

Masto Reader

[Share](#)

Jeff Gilchrist

jeffgilchrist@mstdn.science

1d

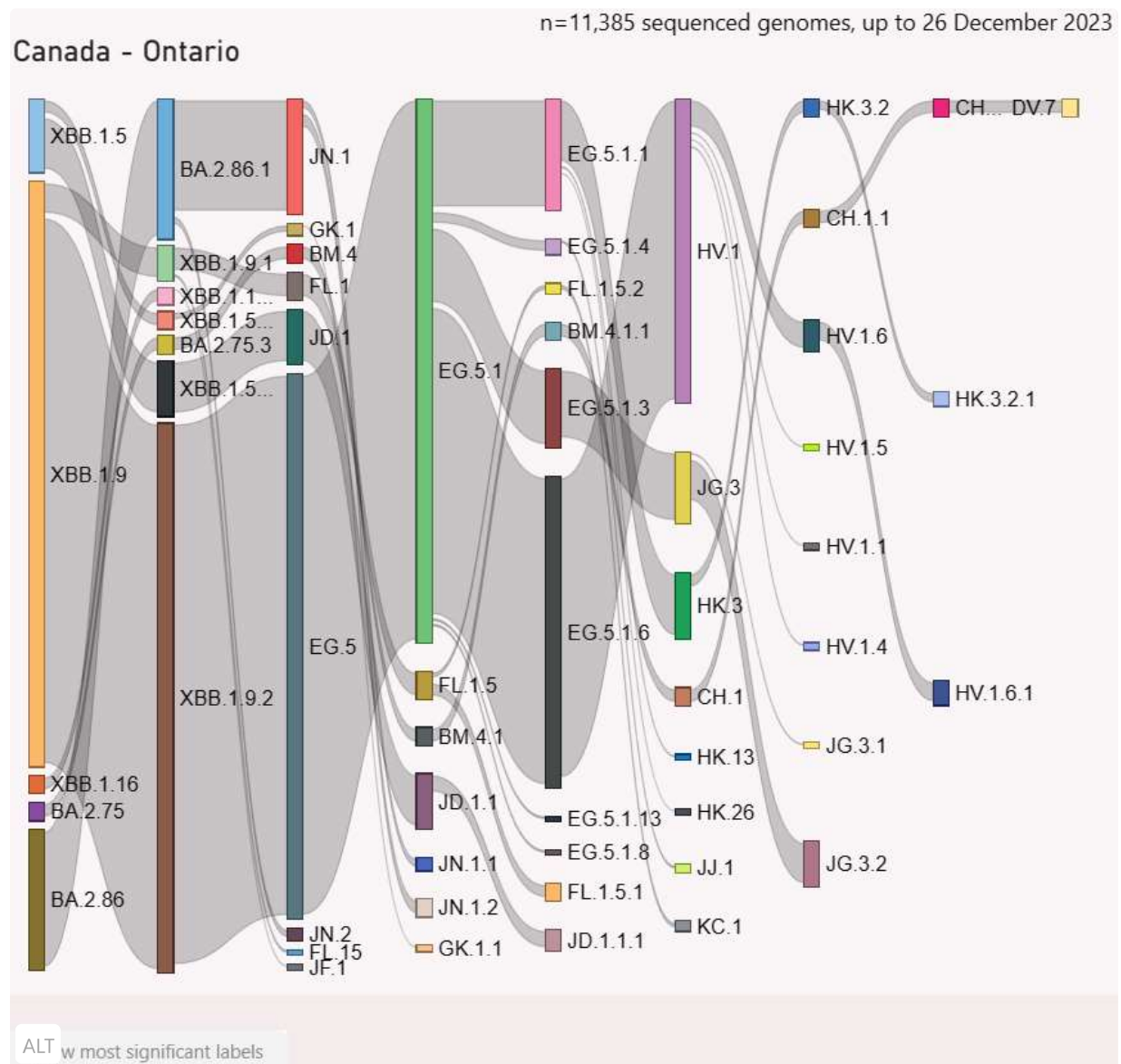
2 12 9

How To: Create a Sankey diagram of variants circulating in your area

This thread will walk you through the process of using @mike_honey_'s visualization tool to generate a Sankey diagram for your country, province or State. 📄 1/

