Transitions Omicron subvariant and amplicon dropouts

Update Feb. 22, 2022

Bette Korber, Will Fischer, Hyejin Yoon, James Theiler

cov.lanl.gov

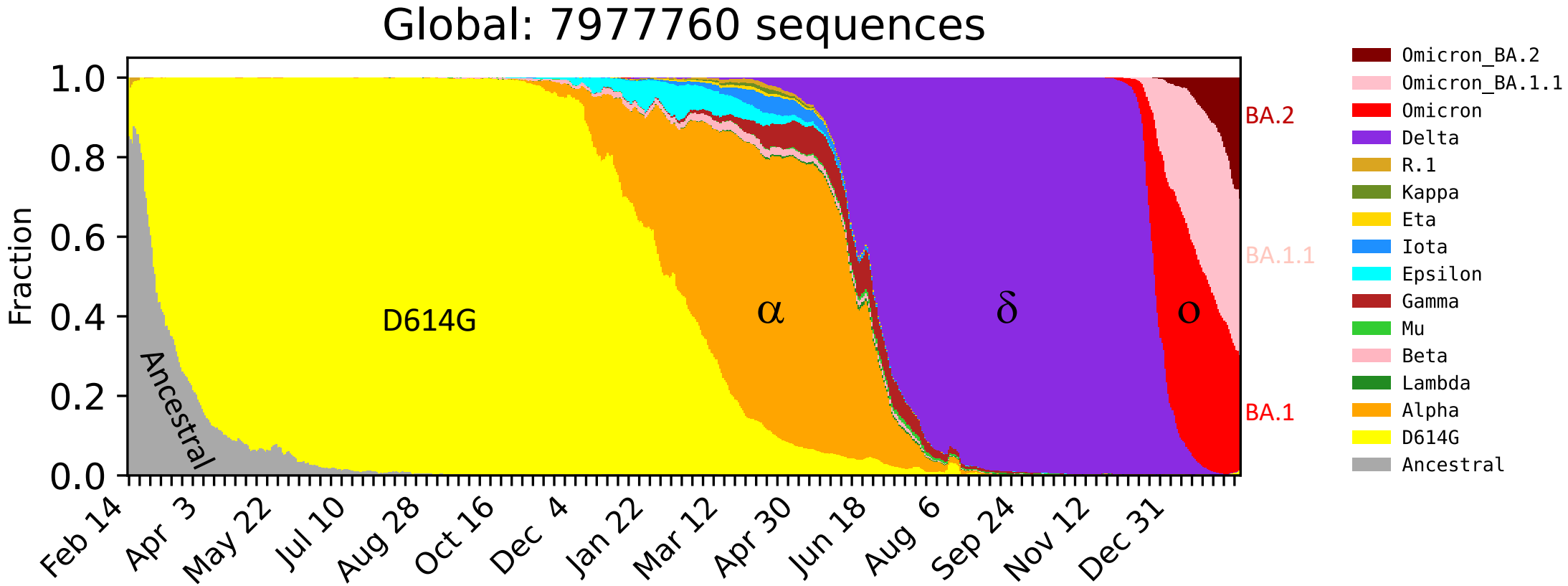
Contents:

Los Alamos NATIONAL LABORATORY

GISAID

LA-UR-21-28226

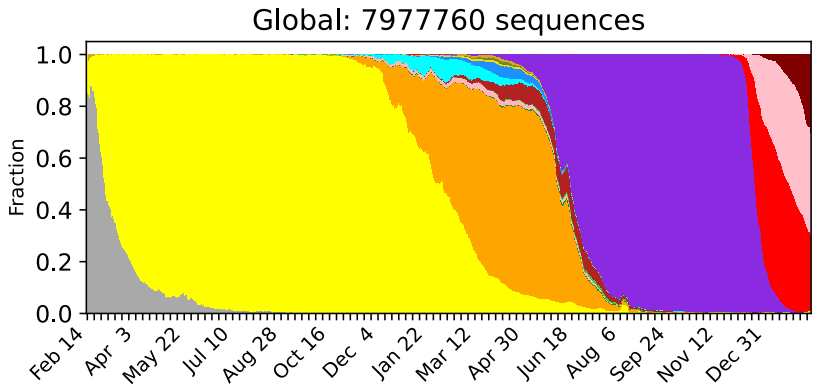
Transitions in major global lineages between Feb 14, 2020 and Feb. 14, 2022

