Thinking about viral mutation & classification through Visualization

April 9, 2024

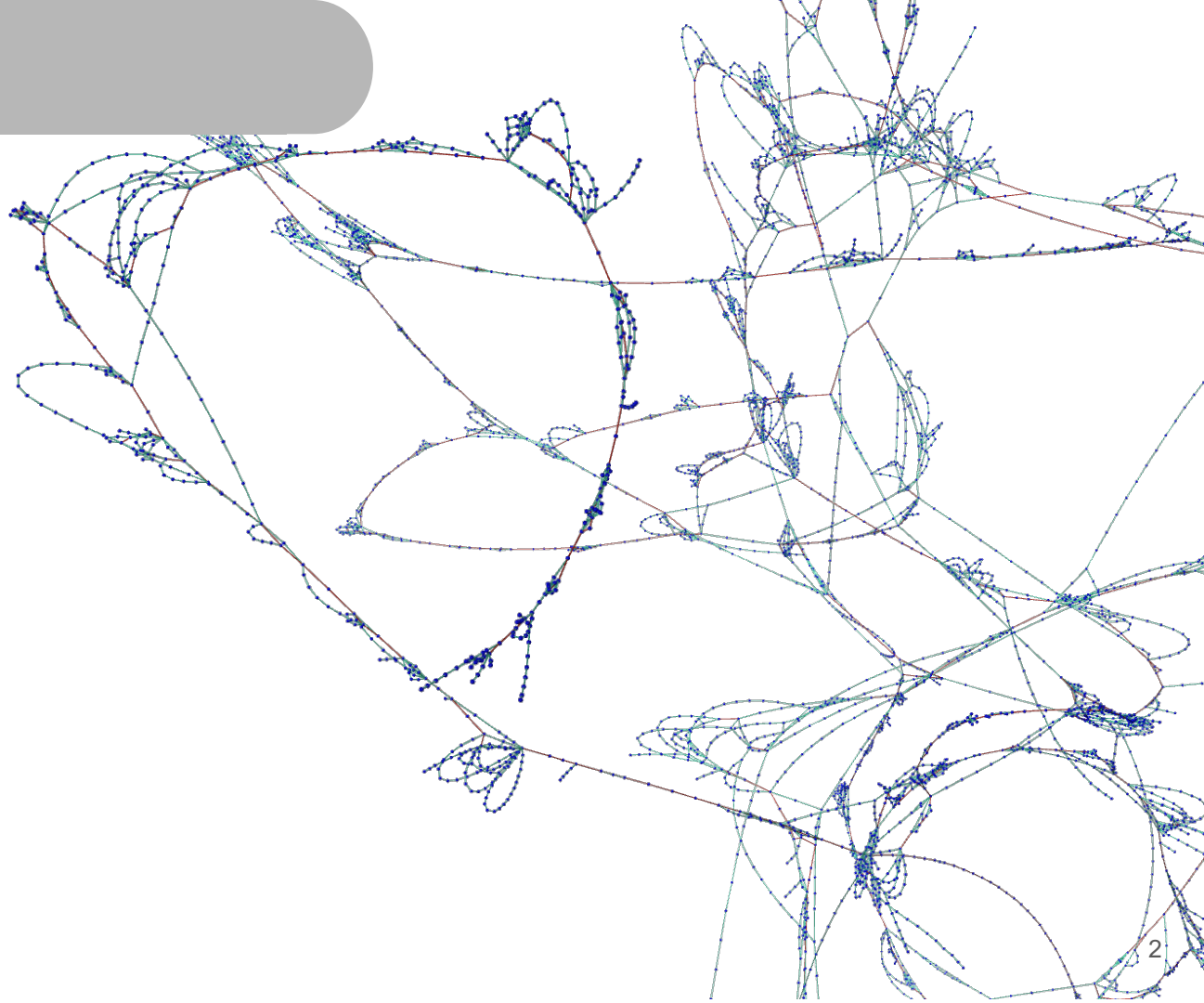
Andrew Warren, UVA, BV-BRC

Key Components

Visualization tools that support working with sub-species evolutionary lineages.

Ingredients
- Scales of classification
- Applications of scales
- Examples of tools & analysis
- Food for thought

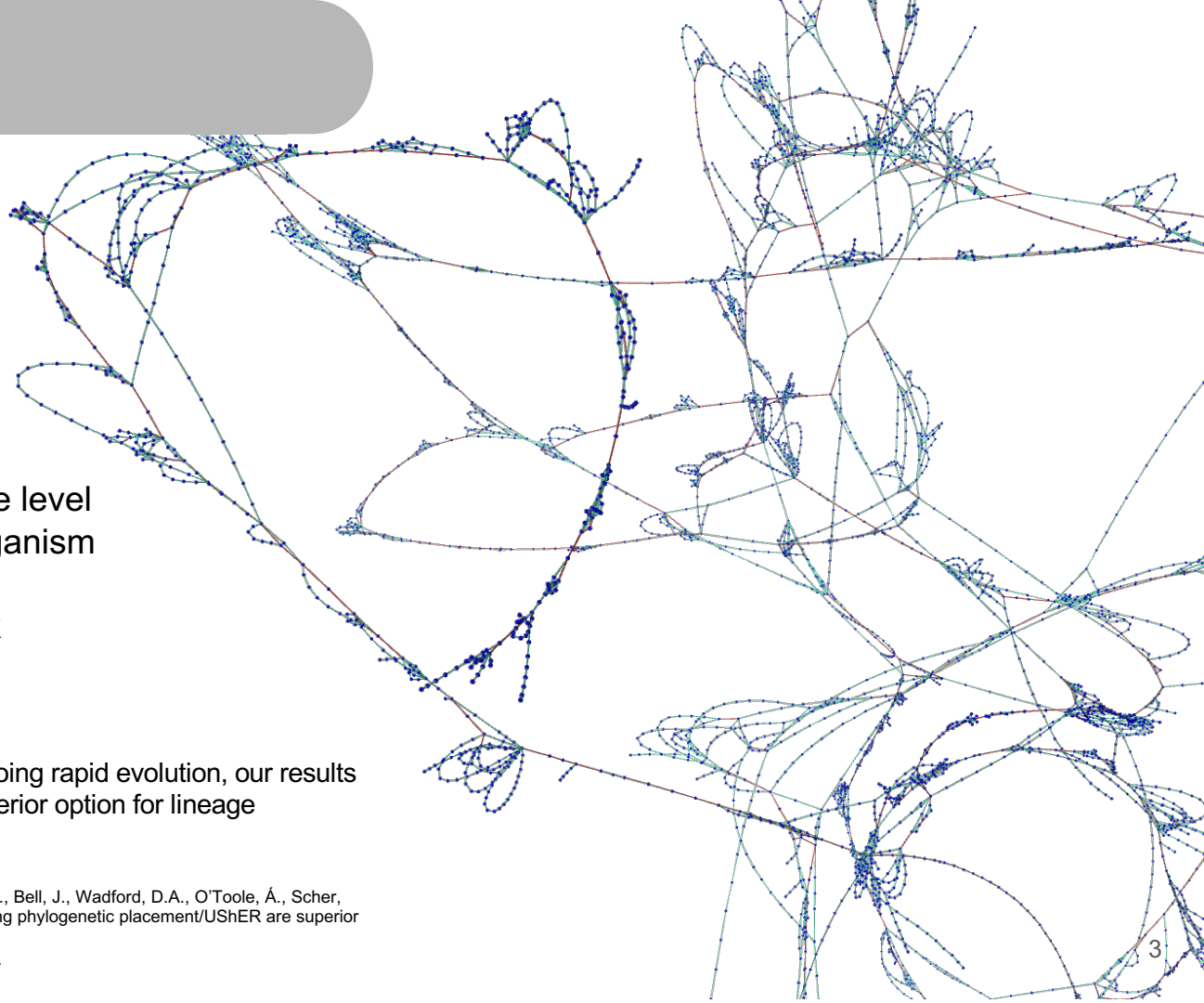
Dimensions
- Accessibility
- Questions answered
- InfoVis Principles
- Interactivity
- Availability of data



Usage

Applications:
- -Communication
- -Analysis
- -Quality Control

Tenant:
- Analytical abilities depend on the level of sequencing available to the organism type
- Complex questions require work

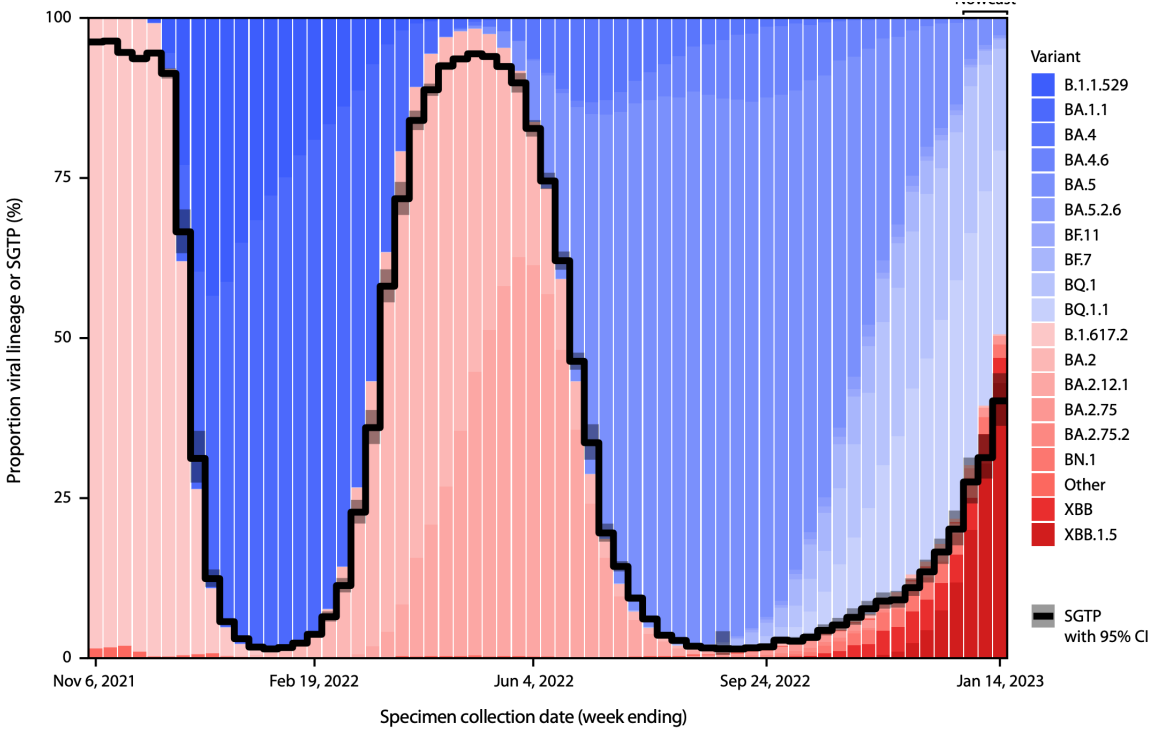


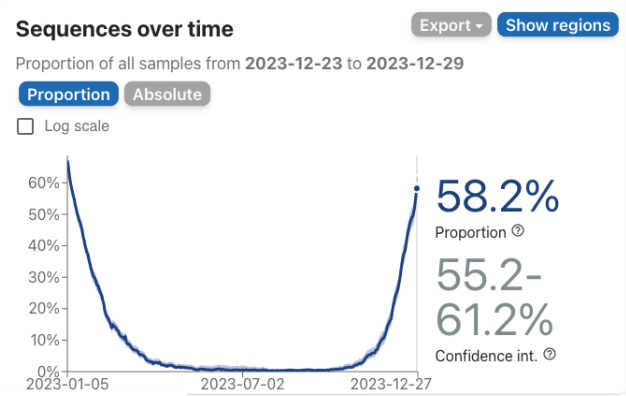
"Thus, for newly emerging pathogens undergoing rapid evolution, our results suggest that phylogenetic placement is a superior option for lineage assignment than machine-learning methods."