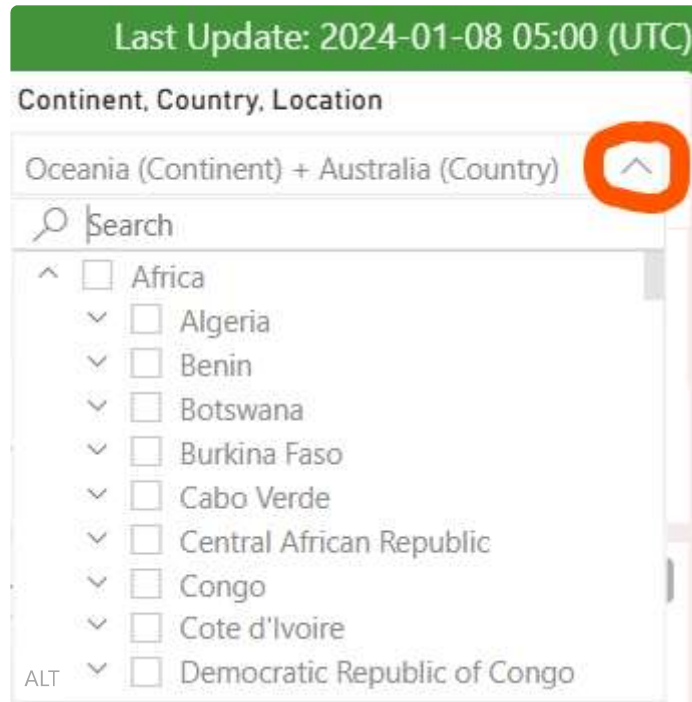
The one page web view is available here:

<https://threadreaderapp.com/thread/1745415266144600073.html>

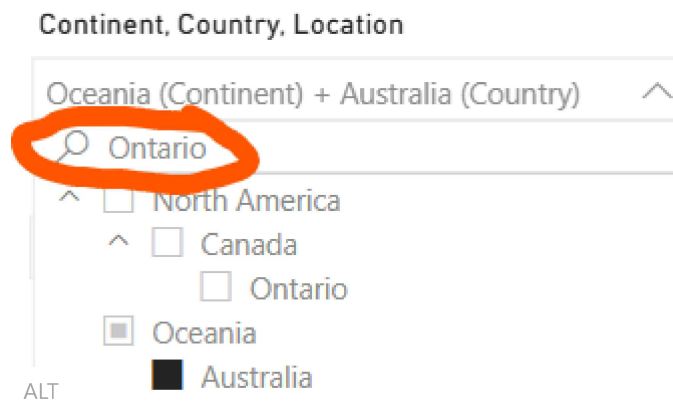


Step 1: Go to Mike's variant visualization tool (<https://app.powerbi.com/view?r=eyJrljoiNzE5YzczODItMDQzMS00M2EzLWJfNWYtMjg3OTY3NTNhZDM3liwidCI6ImRjMWYwNGY1LWMxZTUtdNDQyOS1hODEyLTU3OTNiZTQ1YmY5ZCIsImMiOjEwJmVwQ%3D%3D>). By default, Australia is chosen so you need to select your region of interest.

Step 2: Click on the V shape to the right of the current Continent, County, Location which by default is "Oceania + Australia" and you will see a drop-down list of continents and countries. Scroll down until you find your continent and then country of interest. 2/



