

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, designated “Omicron”.
- 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.

```
## Warning: Transformation introduced infinite values in continuous y-axis
```

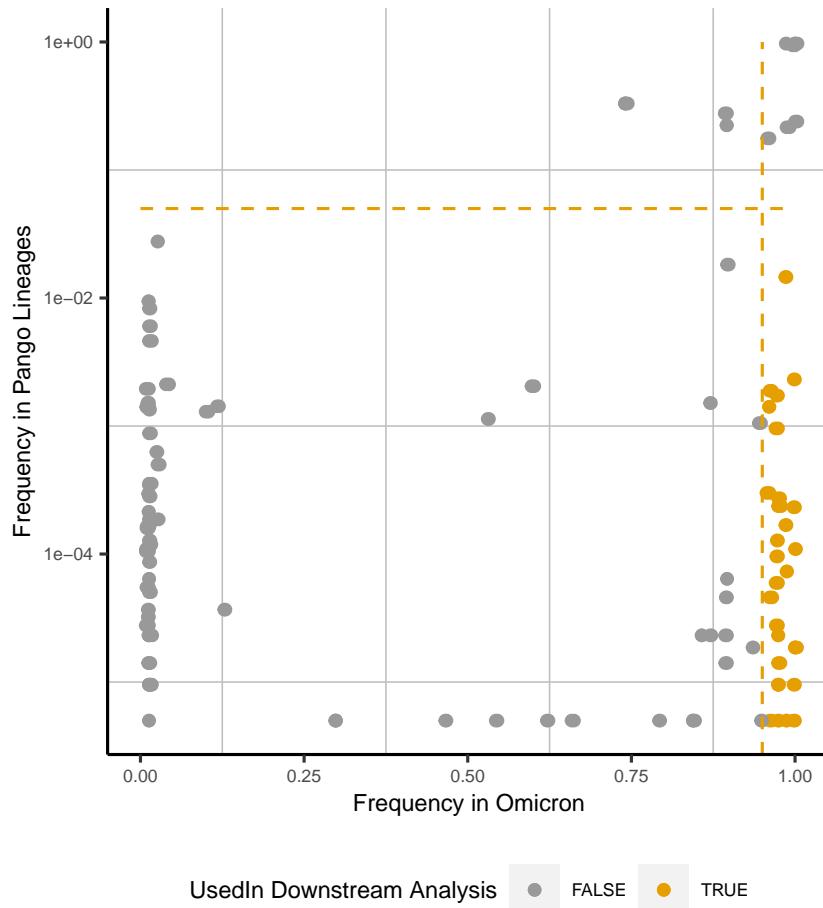


Figure 1 Illustration of 111 mutations found throughout 77 Omicron genomes. Mutations used in downstream analysis were present in >95% of Omicron genomes and <5% of all other Pango lineages. These numbers were chosen arbitrarily and will be explored in further studies.

Table 1 Mutations selected for downstream analysis in wastewater samples.

Mutation	Frequency in Omicron	Frequency in Pango lineages
~ 5386 G	0.987013	0.0000000
aa:orf1a:A2710T	1.000000	0.0000000
aa:S:N856K	0.961039	0.0000000
aa:M:Q19E	0.974026	0.0000000
aa:orf1b:I1566V	1.000000	0.0000045
aa:S:S371F	0.974026	0.0000045
aa:S:N969K	0.974026	0.0000045
~ 27259 C	0.974026	0.0000091
aa:orf1a:K856R	1.000000	0.0000136
aa:orf1a:P3395H	1.000000	0.0000136
aa:S:S371P	0.974026	0.0000182
aa:S:S373P	0.974026	0.0000227
aa:S:N764K	0.961039	0.0000409
aa:S:G446S	0.974026	0.0000545
aa:S:T547K	0.987013	0.0000681
aa:M:A63T	0.974026	0.0000908
~ 13195 C	1.000000	0.0001044
aa:S:G339D	0.974026	0.0001226
aa:S:N679K	0.987013	0.0001635
aa:orf1a:I3758V	1.000000	0.0002270
~ 27807 T	0.974026	0.0002316
aa:S:N440K	0.974026	0.0002679
aa:S:D796Y	0.961039	0.0002951
aa:M:D3G	0.974026	0.0009535
aa:E:T9I	0.961039	0.0014075
~ 25000 T	0.974026	0.0017254
aa:S:K417N	0.961039	0.0018798
~ 15240 T	1.000000	0.0023066
aa:S:H655Y	0.987013	0.0145930

Analysis of Omicron in Ontario Wastewater Samples

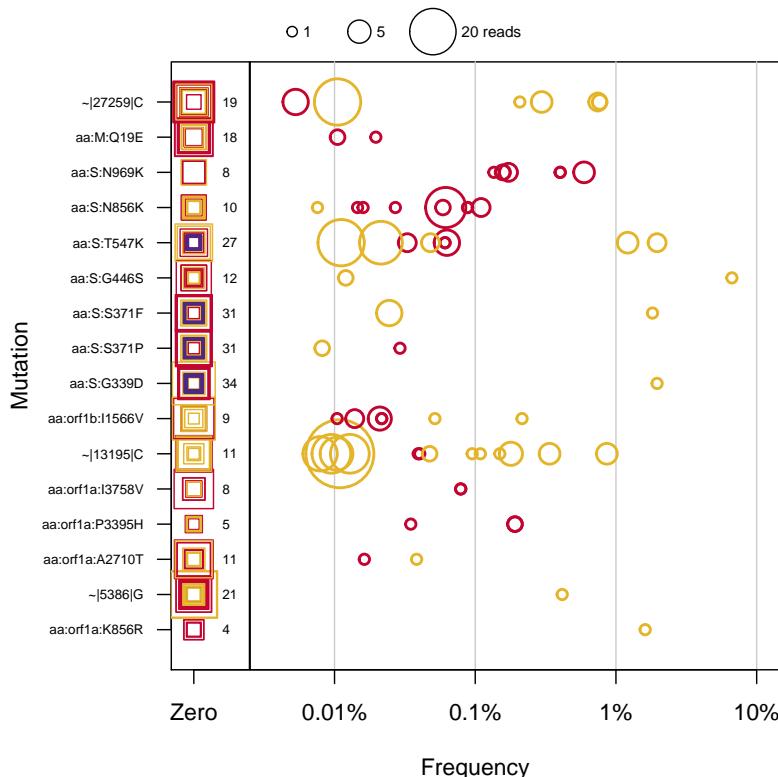
Coverage of genomes located at the positions of selected Omicron mutations was collected from sequence coverage files. Mapped wastewater reads were then probed for selected Omicron mutations. Omicron mutations present in wastewater samples was then scaled by coverage at the site.

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. Low coverage means that we cannot tell if the mutations were not present or were simply not sequenced.



Concluding Remarks

- We have identified mutations that are common in Omicron but rare in all other known variants, and (perhaps overenthusiastically) labelled these as “unique” mutations.
 - Note that these mutations are likely present in unsequenced genomes; there is a lot of unsampled diversity out there.
 - Even if a mutation is completely unique to Omicron compared to all known sequences, this does not imply that it’s presence is a guarantee that we’ve observed Omicron.
 - Even if there existed a mutation that was completely unique to Omicron and not even present in any of the unsampled genomes, our observation of that mutation is subject to sequencing error.
- These “unique” mutations appear in a small number of our samples, and in very small numbers.
 - This is evidence that Omicron may present, but it is far from being conclusive proof.
- To make conclusions from our data, we would like to observe multiple “unique” mutations in one sample, and have this confirmed by another lab sequencing the same sample.