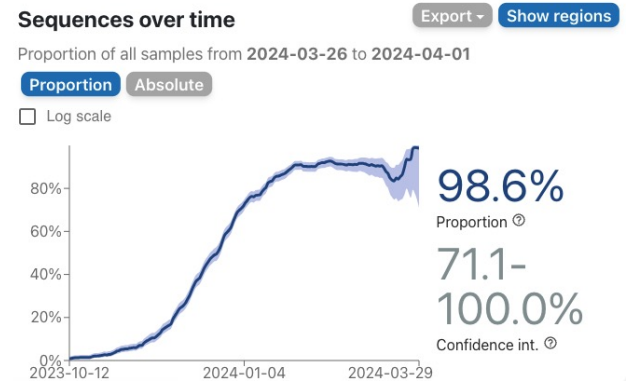
de Bernardi Schneider, A., Su, M., Hinrichs, A.S., Wang, J., Amin, H., Bell, J., Wadford, D.A., O'Toole, Á., Scher, E., Perry, M.D., et al. (2024). SARS-CoV-2 lineage assignments using phylogenetic placement/UShER are superior to pangoLEARN machine-learning method. Virus Evolution *10*, vead085. [10.1093/ve/vead085](https://doi.org/10.1093/ve/vead085).

High level markers

SGTF: Spike Gene Target Failure
!SGTF = SGTP





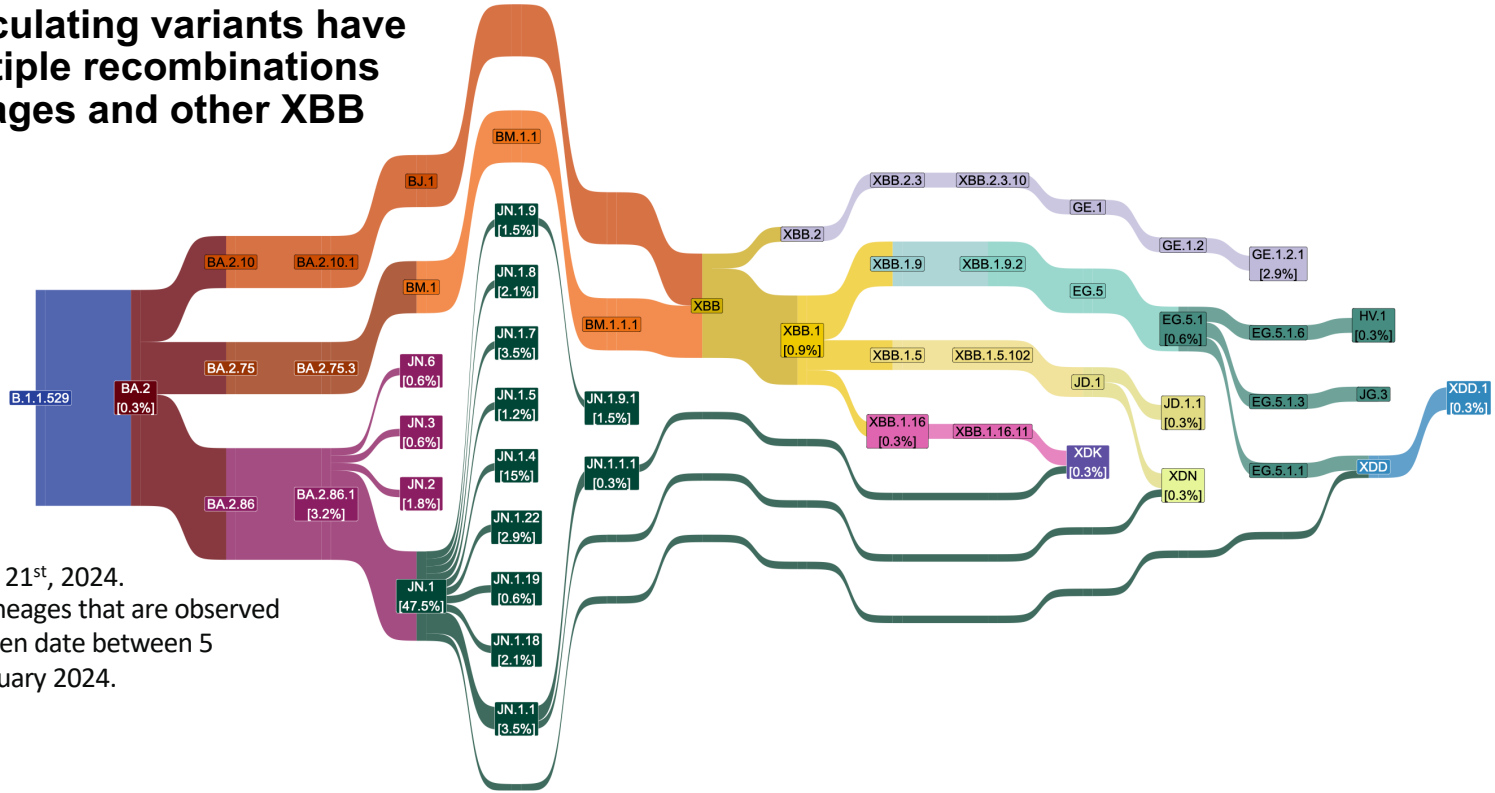


Scobie, H.M., Ali, A.R., Shirk, P., Smith, Z.R., Paul, P., Paden, C.R., Hassell, N., Zheng, X., Lambrou, A.S., Kondor, R., et al. (2023). Spike Gene Target Amplification in a Diagnostic Assay as a Marker for Public Health Monitoring of Emerging SARS-CoV-2 Variants — United States, November 2021–January 2023. MMWR Morb. Mortal. Wkly. Rep. *72*, 125–127. 10.15585/mmwr.mm7205e2.

https://cov-spectrum.org/explore/United%20States/AllSamples/Y2023/variants?variantQuery=S%3A69-+%7C+S%3A70-&

Namespace at Location

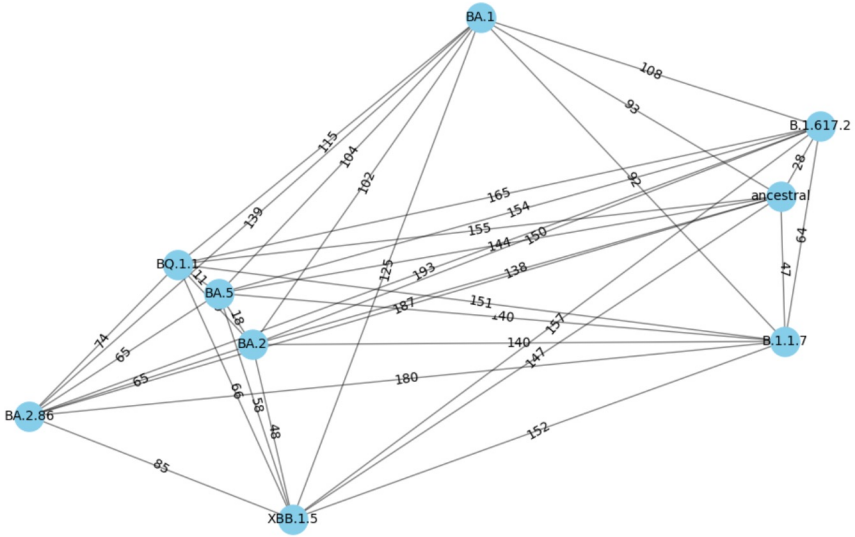
A variety of co-circulating variants have emerged with multiple recombinations between JN.1 lineages and other XBB lineages

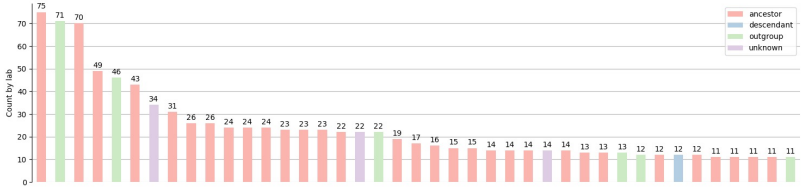


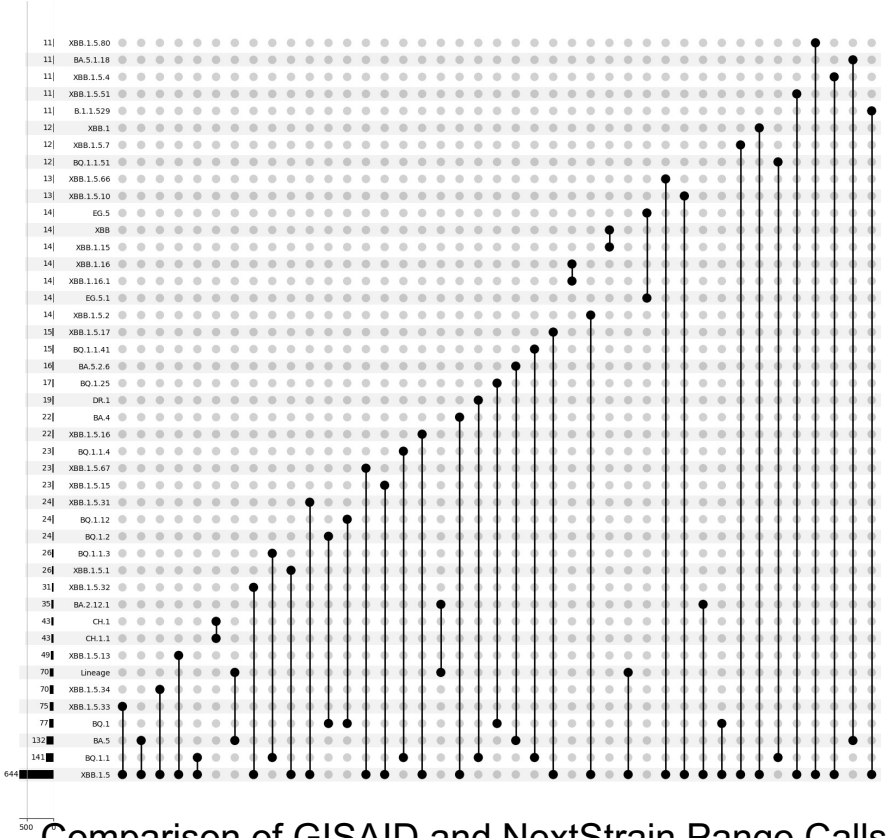
Data shown as of February 21st, 2024.
Proportions are given for lineages that are observed in sequences with a specimen date between 5 February 2024 and 16 February 2024.

https://www.gov.uk/government/publications/sars-cov-2-genome-sequence-prevalence-and-growth-rate/sars-cov-2-genome-sequence-prevalence-and-growth-rate-update-21-february-2024