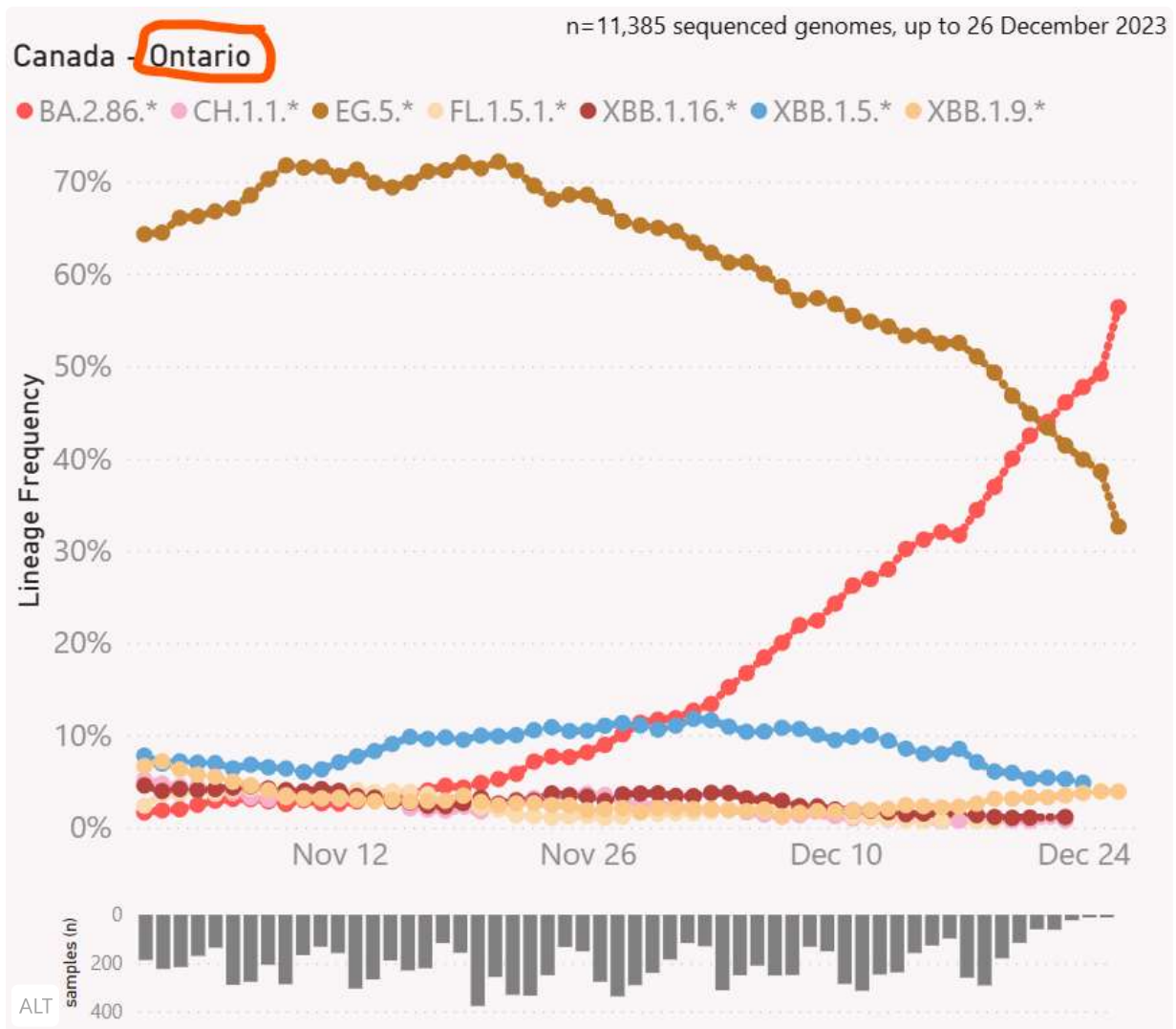
Alternatively, type in the name of the country, province, or State that you are looking for in the search field to quickly narrow down the options. 3/



Step 3: In this example we are going to use the province of Ontario in Canada so you would click on the white box beside Ontario as shown in the image. 4/



After a few seconds the tool will load the sequences from your newly selected location, the default graph will be updated and now display the top 7 variant families for your location (see Date selector box for exact range). 5/

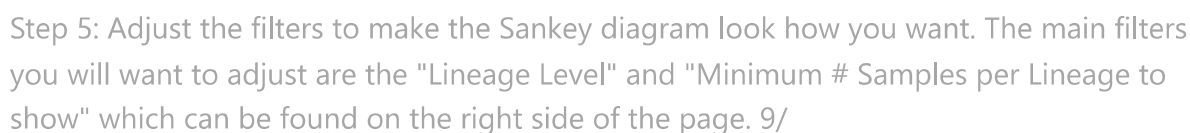
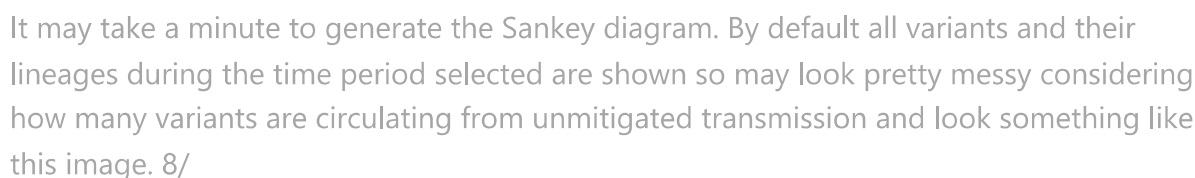


In this example the BA.2.86* family (the * means all variants of BA.2.86 which also includes any children of that variant such as JN.1) currently makes up about 56% of sequences in Ontario.

