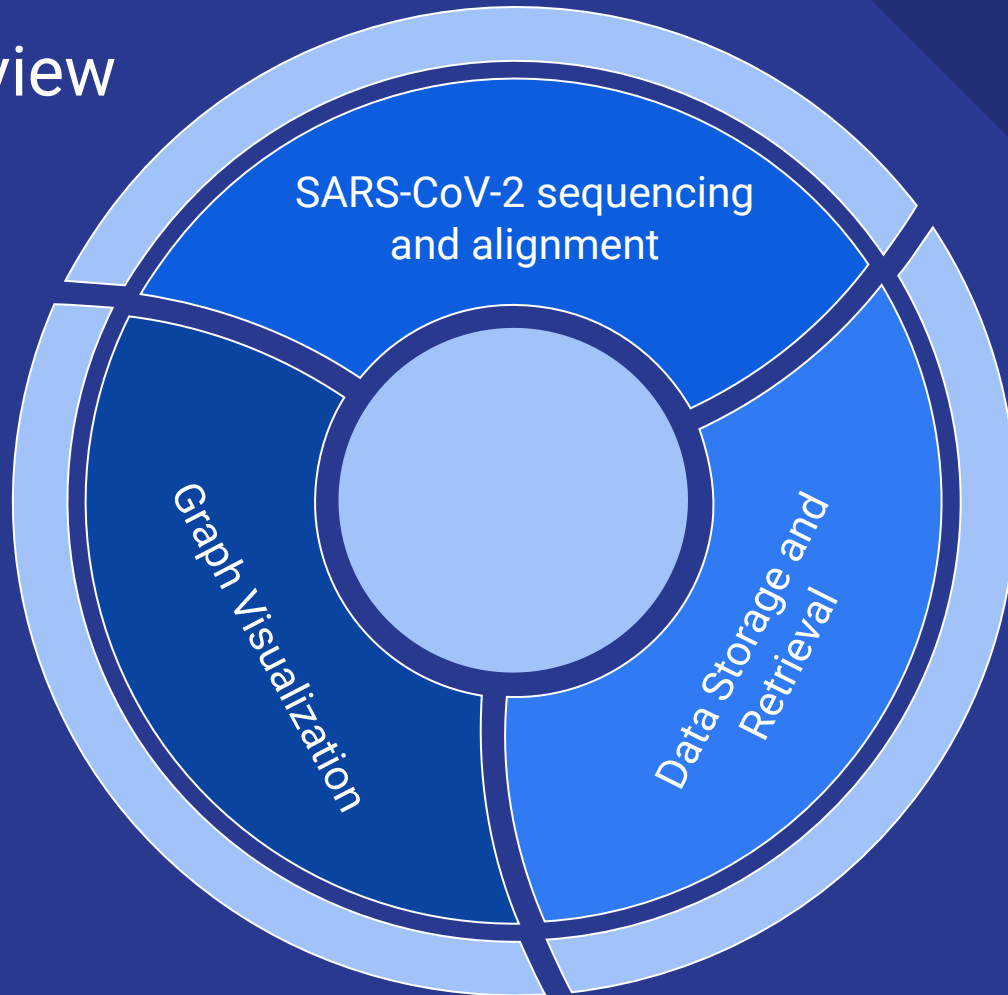
Genome Alignment

Challenges Overcome

Moving Forward

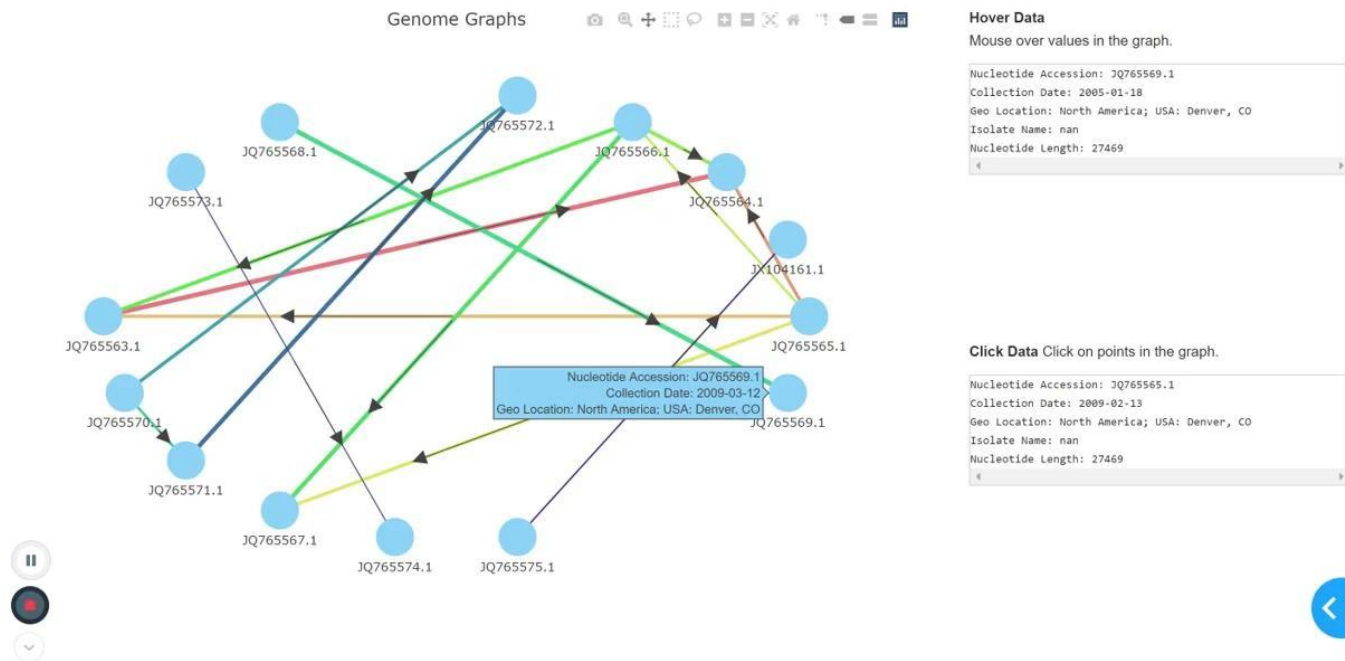
Acknowledgements/References

Project Overview



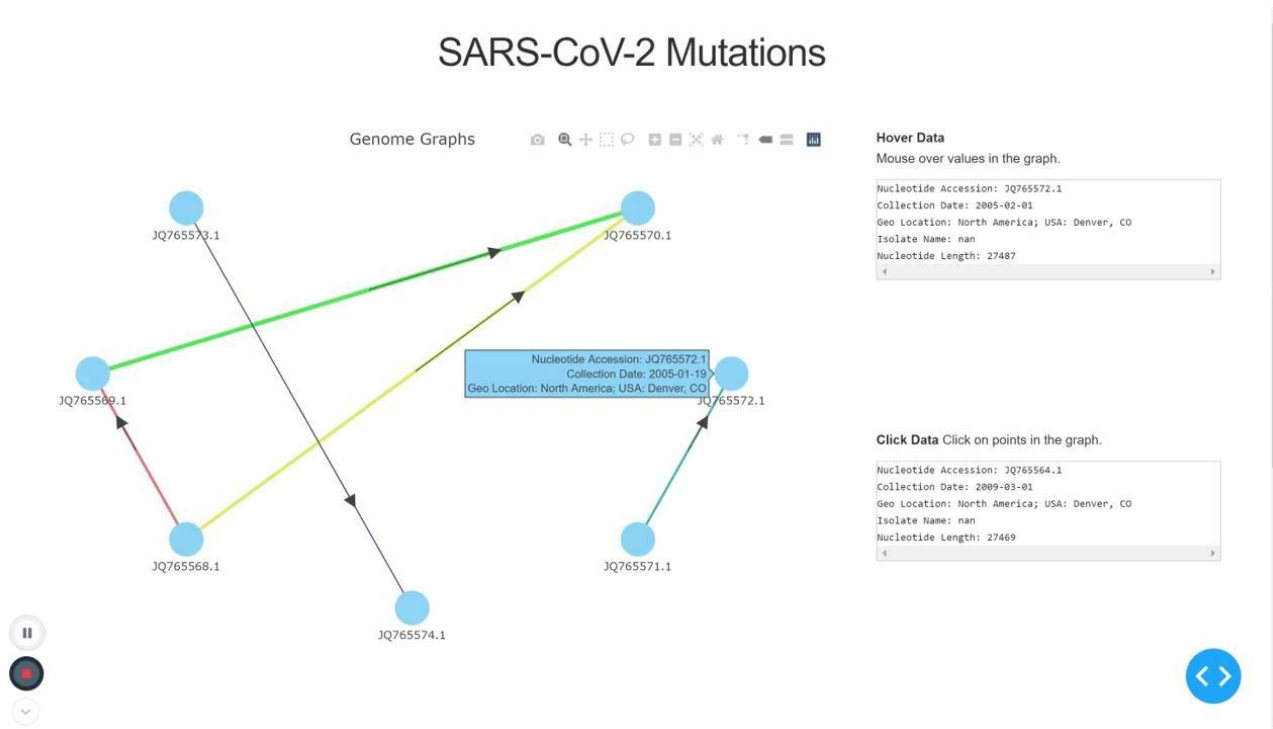
User Walkthrough - Graph Interaction

SARS-CoV-2 Mutations

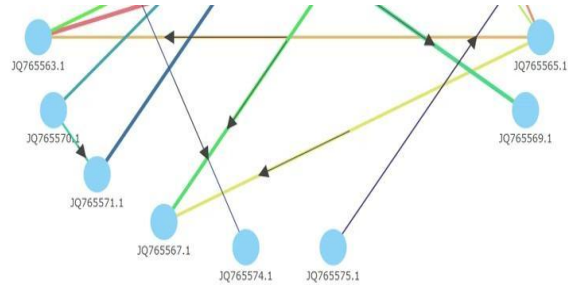


User Walkthrough - Data Manipulation

SARS-CoV-2 Mutations



User Walkthrough - File Import/Export



Click Data Click on points in the graph.