Namespace Creation







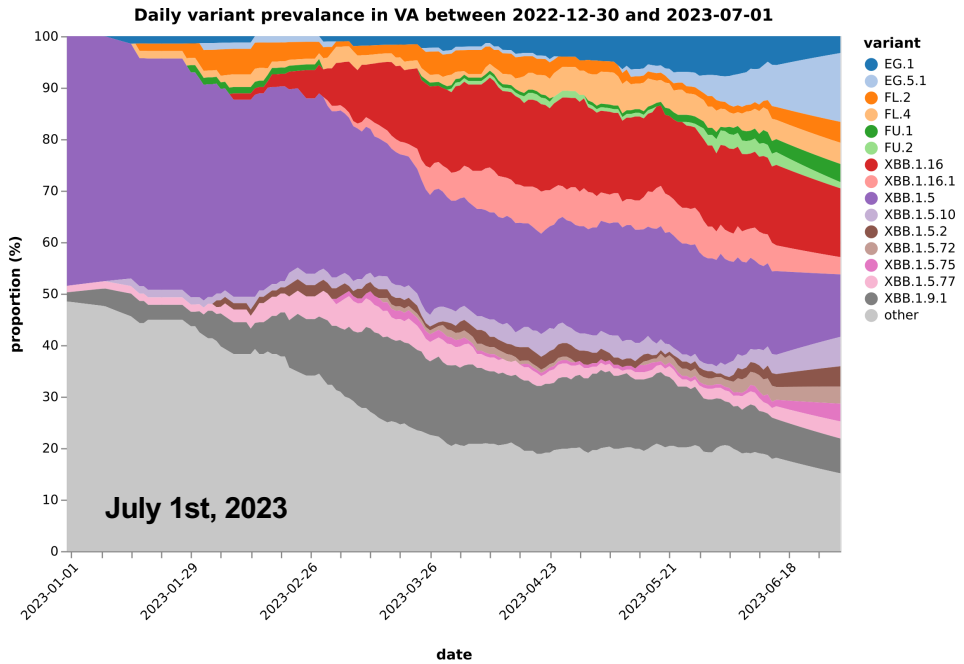
Comparison of GISAID and NextStrain Pango Calls

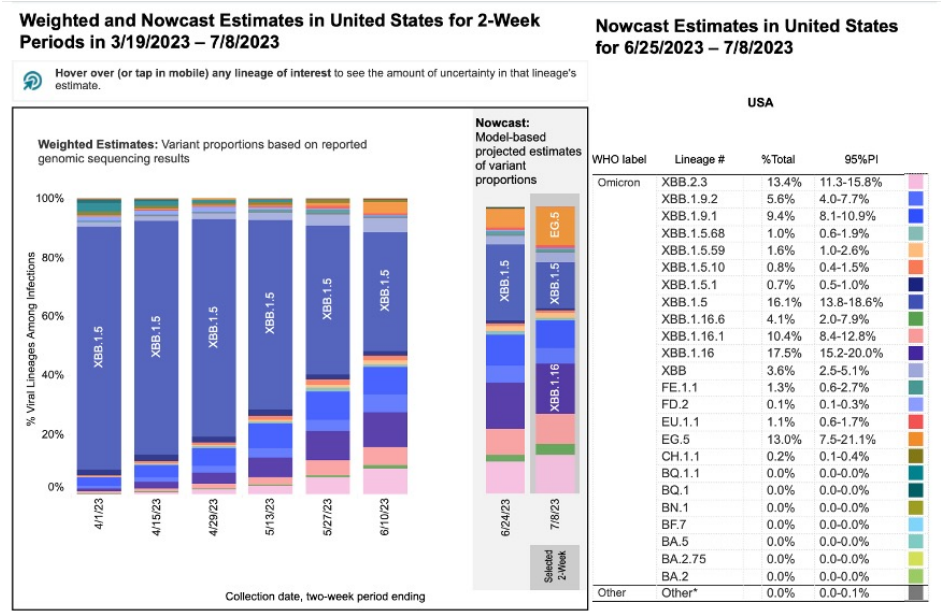
Acknowledgements

We gratefully acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based.

Spatiotemporal Analysis Frame

Parameterization of Optimal Variant List



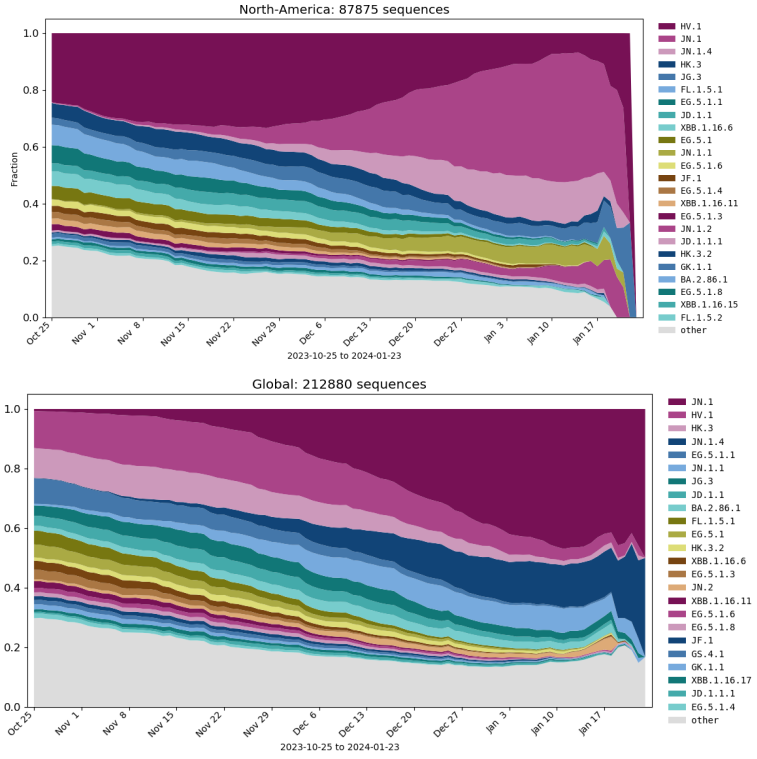


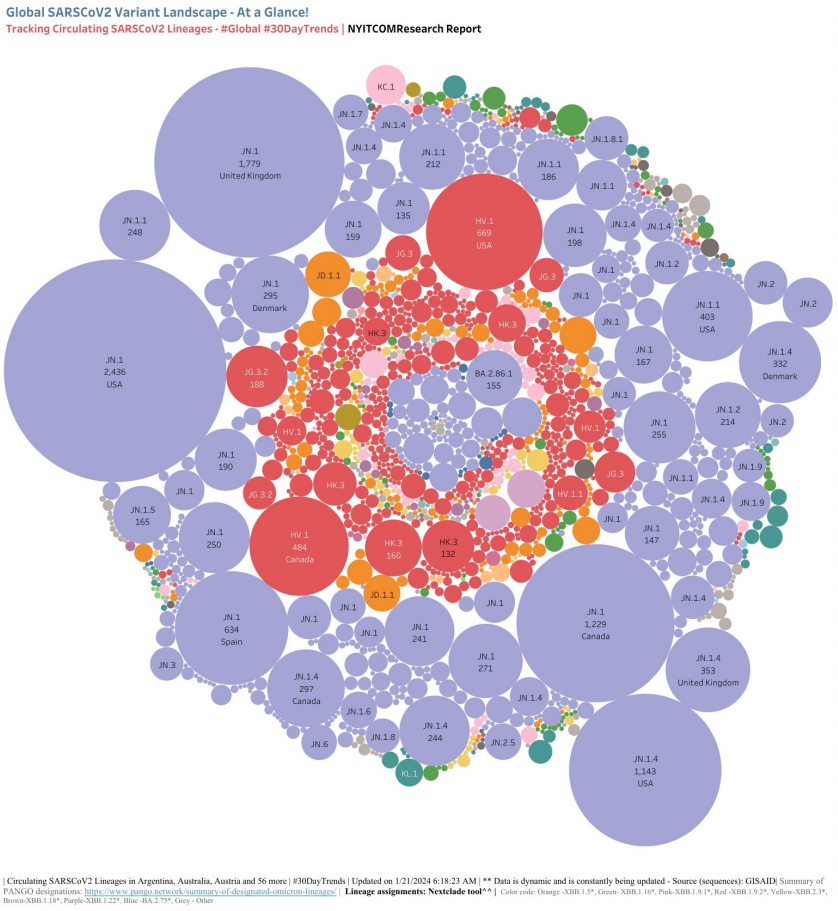
WHO label	Lineage #	%Total	95%PI
Omicron	XBB.2.3	13.4%	11.3-15.8%
	XBB.1.9.2	5.6%	4.0-7.7%
	XBB.1.9.1	9.4%	8.1-10.9%
	XBB.1.5.68	1.0%	0.6-1.9%
	XBB.1.5.59	1.6%	1.0-2.6%
	XBB.1.5.10	0.8%	0.4-1.5%
	XBB.1.5.1	0.7%	0.5-1.0%
	XBB.1.5	16.1%	13.8-18.6%
	XBB.1.16.6	4.1%	2.0-7.9%
	XBB.1.16.1	10.4%	8.4-12.8%
	XBB.1.16	17.5%	15.2-20.0%
	XBB	3.6%	2.5-5.1%
	FE.1.1	1.3%	0.6-2.7%
	FD.2	0.1%	0.1-0.3%
	EU.1.1	1.1%	0.6-1.7%
	EG.5	13.0%	7.5-21.1%
	CH.1.1	0.2%	0.1-0.4%
	BQ.1.1	0.0%	0.0-0.0%
	BQ.1	0.0%	0.0-0.0%
	BN.1	0.0%	0.0-0.0%
	BF.7	0.0%	0.0-0.0%
	BA.5	0.0%	0.0-0.0%
	BA.2.75	0.0%	0.0-0.0%
	BA.2	0.0%	0.0-0.0%
Other	Other*	0.0%	0.0-0.1%