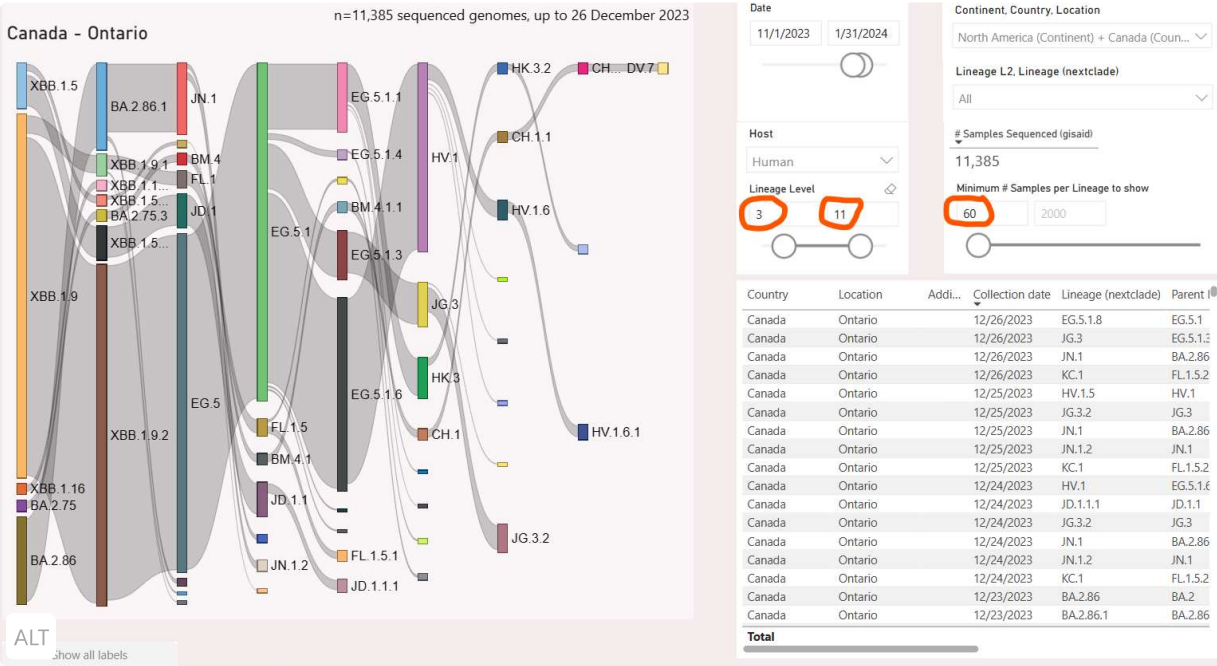
Step 4: Switch to the Sankey diagram by clicking on the page number at the bottom of the web page as shown. 6/



Scroll down until you see "gisaid (nextclade) - sankey" in the page listing and click on that option as shown. 7/