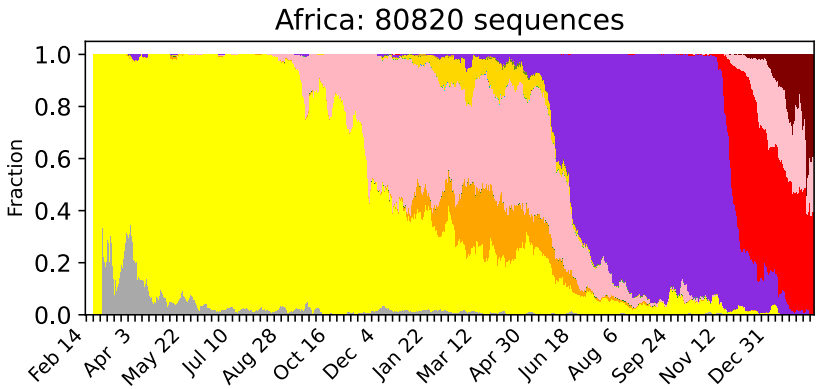


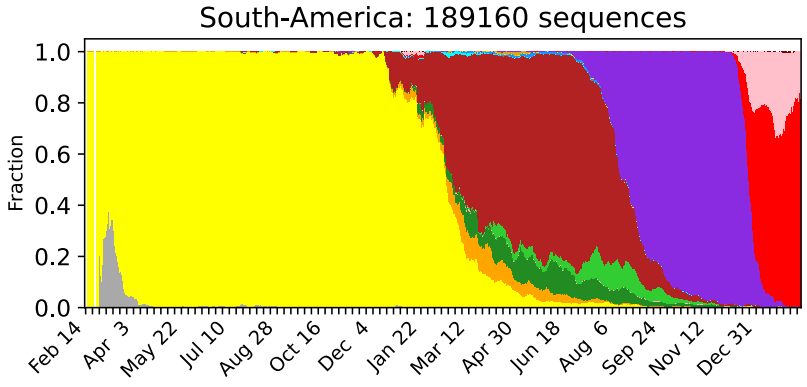
Both BA.1.1 and BA.2 Omicron sublineages are increasingly sampled relative to the original form BA.1.
Early reports suggest BA.2 has different antigenic and phenotypic characteristics relative to BA.1
Yamasoba… Sato, bioRxiv, doi.org/10.1101/2022.02.14.480335
Iketani… Ho, bioRxiv, doi.org/10.1101/2022.02.07.479306

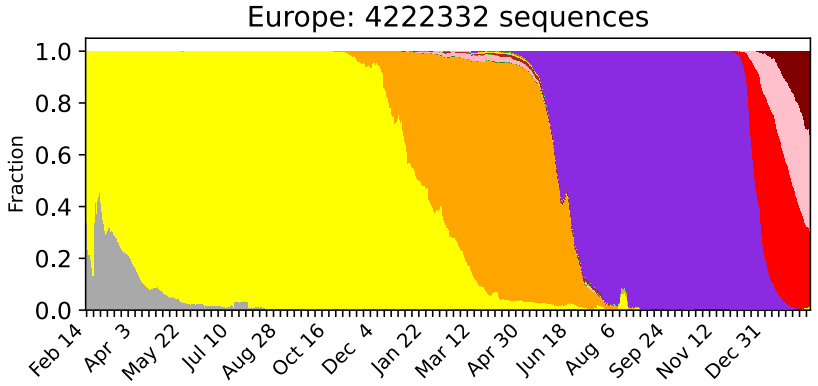
With thanks to those who share sequences through GISAID

