|  |  |
| --- | --- |
| Coding Files | Description |
| [virus\_read\_detection\_simulation.py](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Simulation/virus_read_detection_simulation.py) | Main simulation file. It takes arguments that describe the parameters of the simulation including the selected virus and host genomes, the quantity of each, the error rate, and the reference database. The script generates the reads and then calls a function described in virus\_read\_detection\_blastn\_search.py in a loop to BLAST the reads. |
| [virus\_read\_detection\_blastn\_search.py](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Simulation/virus_read_detection_blastn_search.py) | Describes functions to run BLAST and return the results. |
| [loop\_virus\_read\_detection\_simulation.py](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Simulation/loop_virus_read_detection_simulation.py) | An alternate main script that extends virus\_read\_detection\_simulation.py by taking an additional argument specifying the number of times to repeat the simulation. This script is only for use in simulations where a subset of the reads is mapped. For example, simulations can be run to obtain a distribution for the number of reads sequenced before obtaining the first virus sequence. |
| [dashApp\_fragmentation.py](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Dash%20Apps/dashApp_fragmentation.py) | Simulates ONT sequencing of a chosen reference genome and summarizes the results in terms of the number of reads, the range of read lengths, and a histogram of read lengths. |
| [Hypergeometric app.py](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Dash%20Apps/Hypergeometric%20app.py) | Models ONT sequencing using the hypergeometric probability distribution. Given a prespecified number of virus and background reads, the app plots the likelihood of sequencing a virus read as a function of the total number of reads sequenced. In addition, the app calculates the number of reads that need to be sequenced to achieve threshold levels of sensitivity. |
| [Minimum Read Length Analysis - Sim 2-3.py](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Data/Minimum%20Read%20Length%20Analysis%20-%20Sim%202-3.py) | Analyzes simulation results to identify how long reads need to be to accurately map to host and virus genomes. Generates the data file “Classification Accuracy by Read Length.csv”. |
| [Runtime Analysis.py](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Data/Runtime%20Analysis.py) | Analyzes simulation results **TO DOC**  Generates the data files:  Simulation runtime med + low concentrations.csv  Simulation\_1000x\_runtime.csv  Simulation 1000x Sensitivity by Reads Mapped.csv |
| [Figure\_Detection Time by Viral Concentration.Rmd](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Data/Figure_Detection%20Time%20by%20Viral%20Concentration.Rmd) | Generates a figure showing the time it takes to detect a virus in simulations with 3 different viral concentrations. |
| [Figure\_Classification Accuracy by Read Length.Rmd](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Data/Figure_Classification%20Accuracy%20by%20Read%20Length.Rmd) | Generates a figure showing classification accuracy by read length from the virus detection simulations. |
| [Figure\_1000x Simulation Runtime + Sensitivity.Rmd](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Data/Figure_1000x%20Simulation%20Runtime%20%2B%20Sensitivity.Rmd) | Generates figures showing sensitivity as a function of the number of reads mapped and read mapping rate as a function of read length. |

|  |  |
| --- | --- |
| Data Files | Description |
| [Classification Accuracy by Read Length.csv](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Data/Classification%20Accuracy%20by%20Read%20Length.csv) | A results file that describes the sensitivity of virus detection with varying read error rates and read lengths. |
| [Simulation 1000x Sensitivity by Reads Mapped.csv](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Data/Simulation%201000x%20Sensitivity%20by%20Reads%20Mapped.csv) | A results file that describes the proportion of simulations in which a virus was detected for varying numbers of reads mapped. |
| [Simulation runtime med + low concentrations.csv](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Data/Simulation%20runtime%20med%20%2B%20low%20concentrations.csv) | A results file that describes the runtime, number of hits, and total number of reads mapped during a looped virus detection simulation. |
| [Simulation\_1000x\_runtime.csv](https://github.com/A-Li14/TGT_Virus_Detection_Simulation/blob/main/Data/Simulation_1000x_runtime.csv) | A results file that describes the runtime, number of reads mapped, and average read length during a looped virus detection simulation. |