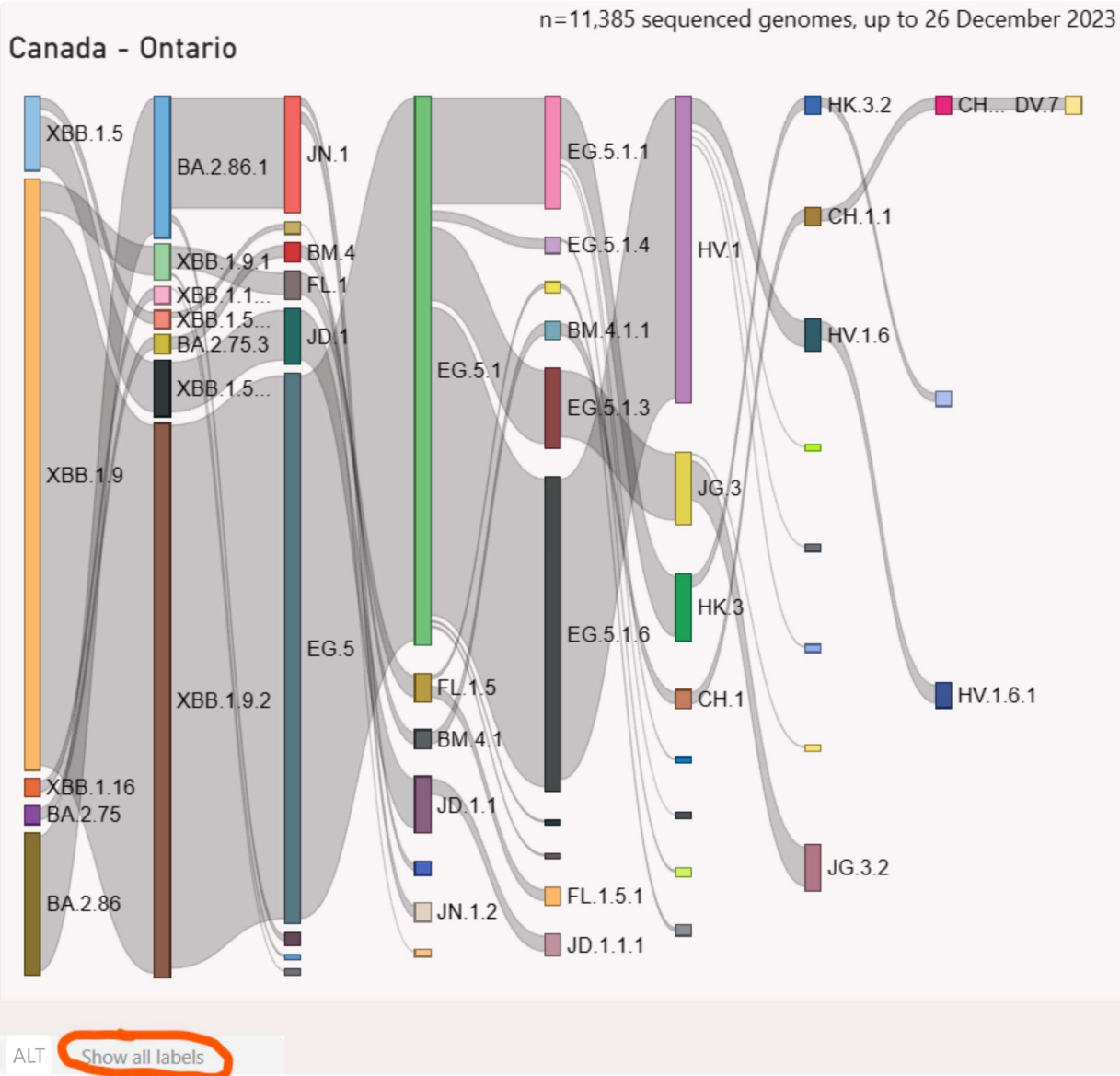


If you set the "Lineage Level" to 3 and 11 and the "Minimum # Samples per Lineage to show" to 60 it will make the Sankey diagram much less complex. The first few parent lineages on the left will be cut out along with a couple of child lineages on the right while only showing variants that have at least 60 samples that have been recorded. 10/

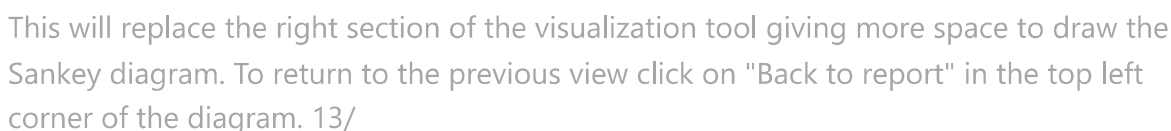
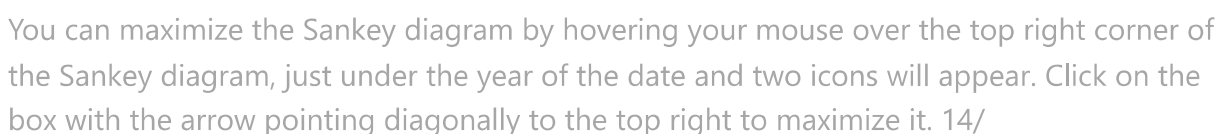