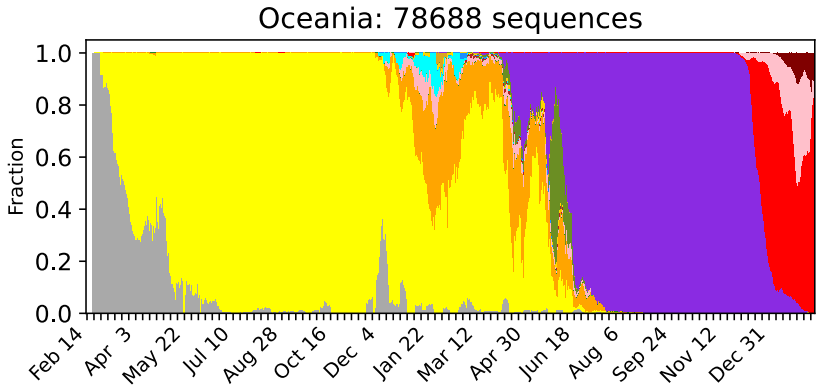
Transitions by continent: BA.2 is rapidly gaining ground, but is only recently sampled in the Americas

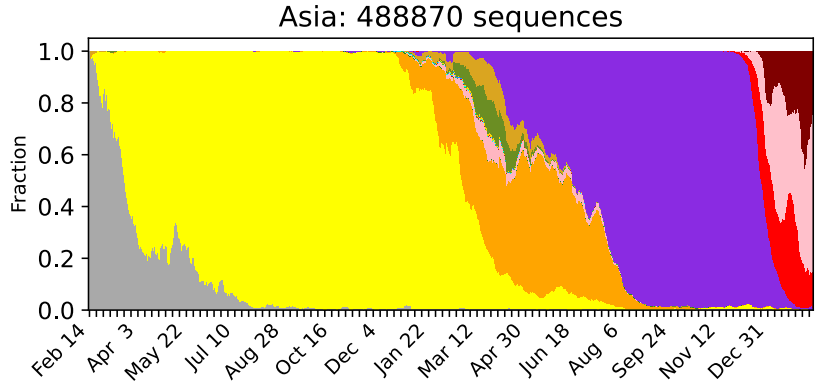


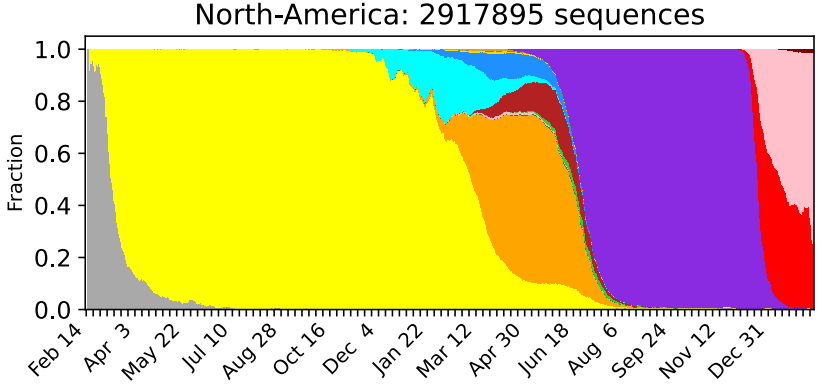


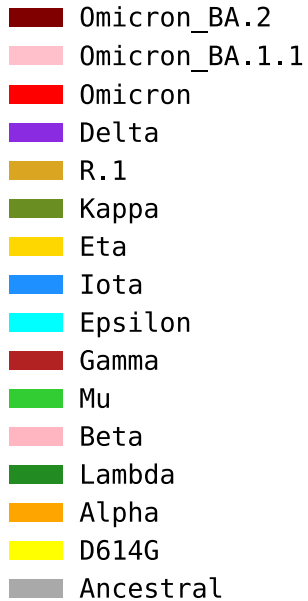












GISAID

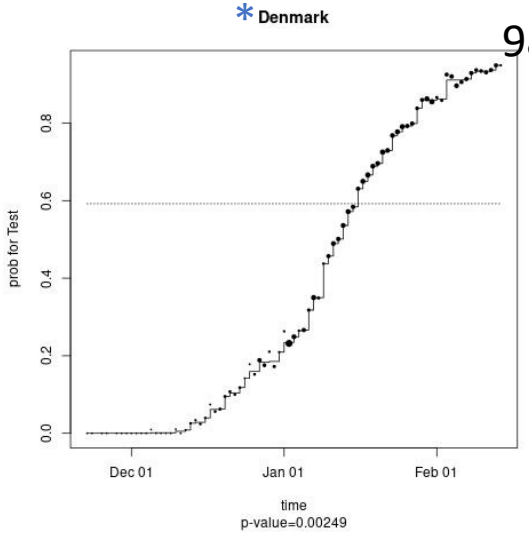
Los Alamos NATIONAL LABORATORY

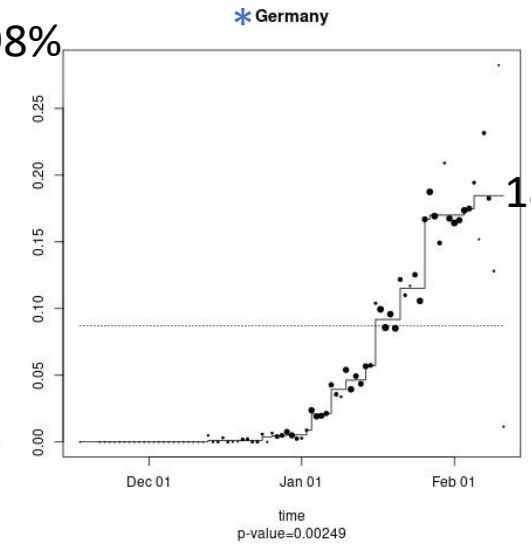
The BA.2 Omicron sub-lineage is increasingly sampled and replacing other Omicron variants across the globe. Data: Feb. 22, 2022

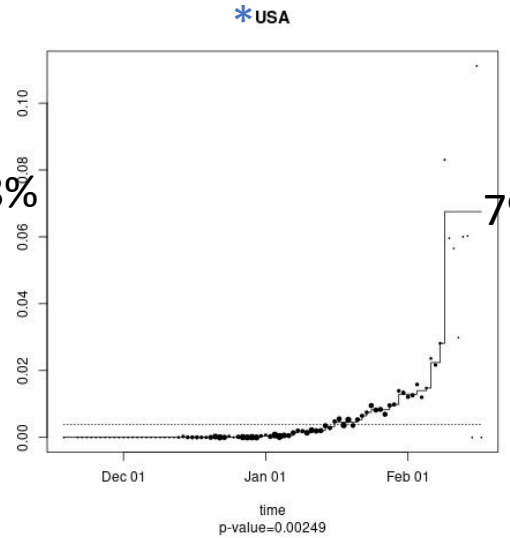
		BA.2	Other Omicron				
	# Test	# Others	Total	Test/Total (%)	# days	Time window	p-val
Australia	256	13832	14088	1.82	82	82	0.00249
Austria	719	1196	1915	37.55	70	71	0.00249
Bangladesh	143	163	306	46.73	44	56	0.00249
Belgium	658	10788	11446	5.75	85	90	0.00249
Botswana	58	1130	1188	4.88	60	66	0.00249
Brazil	12	12486	12498	0.10	72	78	0.00249
Brunei	93	65	158	58.86	20	27	0.00249
Cambodia	180	263	443	40.63	52	62	0.00249
Canada	730	34738	35468	2.06	77	77	0.00249
Croatia	87	5279	5366	1.62	57	65	0.00249
Czech-Republic	158	4785	4943	3.20	70	87	0.00249
