

Southern Ontario Wastewater Analysis Report

Art Poon, Devan Becker, Gopi Gugan and Erin Brintnell

2021-11-30

Background

- On November 23, 2021, a request for a new lineage designation was submitted to the CoV-lineages PANGO designation GitHub site, describing a new sub-lineage of B.1.1 sampled in Botswana, Hong Kong and South Africa with an excessive number of mutations.
- The request was reviewed and accepted on November 24, designating the new lineage as B.1.1.529.
- On November 27, the CoV-lineages team uploaded a ‘constellation’ file, describing a set of mutations for classifying genomes into the newly defined lineage.

Objectives

1. identify common mutations in lineage B.1.1.529 (Omicron) relative to the SARS-CoV-2 reference genome sequence (WH1);
2. determine which of these common mutations are unique to B.1.1.529, in comparison to all other defined lineages;
3. retrospectively screen all available wastewater sample data sets for the presence of these mutations.

Analysis of Omicron Mutations

Sequence data for available Omicron genomes was obtained (n=77).

Gaps in Omicron sequence data were removed and data was processed to generate a list of mutations, indels and sections of low sequence coverage for each genome.

The frequency of selected mutations in the Omicron genomes (n=77) and Pango lineages (n=220242) was calculated by counting the number of occurrences of each mutation. Frequency of mutations in Omicron and Pango lineages were compared and mutations present in >95% of Omicron genomes and <5% of all other Pango lineages were selected for further downstream analysis of wastewater data.

There were 111 mutations present in the Omicron genomes (n=77). Many mutations were present in other Pango lineages (**Figure 1**). Nonetheless, 29 mutations were present in >95% of Omicron genomes

and <5% of all other Pango lineages (**Table 1**) which were used for downstream analyses.

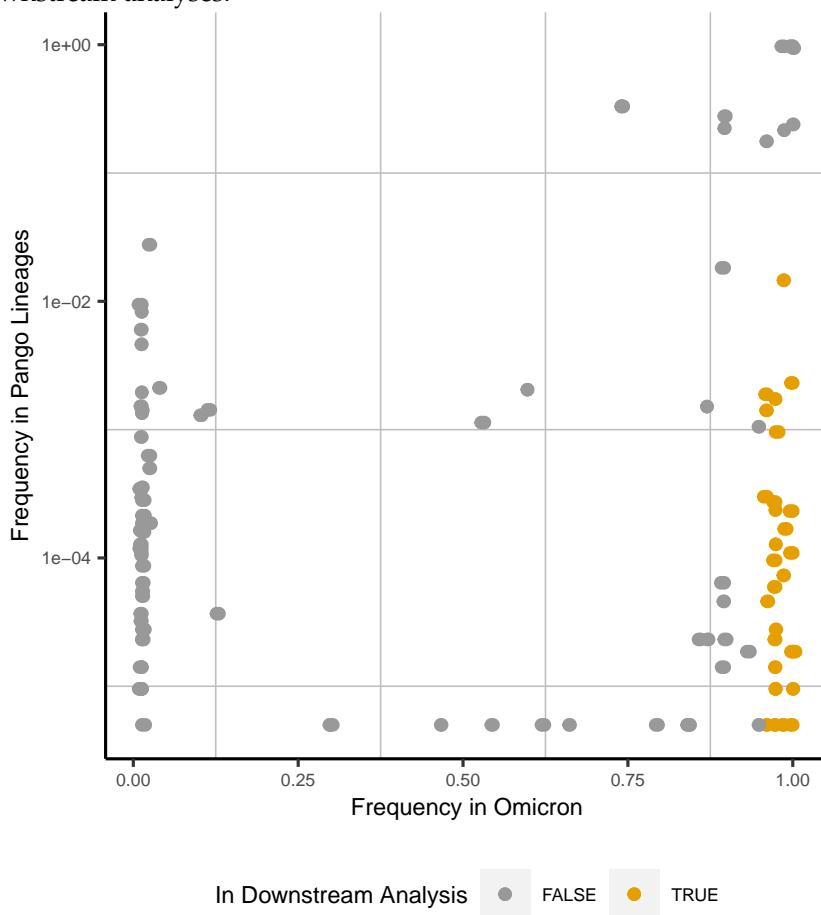


Figure 1 Illustration of 111 mutations found throughout 77 Omicron genomes. Mutations used in downstream analysis were present in >95% of Omicron genomes and <5% of all other Pango lineages.