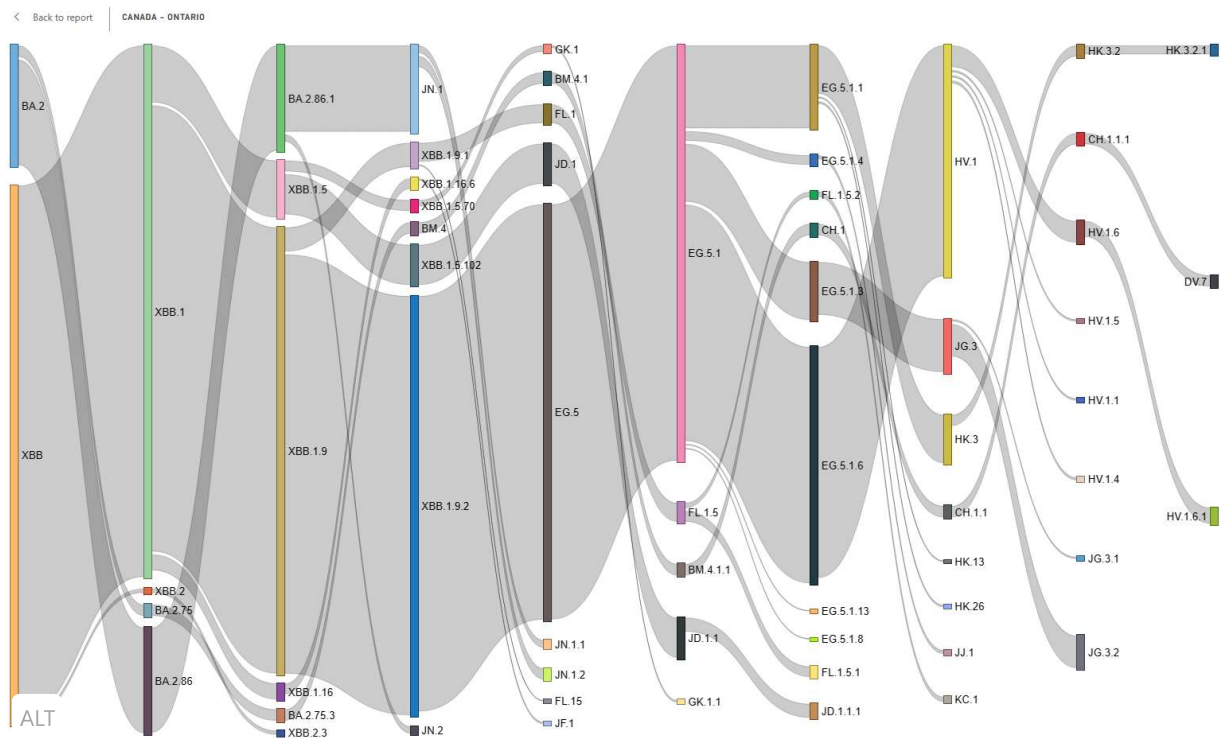
This may take a minute to process as it has to recalculate the entire diagram but you now have a less busy diagram. When the Lineage Level was 1, the Sankey showed BA.2 and XBB on the leftmost column, but now skips ahead two generations and starts with XBB.1.5, XBB.1.9, XBB.1.16, BA.2.75, and BA.2.86. 11/

Step 6: You will notice that most of the variants are labeled but some are not. By default the diagram only labels the most significant variants. If you want to see *all* labels, simply click on the "Show all labels" button in the bottom left corner as shown. 12/



Now all the labels will be shown. Sometimes even a filtered Sankey may seem a bit squished. 13/