* Denmark	51518	35488	87006	59.21	83	84	0.00249
Estonia	46	618	664	6.93	41	42	0.00249
Finland	17	1661	1678	1.01	58	64	0.00249
France	1158	34446	35604	3.25	85	84	0.00249
Georgia	197	357	554	35.56	50	61	0.00249
* Germany	5647	59264	64911	8.70	84	86	0.00249
Hong-Kong	286	260	546	52.38	61	87	0.00249
India	7670	4743	12413	61.79	78	81	0.00249
Indonesia	308	5914	6222	4.95	63	77	0.00249
Ireland	108	6739	6847	1.58	74	75	0.00249
Israel	646	13782	14428	4.48	77	80	0.00249
Italy	154	9833	9987	1.54	76	77	0.00249
Japan	308	11475	11783	2.61	75	78	0.00249
Jordan	17	30	47	36.17	19	36	0.00249
Latvia	14	393	407	3.44	29	37	0.02736
Lithuania	47	743	790	5.95	54	59	0.00249
Luxembourg	132	4287	4419	2.99	60	62	0.00249
Malaysia	34	939	973	3.49	60	79	0.00249
Maldives	51	148	199	25.63	44	62	0.00249
Mexico	13	8296	8309	0.16	72	86	0.02239
Nepal	113	50	163	69.33	29	63	0.00249
Netherlands	440	9923	10363	4.25	84	87	0.00249
New-Zealand	202	828	1030	19.61	58	60	0.00249
Norway	832	5492	6324	13.16	75	76	0.00249
Pakistan	14	74	88	15.91	22	61	0.00249
Philippines	974	250	1224	79.58	59	80	0.00249
Poland	1050	18439	19489	5.39	65	70	0.00249
Portugal	84	2996	3080	2.73	64	75	0.00249