Right Sizing Variant Prevalence Analysis Frames
Warren et al. CSTE 2024 Accepted

https://covid.cdc.gov/covid-data-tracker/#variant-proportions

Spatiotemporal Analysis Frame



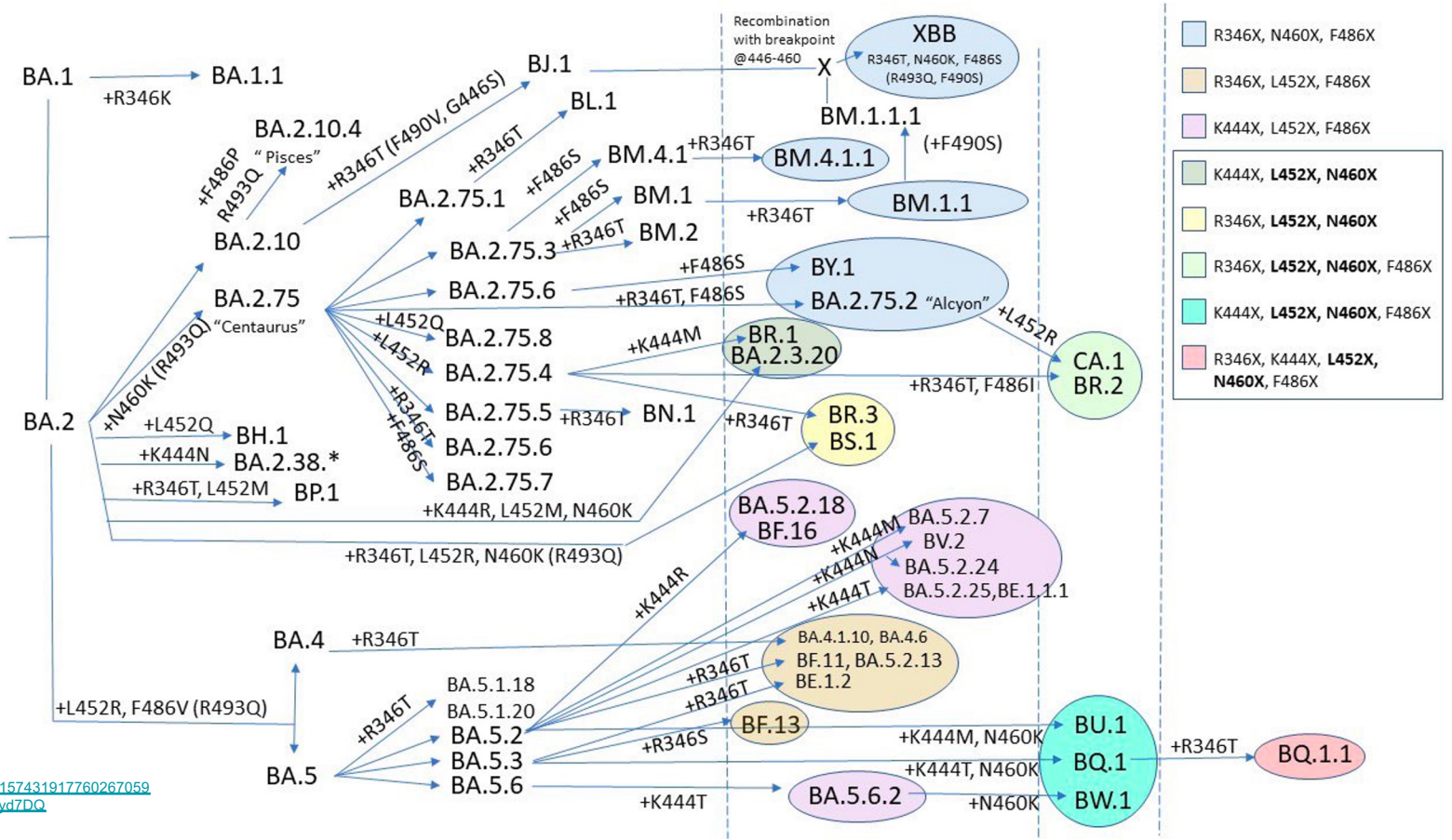


Pangolin Li..	% of Total	Count
JN.1	38.8%	9,270
JN.1.4	13.4%	3,198
JN.1.1	7.1%	1,690
HV.1	5.9%	1,417
HK.3	2.6%	624
BA.2.86.1	2.6%	620
JG.3	1.5%	361
JN.1.2	1.5%	359
JD.1.1	1.4%	341
JN.2	1.4%	334
JN.1.5	1.1%	260
JG.3.2	1.0%	248
JN.1.9	0.8%	197
JN.1.8	0.8%	194
HK.3.2	0.8%	181
JD.1.1.1	0.6%	143
GE.1.2.1	0.6%	139
XDD	0.6%	139
JN.1.8.1	0.5%	129
JN.3	0.5%	128
JN.1.7	0.5%	125
HV.1.1	0.5%	110
JN.1.6	0.4%	103
KC.1	0.4%	98
EG.5.1.1	0.4%	91
XDA	0.3%	79
EG.5.1.6	0.3%	78
JN.2.5	0.3%	76
JN.6	0.3%	75
JN.1.1.1	0.3%	67
XBB.1.16.6	0.3%	66
GS.4.1	0.3%	63
FL.1.5.1	0.3%	60
FL.15.1.1	0.3%	60
HK.20.1	0.2%	59
JF.1.1	0.2%	58
JG.3.1	0.2%	56
KL.1	0.2%	55
JN.4	0.2%	54
EG.5.1.8	0.2%	51
BA.2.86	0.2%	49
JD.1.1.8	0.2%	49
XDK	0.2%	48
JN.10	0.2%	46
XBB.1.16.11	0.2%	45
JN.1.3	0.2%	44
JR.1.1	0.2%	44
HK.26	0.2%	42
EG.5.1	0.2%	40
GK.1.1	0.2%	40
EG.5.1.3	0.2%	39
HK.3.1	0.2%	38
FW.1.1.1	0.2%	36
GK.1.1.1	0.2%	36
HK.27.1	0.1%	35
JC.5.1	0.1%	35
JN.1.11	0.1%	35
BA.2.86.3	0.1%	34
HK.13	0.1%	34
XBB.1.16.15	0.1%	34
JD.1.1.7	0.1%	32
JF.1	0.1%	32
XBB.1.5	0.1%	27
HV.1.4	0.1%	25
JN.2.1	0.1%	24

JN.1.4 ORF1a:T170I signature

https://cov.lanl.gov/components/sequence/COV/sparks.comp
https://twitter.com/RailabN/status/1748226226915787153

Convergent Evolution & Immune Escape

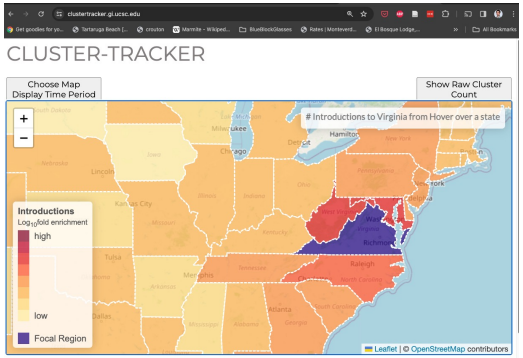


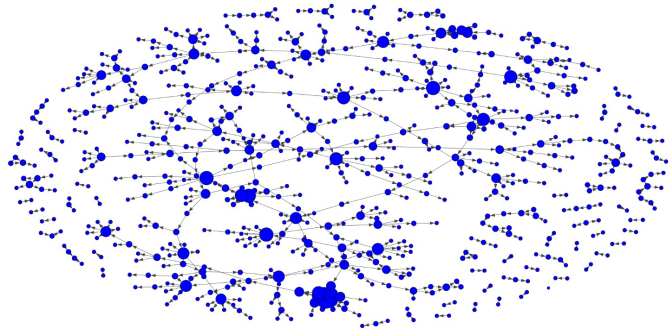
https://twitter.com/dfocosi/status/1574319177602670592?s=12&t=myre59DuQ1H31TI_-yd7DQ

Imputed Clusters / Cascades

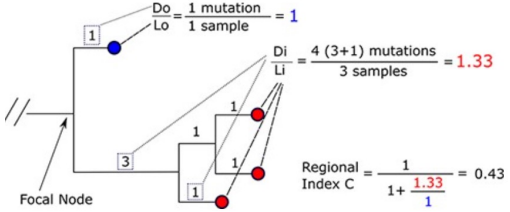
Imputed Contact Network:

https://github.com/ASU-Lim-Lab/Network_analysis



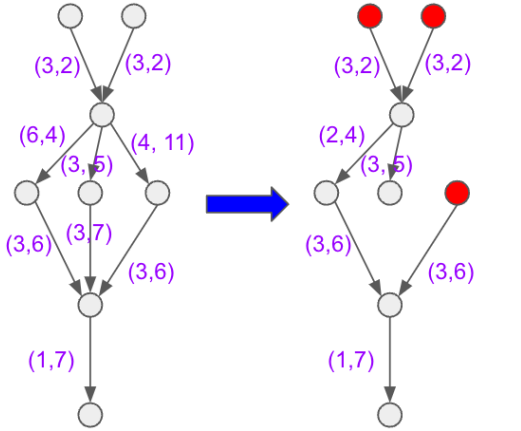


Directed edge:

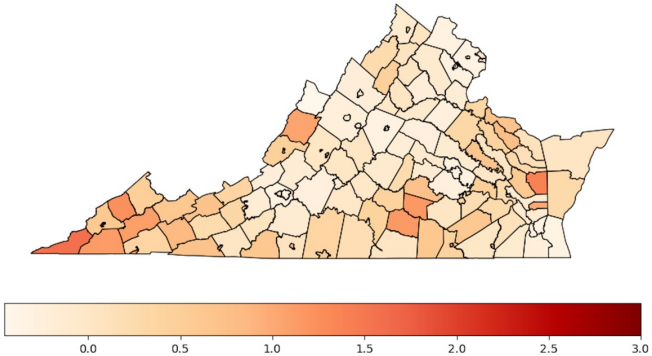


Candidate relationships

Imputed contact network

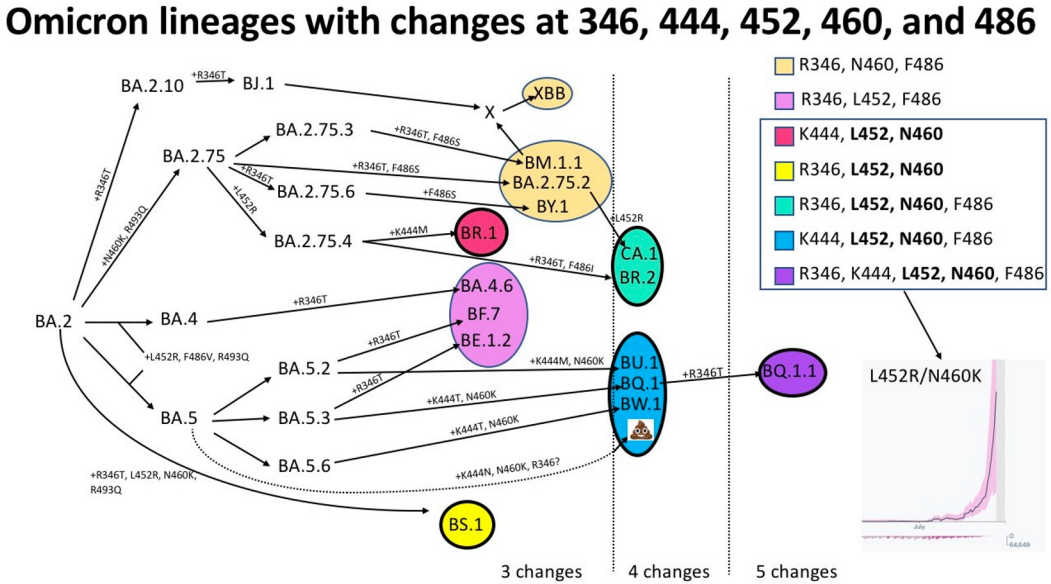


Importations in **red**

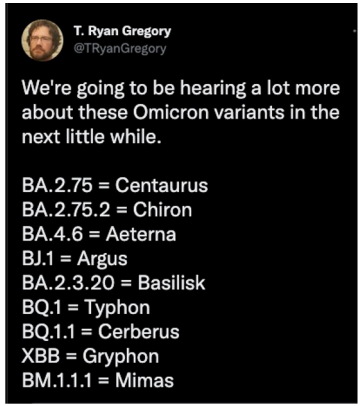


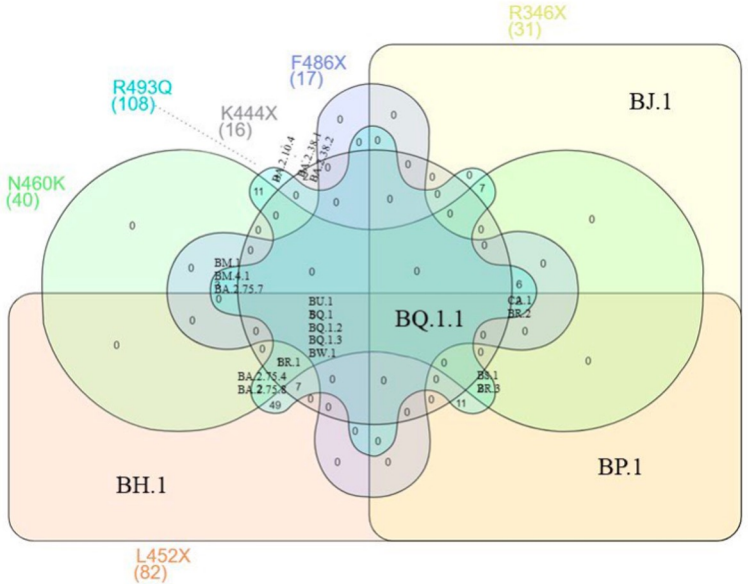
1. McBroome, J., Martin, J., de Bernardi Schneider, A., Turakhia, Y., and Corbett-Detig, R. (2022). Identifying SARS-CoV-2 regional introductions and transmission clusters in real time. Virus Evolution *8*, veac048. 10.1093/ve/veac048.