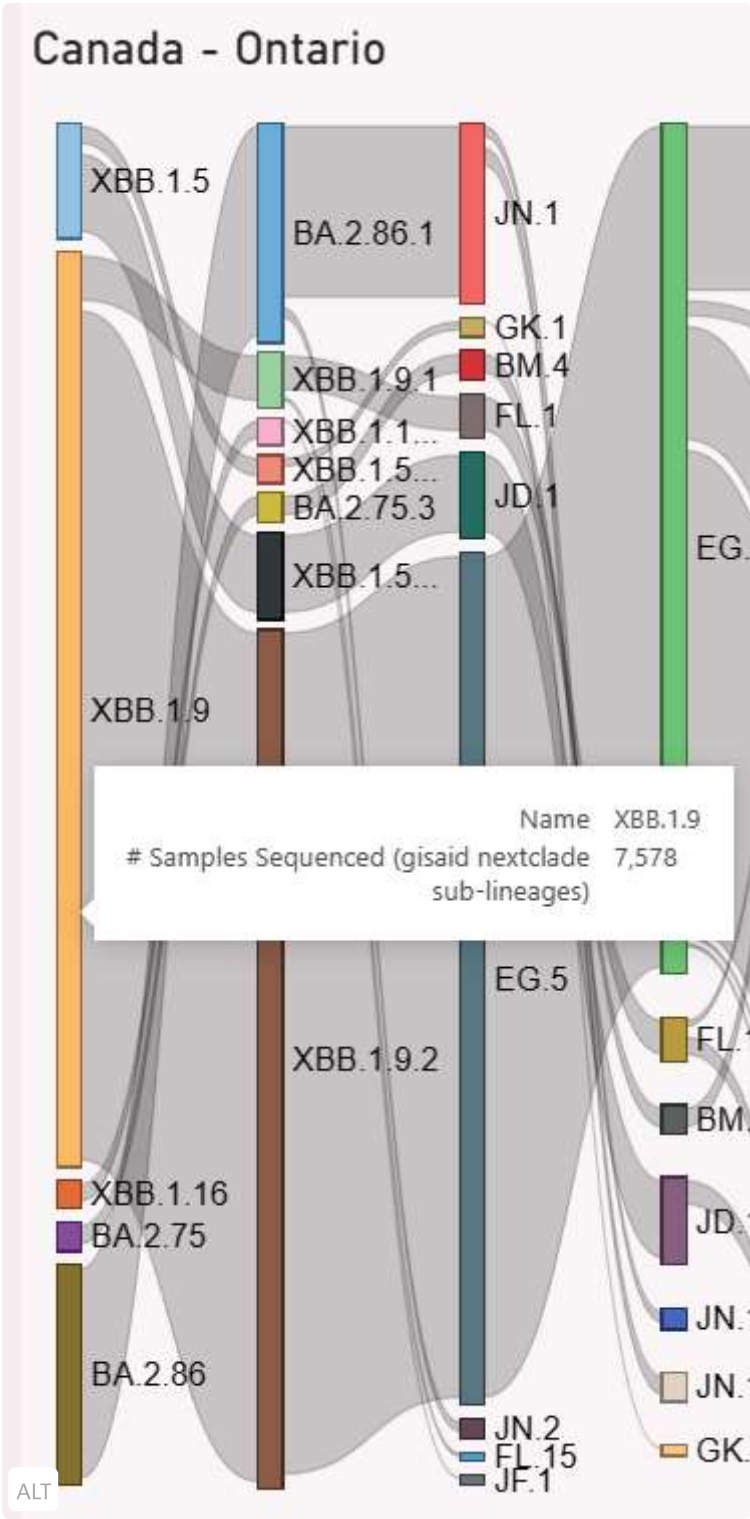




Step 7: Enjoy your filtered Sankey diagram.

Now that you have your filtered Sankey diagram, how do you interpret it? The height of each bar represents the relative difference in how many sequences there are for each variant in the time period that was selected.