		BA.2	Other Omicron				
	# Test	# Others	Total	Test/Total (%)	# days	Time window	p-val
Qatar	191	64	255	74.90	46	56	0.00249
Reunion	34	1514	1548	2.20	60	78	0.00249
Romania	82	1679	1761	4.66	54	64	0.00249
Russia	37	421	458	8.08	38	53	0.00249
Singapore	667	1441	2108	31.64	71	74	0.00249
Slovakia	76	1632	1708	4.45	68	77	0.00249
Slovenia	54	7194	7248	0.75	59	64	0.00249
South-Africa	689	5589	6278	10.97	83	83	0.00249
South-Korea	50	2139	2189	2.28	59	63	0.00249
Spain	215	11743	11958	1.80	81	82	0.00249
Sri-Lanka	232	388	620	37.42	62	84	0.00249
Sweden	2490	4581	7071	35.21	81	88	0.00249
Switzerland	431	16208	16639	2.59	86	85	0.00249
Thailand	188	2732	2920	6.44	67	77	0.00249
Turkey	16	5125	5141	0.31	27	71	0.00249
* USA	1767	457018	458785	0.39	90	91	0.00249
United-Kingdom	25930	463863	489793	5.29	91	92	0.00249
Vietnam	13	198	211	6.16	38	49	0.00249

The table lists *all* countries where BA.2 has been sampled ≥10 times in GISAID. The low p-values indicate BA.2 is increasing in sampling frequency relative to other Omicron variants in every nation where it has been sampled. The plots show examples from Denmark, the nation first to show the transition, and Germany and the USA, examples of two nations where the transition started later. In the plots, each point represents samples from a given day. The size of the dot indicates the number sampled that day, with dates along the x axis, the proportion of BA.2 lineage sequences relative to other Omicron variants on the y axis. Note that the ranges on the y-axis are adjusted to best display the variation.