Communication



https://twitter.com/dfocosi/status/1574319177602670592?s=12&t=myre59DuQ1H31TI_-vd7DQ

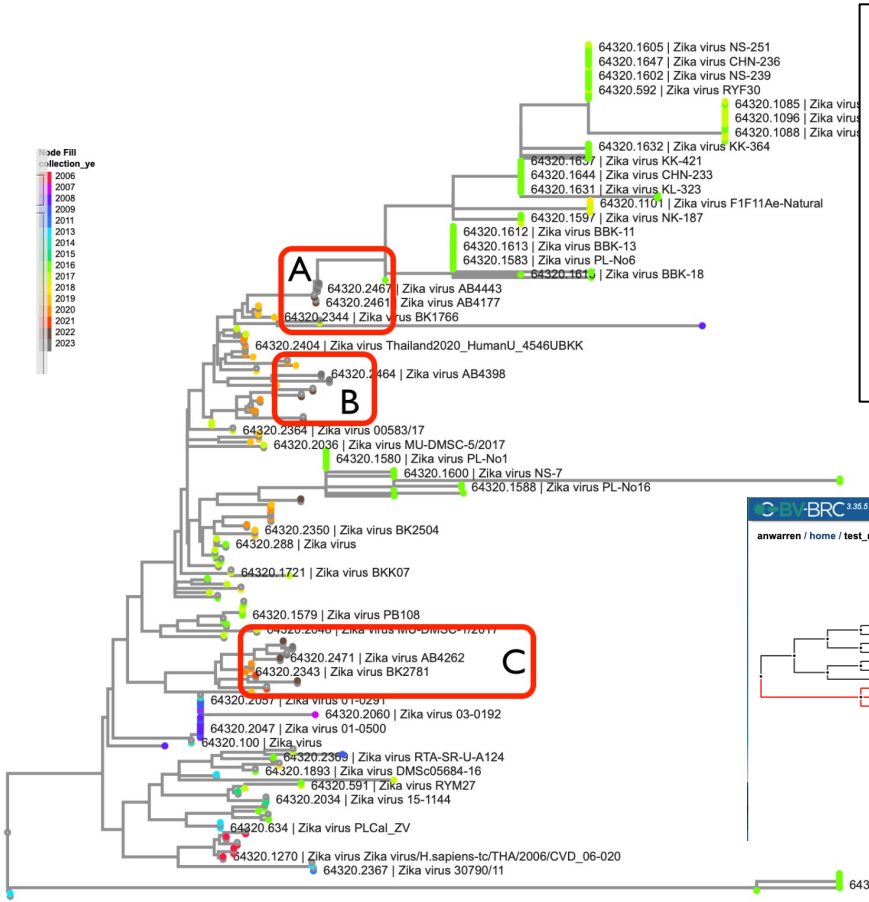


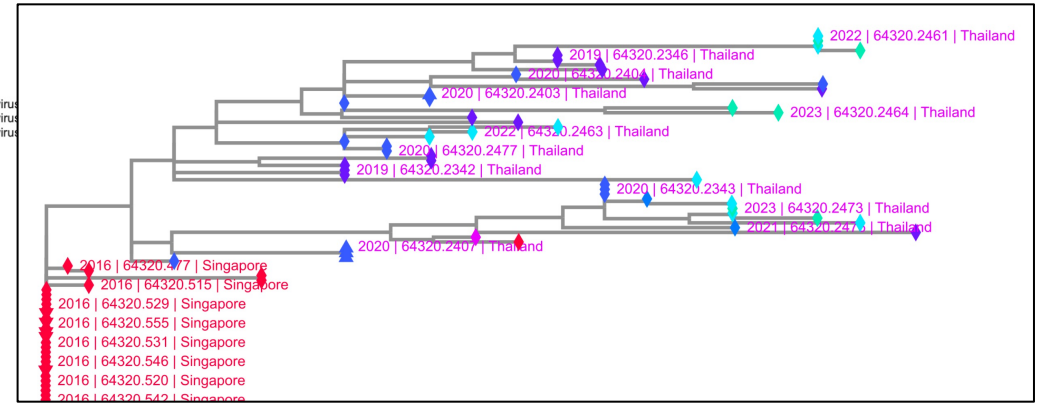


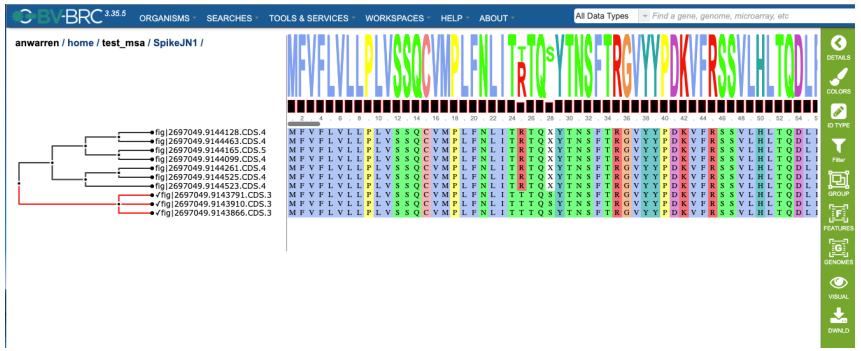
https://twitter.com/dfocosi/status/1574351472787165185

Useful Tools

Archaeopteryx and MSA viewer BV-BRC







https://www.bv-brc.org/

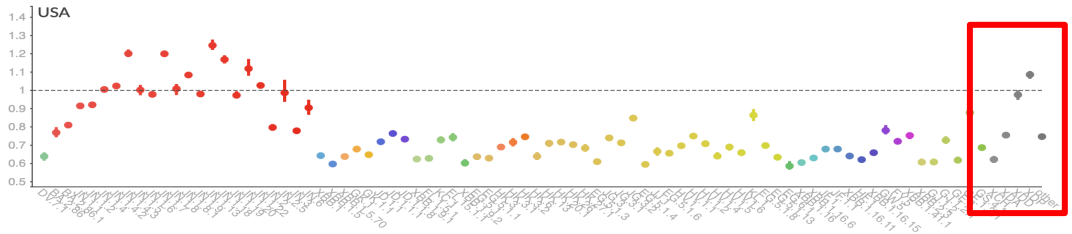
Label Color
isolation_country
- Singapore
- Thailand

Node Fill
collection_year
- 2016
- 2017
- 2019
- 2020
- 2021
- 2022
- 2023

Node Shape
host_common_name
- Asian tiger mosquito
- Human
- Southern House Mosquito
- Yellow fever mosquito

NextStrain / NextClade

https://nextstrain.org/sars-cov-2/forecasts/



Lineage growth advantage

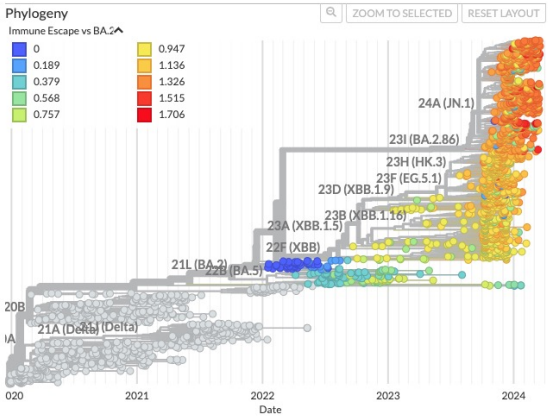
These plots show the estimated growth advantage for given Pango lineages relative to lineage JN.1. This describes how many more secondary infections a variant causes on average relative to lineage JN.1. Vertical bars show the 95% HPD. The "hierarchical" panel shows pooled estimate of growth rates across different locations. Results last updated 2024-03-19.

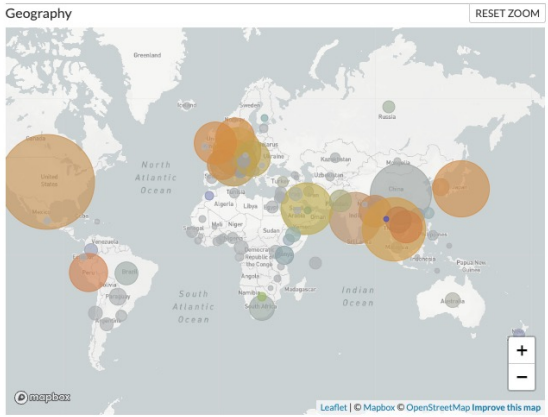
DV.7.1 BA.2.86 BA.2.86.1 JN.1.1 JN.1.2 JN.1.4 JN.1.4.2 JN.1.4.3 JN.1.5 JN.1.6
JN.1.7 JN.1.8 JN.1.8.1 JN.1.9 JN.1.13 JN.1.18 JN.1.19 JN.1.20 JN.1.22 JN.2
JN.2.5 JN.3 JN.6 XBB.1 XBB.1.5 XBB.1.5.70 GK.1 GK.1.1 JD.1.1 JD.1.1.1
JD.1.1.8 XBB.1.9.1 FL.1.5.1 KC.1 FL.15.1.1 XBB.1.9.2 EG.5.1 EG.5.1.1 HK.3
HK.3.1 HK.3.2 HK.6 HK.13 HK.20.1 HK.26 HK.27.1 EG.5.1.3 JG.3 JG.3.1
JG.3.2 EG.5.1.4 JJ.1 EG.5.1.6 HV.1 HV.1.1 HV.1.2 HV.1.4 HV.1.5 HV.1.6
KL.1 EG.5.1.8 EG.5.1.13 EG.6.1 XBB.1.16 XBB.1.16.6 JF.1 JF.1.1 XBB.1.16.11
HF.1 XBB.1.16.15 GW.5 FY.5 XBB.1.41.1 XBB.2.3 GJ.1.2 GJ.1.2.1 GE.1 GE.1.2.1
GS.4.1 XCH XDA XDD XDP other

Genomic epidemiology of SARS-CoV-2 with subsampling focused globally over the past 6 months

Built with nextstrain/ncov. Maintained by the Nextstrain team. Data updated 2024-04-06. Enabled by data from GenBank.