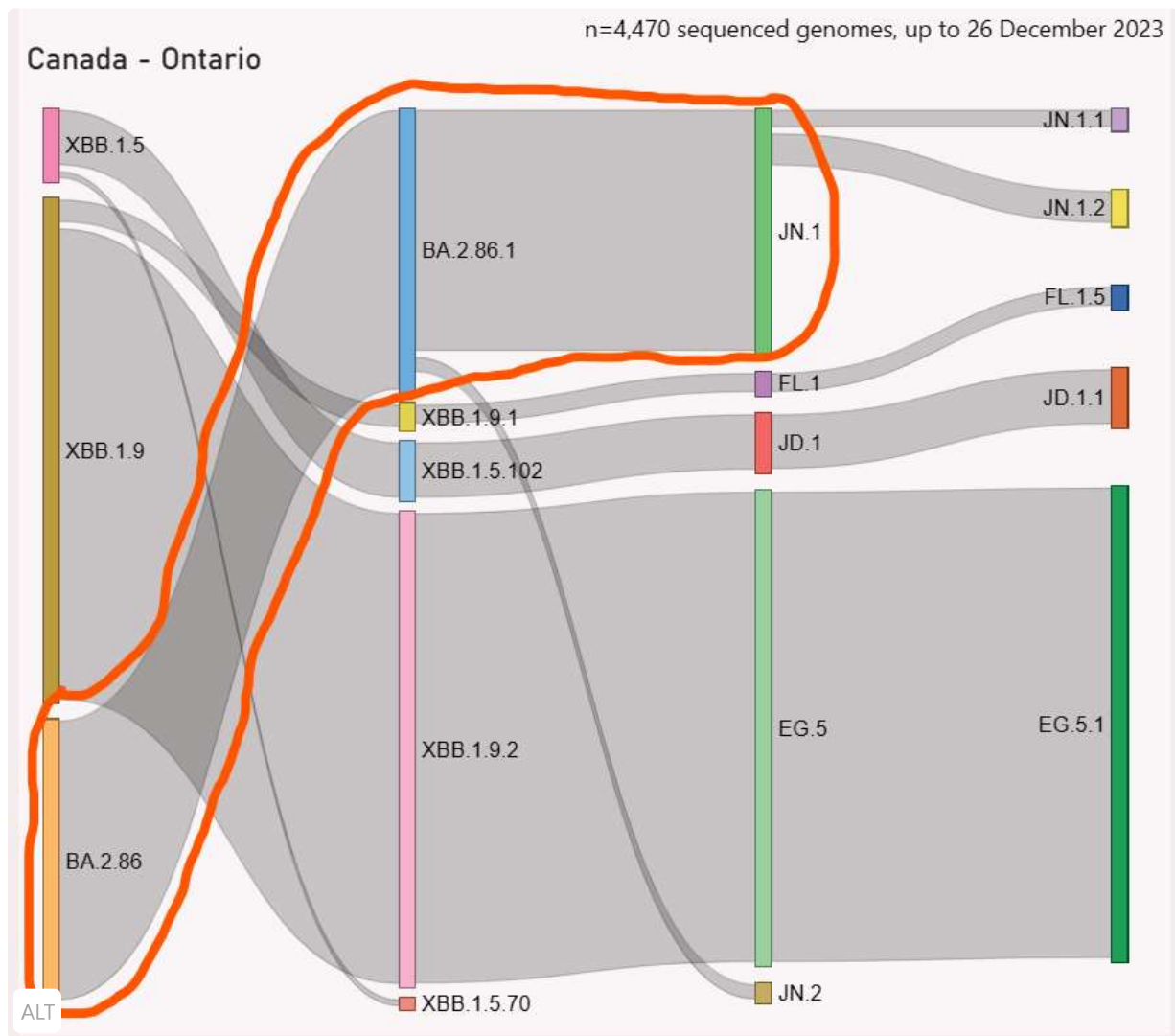
The longest bar in this example is XBB.1.9 "Hyperion" in the first column. Hovering over that bar shows there are 7,578 sequences found of XBB.1.9 including all of its descendants. 14/



While the original XBB.1.9 "Hyperion" variant itself is no longer circulating in Ontario, you can quickly see that the most common variants sequenced over the past two months in Ontario were based on the XBB.1.9 variant. Looking in column 4 you can see that most of those descendants come from the EG.5.1 "Eris" clan of variants. 15/

Remember that the Sankey diagram shows all the variants sequenced (except those filtered out) over the entire time period (the history of variants in the region) but does not identify the current most prevalent variant. To see the current most prevalent variants, you can use pages 1 and 2 of the tool using this guide (<https://mstdn.science/@jeffgilchrist/110203494274348811>). 16/

The Sankey diagram also helps you trace the lineage of a variant. Changing the filter to set the max lineage level to 6 and the date range from Dec. 1, 2023 to present simplifies the diagram even further. If you look at the current most prevalent variant JN.1, you can trace its lineage back to BA.2.86.1 and then BA.2.86 to learn where this variant came from. 17/



Now that you know how to use @mike_honey_'s visualization tool, you can generate Sankey diagrams for almost anywhere in the world that still does COVID-19 genomic sequencing, which unfortunately has become less and less over time.

You can find code for his tool on GitHub (<https://github.com/Mike-Honey/covid-19-genomes#readme>). 18/

Mike-Honey/covid-19-genomes

COVID-19 topic of genomic sequencing

0 Issues 43 Stars 0 Forks

GitHub - Mike-Honey/covid-19-genomes: Projects on COVID-19 topic of ge...

Projects on COVID-19 topic of genomic sequencing - mostly DataViz - GitHub - ...

github.com