Nucleotide Accession: JQ765565.1
Collection Date: 2009-02-13
Geo Location: North America; USA; Denver, CO
Isolate Name: nan
Nucleotide Length: 27469

Data Modifications

UPLOAD GENOME FILE

UPLOAD EDGES FILE

EXPORT DATA TO CSV

Date Picker Enter the range of dates of the genomes you want to list. Use the format 01/01/2021.

Start Date → End Date

Select a date to see it displayed here

Limit by Location Enter the location you want to find genomes from.

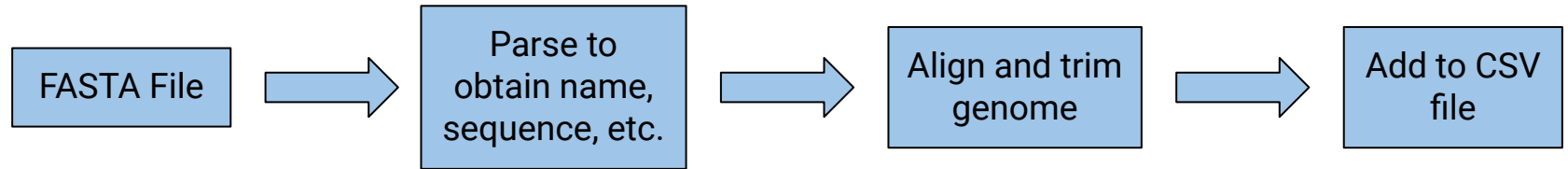


Limit by Nucleotide Length Enter the lowest size of genomes you want to search for.



Back-End Work

- Parse through imported FASTA file to obtain genome names and sequences
- Implementation to add, remove, search of genomes in CSV file
- Able to search and obtain genome by type, name, sequence
- Alignment



Genome Alignment

- The arrangement of genome sequences to identify regions of similarity
- Purpose: Aid in function prediction
 - Show homology between related members
 - Investigation or prediction possibilities
- MUMmer4



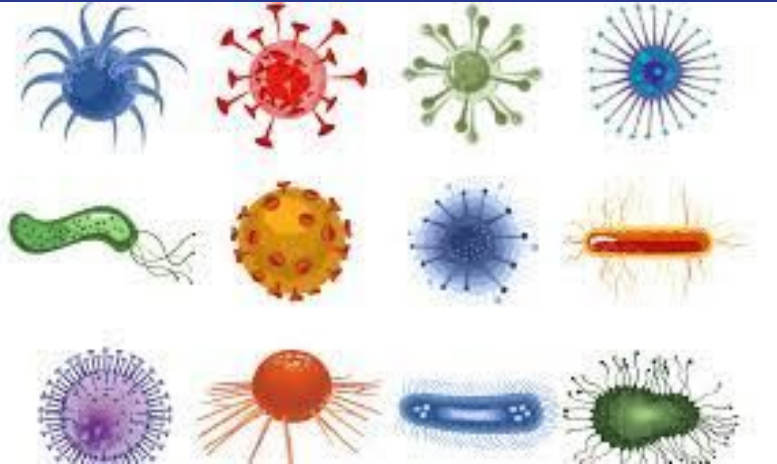
Challenges Overcome

- Lack of knowledge about genomes
- Learning new coding languages and APIs
- Issues with working online



Moving Forward

- Applied to different viruses and types of genomes
- Auto-populate extraneous columns
 - Allow for choice of data
 - Multichoice menus
- Increased speed of trimming and aligning



<input checked="" type="checkbox"/>	Genome Name:	JQ3756f.1
<input type="checkbox"/>	Isolate Name:	
<input checked="" type="checkbox"/>	Bioproject:	PD2710
<input type="checkbox"/>	US State:	
<input type="checkbox"/>	Taxonomy Number:	

Acknowledgements

- **Client:** Dr. Lenwood S. Heath -
Virginia Tech Computer Science
Department



References

- Resources: Python, NCBI, BLAST, MUSCLE
- <https://towardsdatascience.com/python-interactive-network-visualization-using-networkx-plotly-and-dash-e44749161ed7>
- <https://labarchives.kayako.com/Knowledgebase/Article/View/407>
- <https://imb.uq.edu.au/article/2020/04/difference-between-bacteria-and-viruses>
- [https://en.wikipedia.org/wiki/Pandas_\(software\)](https://en.wikipedia.org/wiki/Pandas_(software))
- <https://rapids.ai/plotly.html>