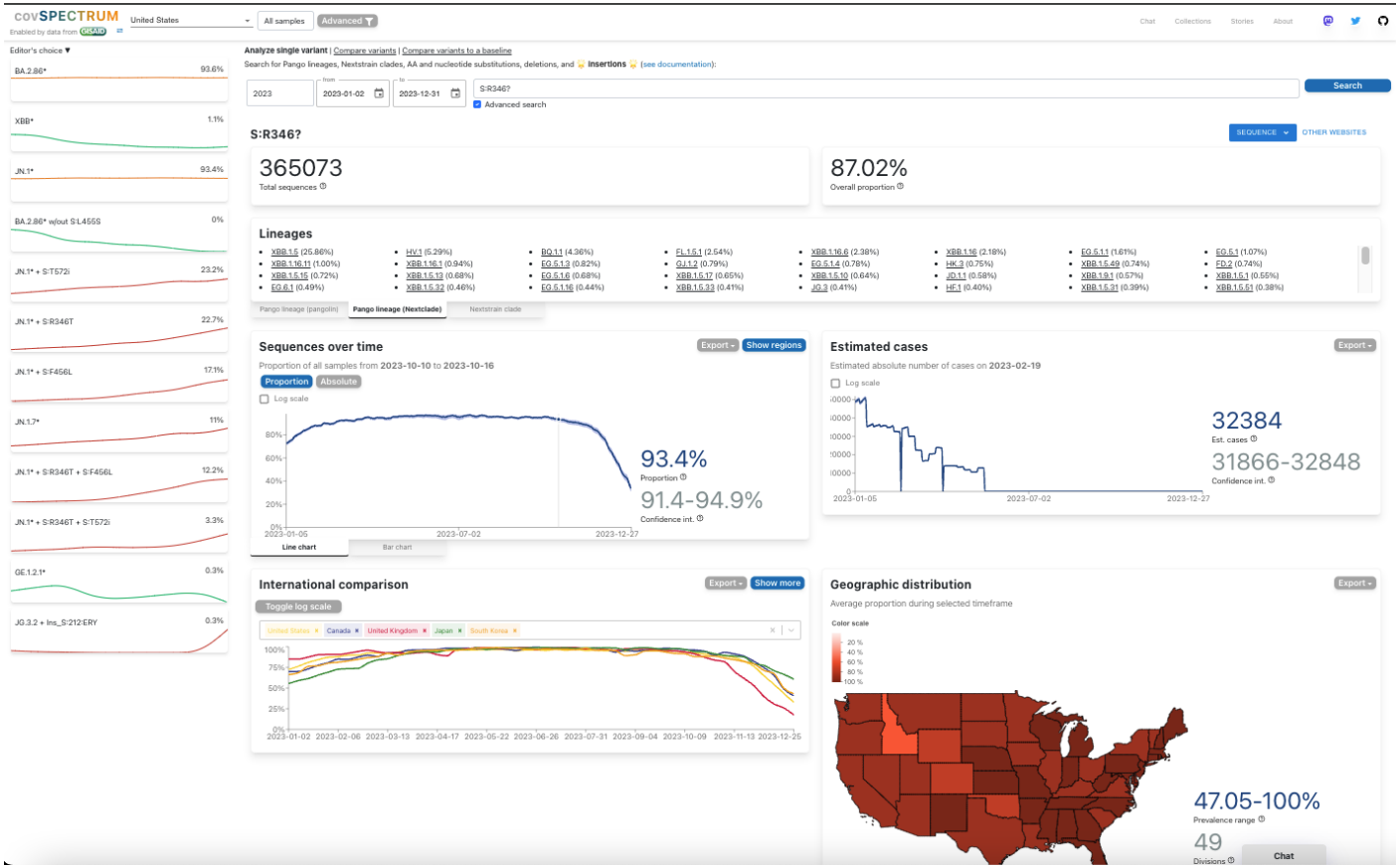
Showing 2016 of 2016 genomes sampled between Dec 2019 and Mar 2024.





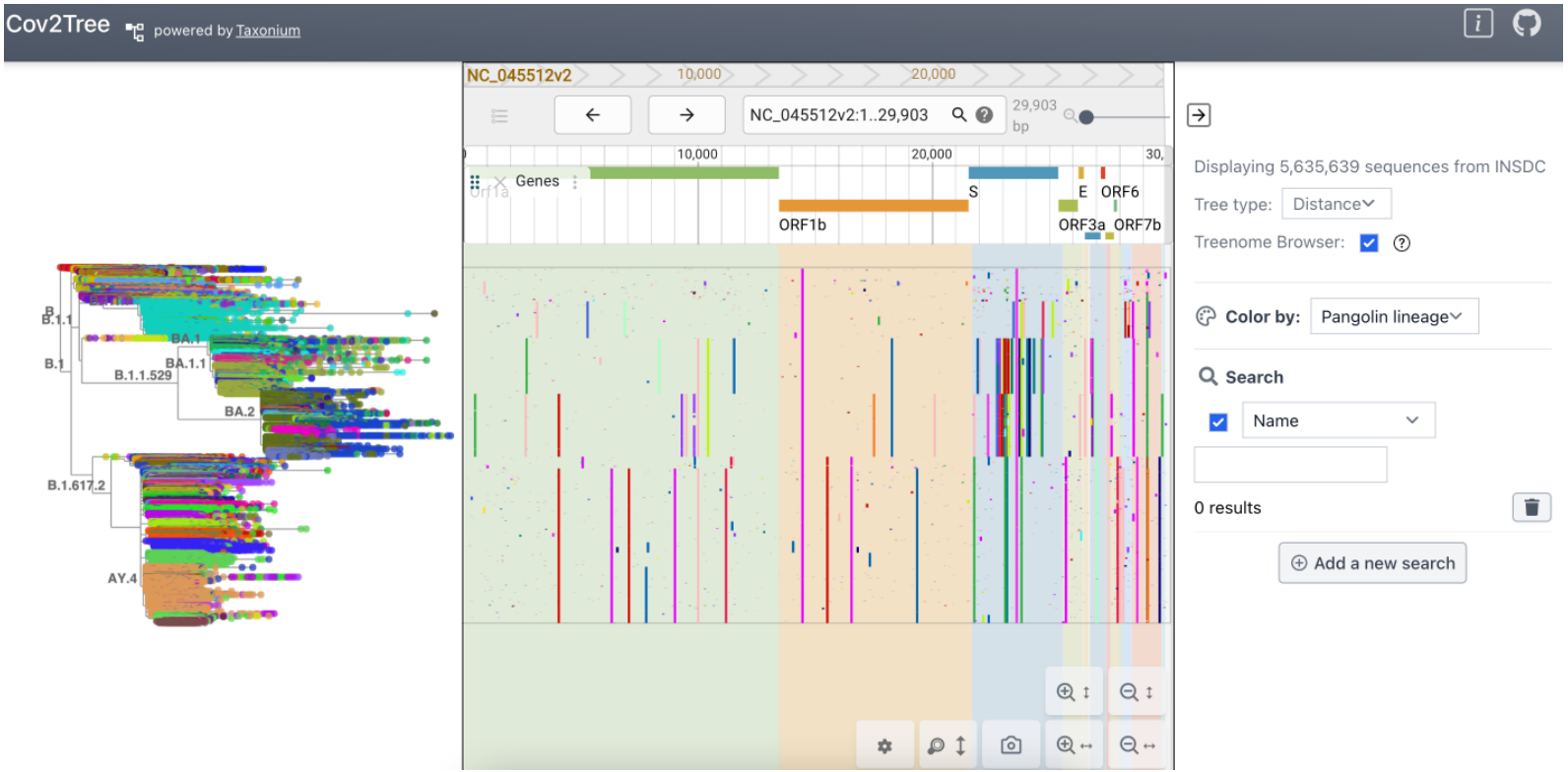
CovSpectrum

https://cov-spectrum.org/



Treenome / Taxonium

https://docs.taxonium.org/en/latest/treenome.html
https://genomium.org/



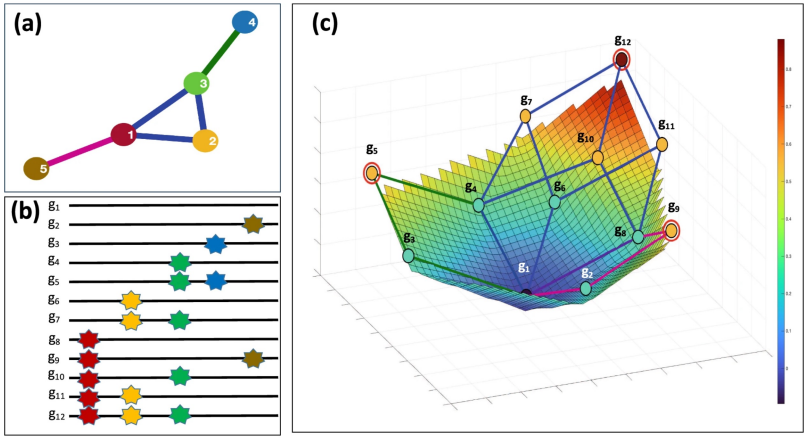
Wasabi

http://was.bi/

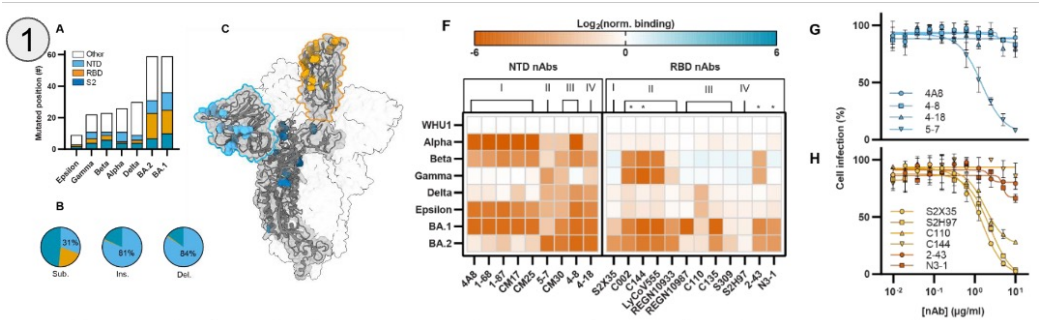
Potential Gaps

Epistatic interactions

Fig. 5: The model of an epistatically-constrained sequence space and fitness landscape.

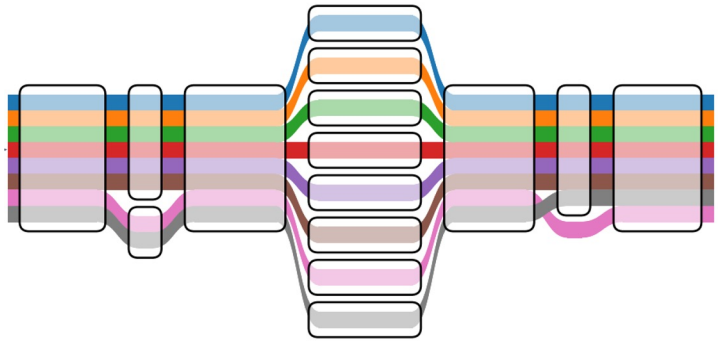


Mohebbi, F., Zelikovsky, A., Mangul, S., Chowell, G., and Skums, P. (2024). Early detection of emerging viral variants through analysis of community structure of coordinated substitution networks. Nat Commun *15*, 2838. 10.1038/s41467-024-47304-6.