Table 1 Mutations selected for downstream analysis in wastewater samples.

Mutation	Frequency in Omicron	Frequency in Pango Lineages
~8392A	1.000000	0.0000000
~18162G	1.000000	0.0000045
~2831G	1.000000	0.0000136
~10448A	1.000000	0.0000136
~13194C	1.000000	0.0001044
~11536G	1.000000	0.0002270
~15239T	1.000000	0.0023066
~5385G	0.987013	0.0000000
~23201A	0.987013	0.0000681
~23598G	0.987013	0.0001635
~23524T	0.987013	0.0145930
~26576G	0.974026	0.0000000
~22673T	0.974026	0.0000045
~24468A	0.974026	0.0000045
~27258C	0.974026	0.0000091
~22672C	0.974026	0.0000182
~22678C	0.974026	0.0000227
~22897A	0.974026	0.0000545
~26708A	0.974026	0.0000908
~22577A	0.974026	0.0001226
~27806T	0.974026	0.0002316
~22881G	0.974026	0.0002679
~26529G	0.974026	0.0009535
~24999T	0.974026	0.0017254
~24129A	0.961039	0.0000000
~23853A	0.961039	0.0000409
~23947T	0.961039	0.0002951
~26269T	0.961039	0.0014075
~22812T	0.961039	0.0018798

Analysis of Omicron in Ontario Wastewater Samples

Coverage of genomes located at the positions of selected Omicron mutations was collected from sequence coverage files. Mapped wastewater reads were then probed for selected Omicron mutations. Omicron mutations present in wastewater samples was then scaled by coverage at the site.

Many wastewater samples had very low coverage (**Figure 2**), however there was evidence of some Omicron mutations in wastewater

samples when scaled by coverage (**Figure 3**). Of note, aa:S:T547K, aa:S:G446S, aa:S:G339D, aa:orf1a:K856R were all present in >1% of genomes from samples processed by laboratories at the University of Guelph. Additionally, there was no evidence of any Omnicron mutations in samples processed at Western University of laboratories.