Texas researchers find that "stabilizing mutations in the N-terminal and S2 domains of the spike protein compensate for destabilizing mutations in the receptor binding domain, thereby enabling the record number of mutations in Omicron sub-lineages." The compensating region, N-terminal and S2 domains, are highlighted in shades of blue, in panels A&C. Panel F shows monoclonal binding affinity to the receptor binding domain. Panels G and H compare virus neutralization of NTD and RBD directed monoclonal antibodies respectively.
https://www.biorxiv.org/content/10.1101/2022.04.18.488614v1

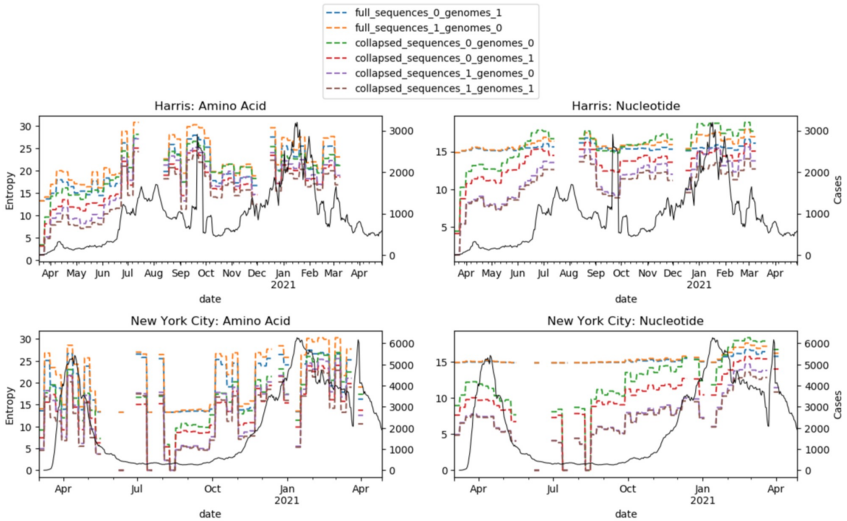
Compensatory mutations

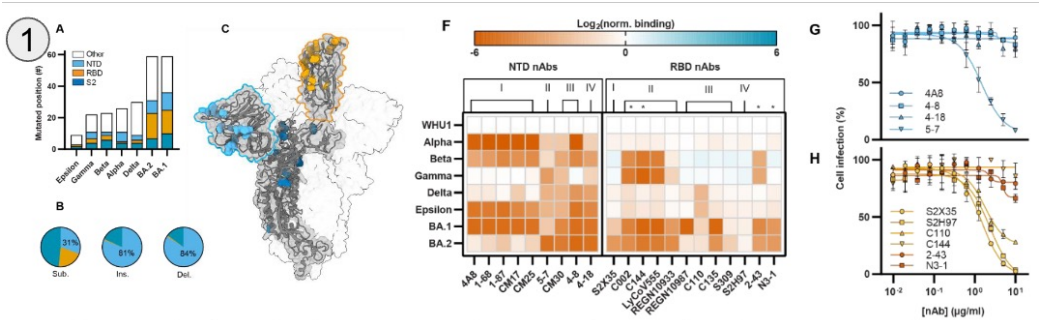


https://github.com/vgteam/sequenceTubeMap
Hickey, G., Monlong, J., Ebler, J., Novak, A.M., Eizenga, J.M., Gao, Y., Marschall, T., Li, H., and Paten, B. (2023). Pangenome graph construction from genome alignments with Minigraph-Cactus. Nat Biotechnol, 1–11. 10.1038/s41587-023-01793-w.

Potential Gaps

Epistatic interactions



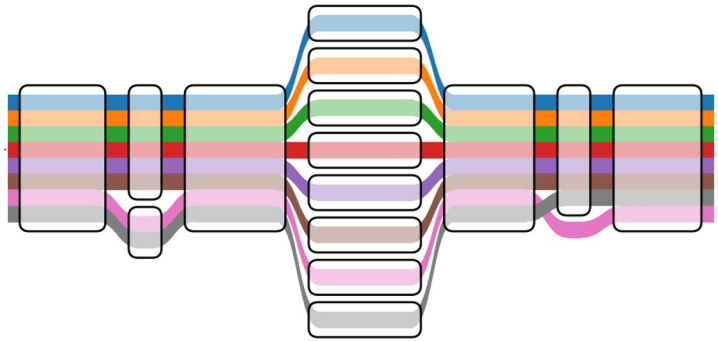


Texas researchers find that “stabilizing mutations in the N-terminal and S2 domains of the spike protein compensate for destabilizing mutations in the receptor binding domain, thereby enabling the record number of mutations in Omicron sub-lineages.” The compensating region, N-terminal and S2 domains, are highlighted in shades of blue, in panels A&C. Panel F shows monoclonal binding affinity to the receptor binding domain. Panels G and H compare virus neutralization of NTD and RBD directed monoclonal antibodies respectively.

https://www.biorxiv.org/content/10.1101/2022.04.18.488614v1

United States: 1/9/2022 – 4/16/2022 | United States: 4/10/2022 – 4/16/2022 NOW

Compensatory mutations



https://github.com/vgteam/sequenceTubeMap

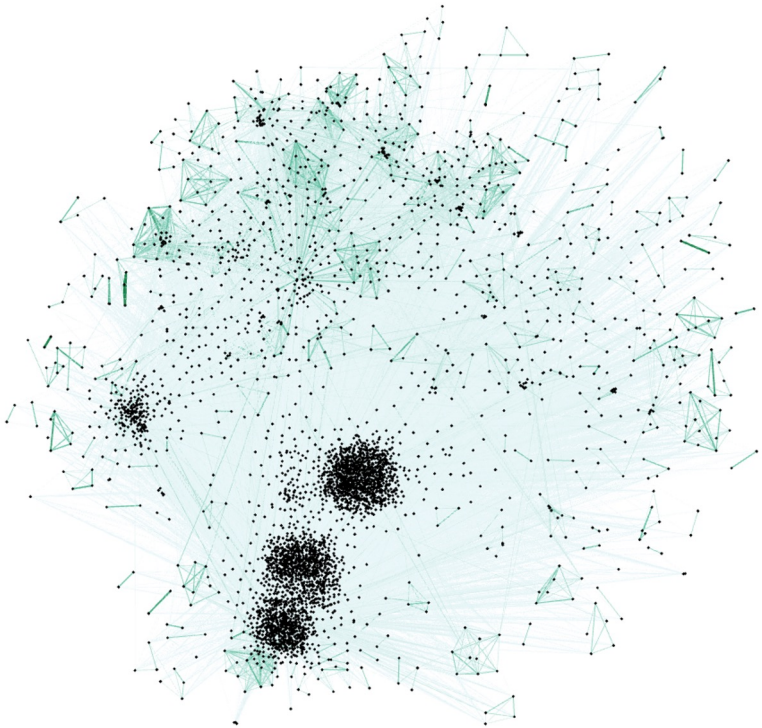
Hickey, G., Monlong, J., Ebler, J., Novak, A.M., Eizenga, J.M., Gao, Y., Marschall, T., Li, H., and Paten, B. (2023). Pangenome graph construction from genome alignments with Minigraph-Cactus. Nat Biotechnol, 1–11. 10.1038/s41587-023-01793-w.

Epistatic Interactions

Clustering mutations based on coordinated substitutions

- Mutations as nodes and Pearson correlation coefficients as edge weights in graph
- Based on NCBI ACTIVETrace
- Clustered using Markov clustering algorithm
- Input
  - 4,676 nodes i.e. mutations
  - 134,895 mutation pairs

- AWS Athena query to count co-occurring mutations
  - ~6 million SRA datasets
  - ~11 Gb of VCF data processed
- Calculate pearson correlation coefficient between mutations
  - 134,895 mutation pairs with pearson coefficient greater than cutoff >0.1 < -0.1 across whole genome
  - 7768 non-synonymous mutation pairs in Spike gene



Connor, R., Shakya, M., Yarmosh, D.A., Maier, W., Martin, R., Bradford, R., Brister, J.R., Chain, P.S.G., Copeland, C.A., di Iulio, J., et al. (2024). Recommendations for Uniform Variant Calling of SARS-CoV-2 Genome Sequence across Bioinformatic Workflows. Viruses *16*, 430. 10.3390/v16030430.

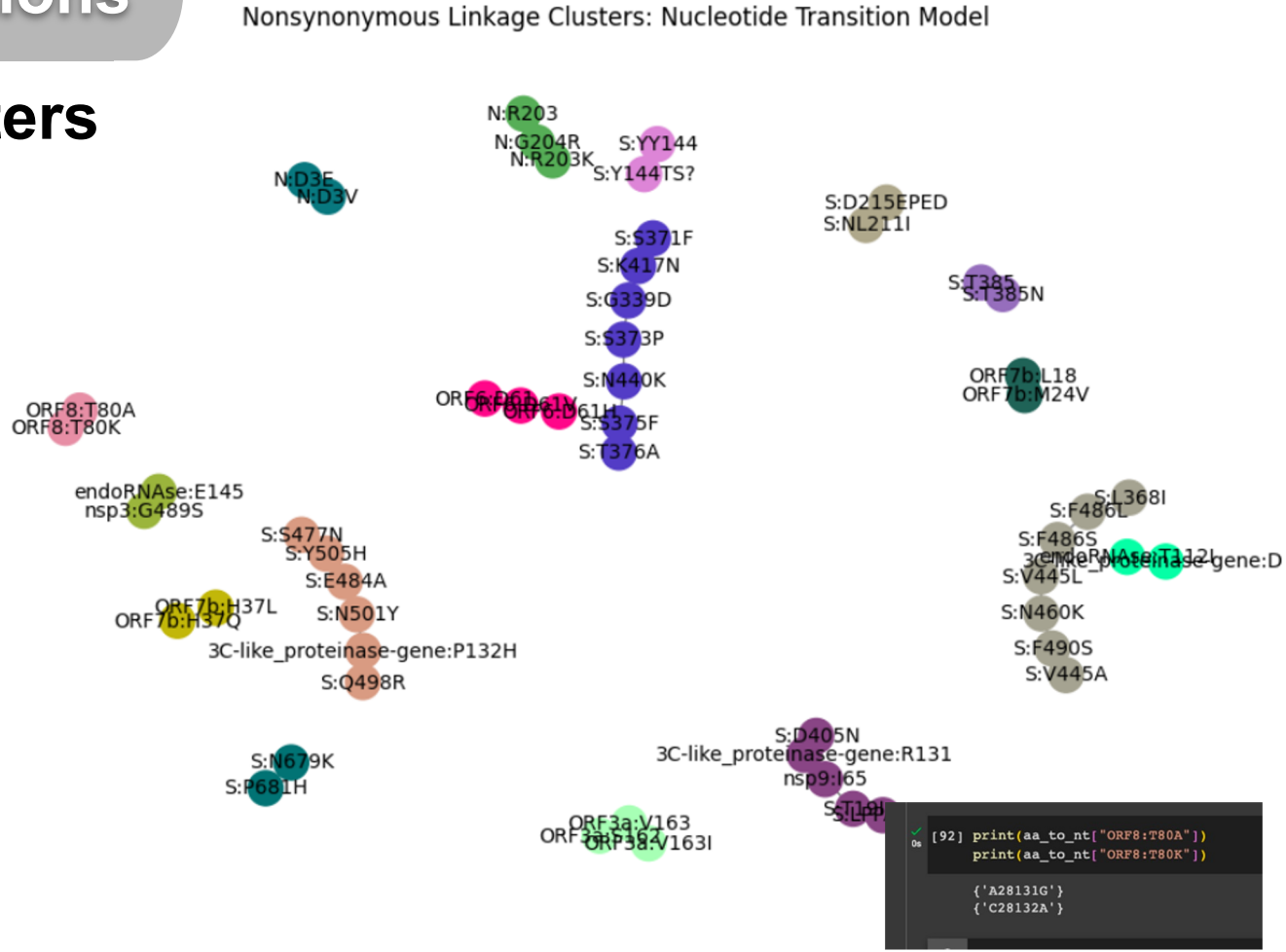
Epistatic Interactions

Significant clusters

Many co-located mutations

Useful unit test, that our processing and transformations are working

Negative edge weights likely contributing to some missing mutations here

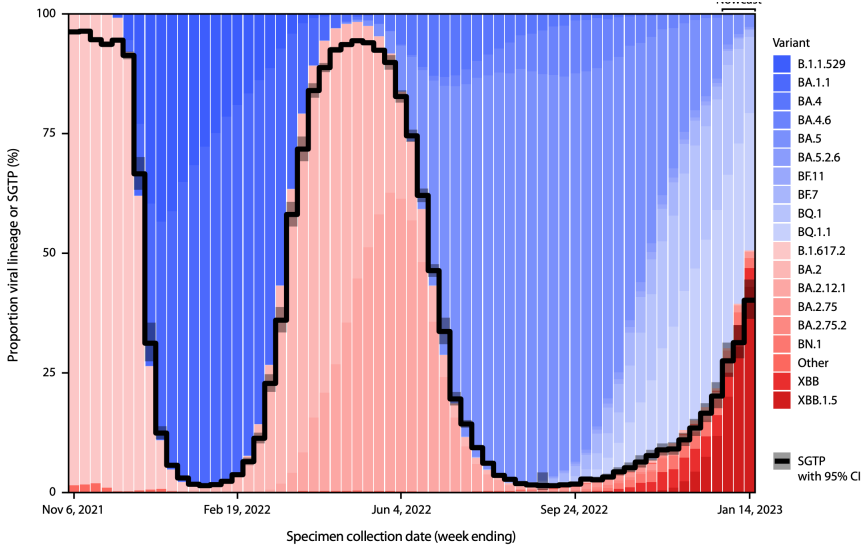


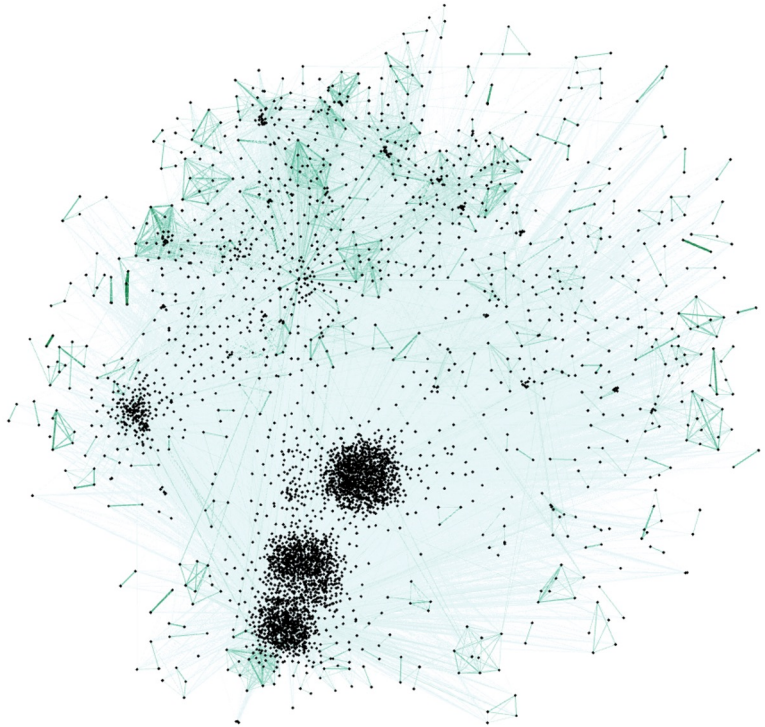
Epistatic Interactions

NIAID / NIH Codeathon Summer 2023
VCF Files for Population Genomics: Scaling to Millions of Samples

Clustering mutations based on coordinated substitutions

- AWS Athena query to count co-occurring mutations
  - ~6 million SRA datasets
  - ~11 Gb of VCF data processed
- Calculate pearson correlation coefficient between mutations
  - 134,895 mutation pairs with pearson coefficient greater than cutoff >0.1 < -0.1 across whole genome
  - 7768 non-synonymous mutation pairs in Spike gene

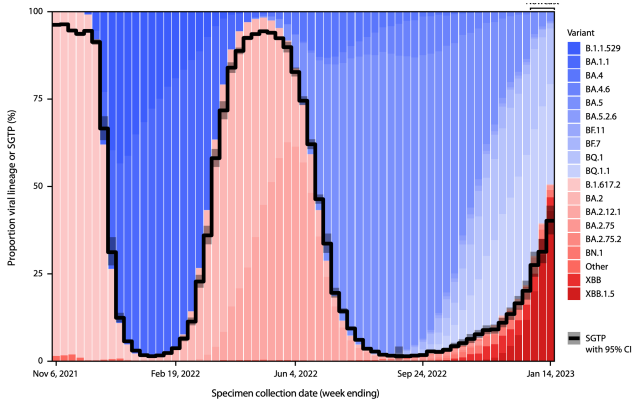




Connor, R., Shakya, M., Yarmosh, D.A., Maier, W., Martin, R., Bradford, R., Brister, J.R., Chain, P.S.G., Copeland, C.A., di Iulio, J., et al. (2024). Recommendations for Uniform Variant Calling of SARS-CoV-2 Genome Sequence across Bioinformatic Workflows. Viruses *16*, 430. 10.3390/v16030430.

Epistatic Interactions

Clustering mutations based on coordinated substitutions



aa_pos1	aa_pos2	aa_change_1	aa_change_2	record_count	samples	projects	record_freq	rel_record_freq	jaccard	pearson
68	144	S: IHV68I	S: YY144Y	142158	138841	173	0.07183699	7.94080589	0.60405628	0.70238209
68	501	S: IHV68I	S: N501Y	162140	159789	206	0.08193453	6.15353172	0.54454531	0.63259802
68	890	S: IHV68I	nsp3: A890D	141533	138296	168	0.07152116	6.64180292	0.52575409	0.61470751
68	183	S: IHV68I	nsp3: T183I	144543	141166	169	0.0730422	6.525839	0.52679284	0.61456637
403	68	RNA-dependent_RNA_polymerase-gene: P403	S: IHV68I	141652	138205	170	0.07158129	6.52345531	0.51881478	0.6066465
1118	68	S: D1118H	S: IHV68I	146067	142569	172	0.07381233	6.35028319	0.51949156	0.60583962
27	68	ORF8: Q27*	S: IHV68I	143342	139902	173	0.0724353	6.43282118	0.51769133	0.60465188
68	36	S: IHV68I	nsp2: S36	141548	138160	169	0.07152874	6.49245863	0.51661928	0.60415449
68	681	S: IHV68I	S: P681H	177277	173819	210	0.08958374	5.35617206	0.50876463	0.60289064

Connor, R., Shakya, M., Yarmosh, D.A., Maier, W., Martin, R., Bradford, R., Brister, J.R., Chain, P.S.G., Copeland, C.A., di Iulio, J., et al. (2024). Recommendations for Uniform Variant Calling of SARS-CoV-2 Genome Sequence across Bioinformatic Workflows. Viruses *16*, 430. 10.3390/v16030430.

Potential Gaps

Structural Annotation / Proximity related to observed mutations

Proximity of 27 to other positions

Computed A Spike Open

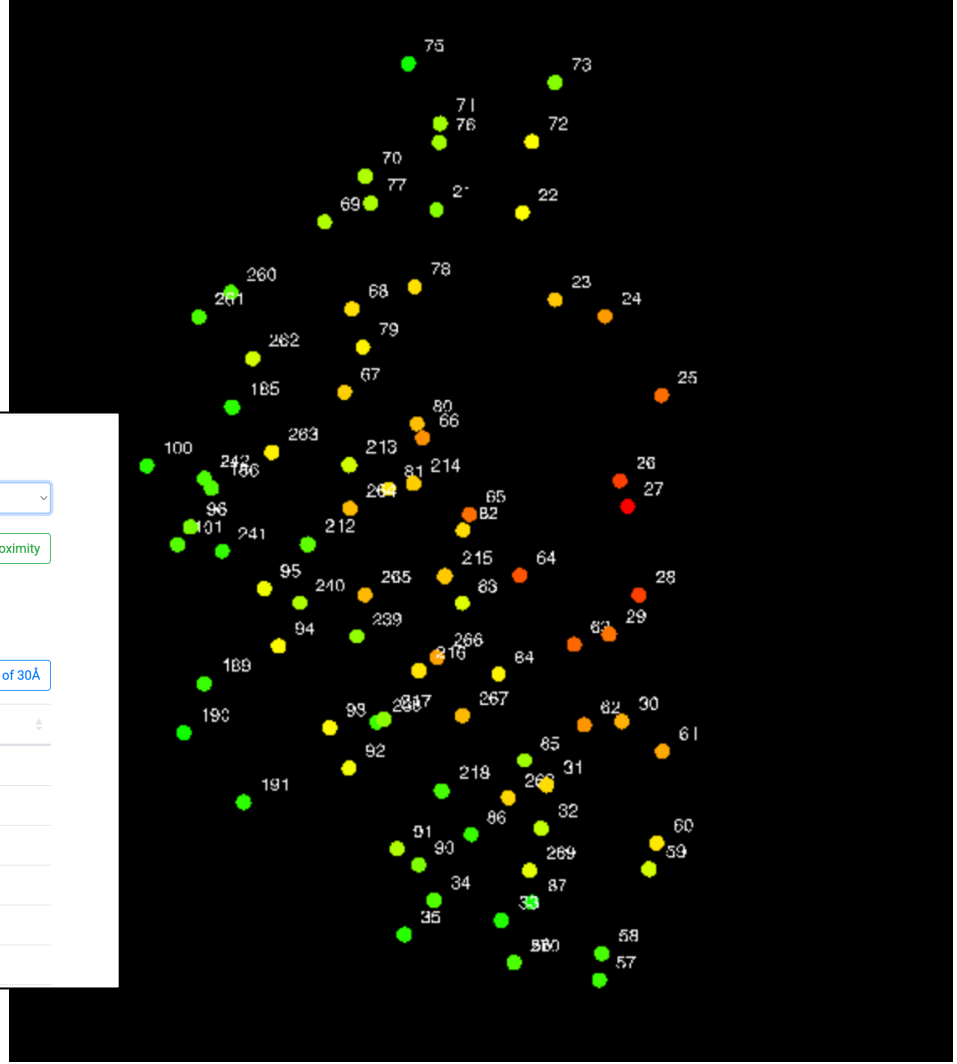
27 | Show Proximity

View 3D model of spike centred on position 27

Close (within 15Å) to position 27: 23-31, 60-68, 72-72, 78-82, 84-84, 93-94, 214-216, 263-268

Filter table to max of 30Å

Position	Chain A	Chain B	Chain C	Smallest
1	44.2	111.9	94.6	44.2
2	43.6	111.9	91.6	43.6
3	40.4	112.4	93.1	40.4
4	40	109.9	93.8	40
5	40	112.5	97.4	40
6	38.5	110	98.3	38.5



https://ukcovid.xyz/proximity.php

Support & Thank you

Funding:
BV-BRC NIAID
PGCOE CDC
Virginia Department of Health

Ryan Connor (ACTIVE trace SRA, Query supply/assist)
Elliot Lefkowitz & BV-BRC Team
UVA Biocomplexity Institute
SPHERES Team
PGCOE Team
VDH Team

BV-BRC