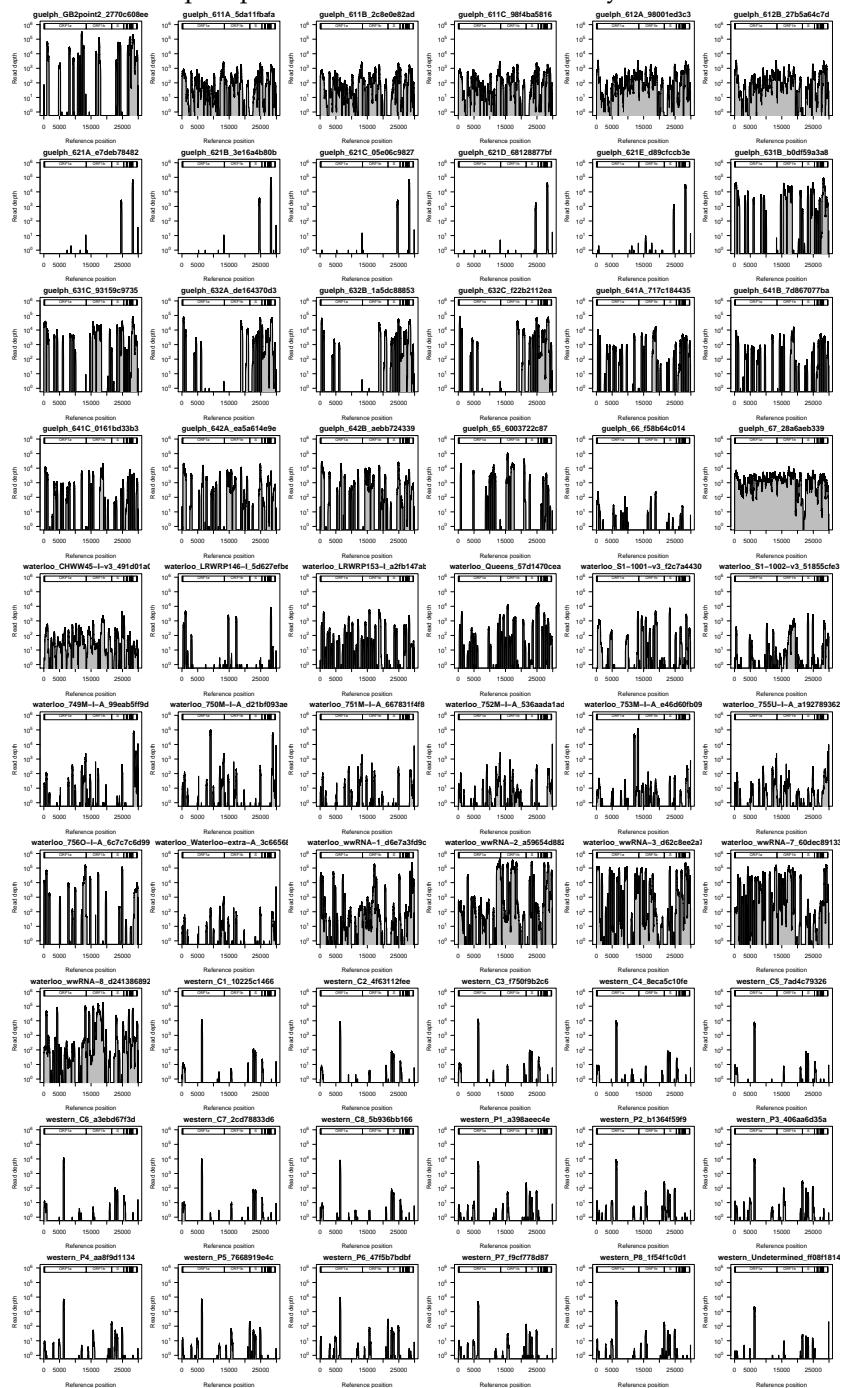


Figure 2 Coverage of wastewater samples across the SARS-CoV-2

genome.

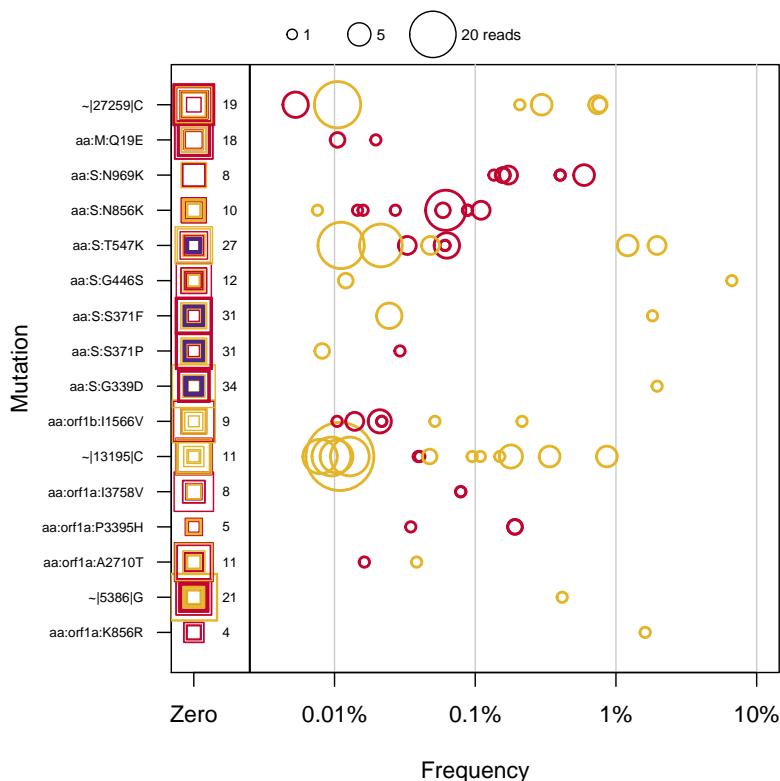


Figure 3 Illustration of the frequency of selected Omicron mutations in wastewater samples. The size of the bubble represents the number of reads containing that mutation in a particular sample. The colours are the official colours of Western, Guelph, and Waterloo, respectively, and represent the lab that sequenced the data (not the location of the sample - this data is currently unavailable to us). The squares represent mutations that were not present in the samples OR occurred in a location where the sample had zero coverage, so it would not be possible to tell whether these mutations were not present or merely not sequenced. The size of the square represents the log of the coverage (smallest squares have lowest coverage).

Concluding Remarks