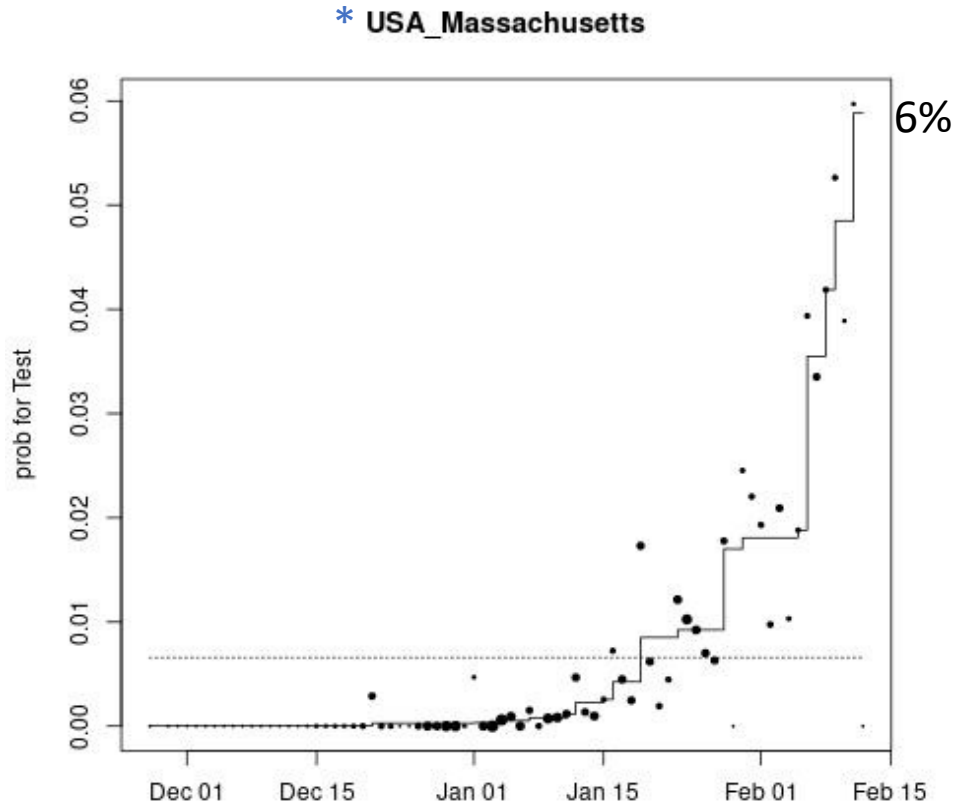


The BA.2 Omicron sub-lineage is still rare but increasingly sampled relative to other Omicron variants in states across the USA.

Data: Feb. 23, 2022

		BA.2 # Test	Other Omicron # Others	Total	Test/Total (%)	# days	Time window	p-val
	USA_Alaska	18	1124	1142	1.58	62	74	0.00498
	USA_Arizona	168	18856	19024	0.88	73	72	0.00249
	USA_California	345	74598	74943	0.46	83	85	0.00249
	USA_Colorado	53	32315	32368	0.16	75	79	0.00249
	USA_Connecticut	30	8840	8870	0.34	78	80	0.00249
	USA_Florida	20	16860	16880	0.12	74	76	0.00249
	USA_Georgia	29	10932	10961	0.26	69	71	0.00249
	USA_Hawaii	38	2086	2124	1.79	72	75	0.00249
	USA_Illinois	11	7854	7865	0.14	76	76	0.00249
	USA_Indiana	23	3480	3503	0.66	65	67	0.00249
	USA_Maryland	34	8474	8508	0.40	78	81	0.00249
*	USA_Massachusetts	258	38929	39187	0.66	77	77	0.00249
	USA_Minnesota	39	11202	11241	0.35	70	82	0.00249
	USA_Nevada	15	4365	4380	0.34	71	72	0.00249
	USA_New-Jersey	44	11056	11100	0.40	73	78	0.00249
	USA_New-York	119	32369	32488	0.37	83	87	0.00249
	USA_North-Carolina	49	12361	12410	0.39	74	75	0.00249
	USA_Northern-Mariana-Islands	17	56	73	23.29	19	20	0.36816
	USA_Ohio	41	6153	6194	0.66	73	76	0.00249
	USA_Oregon	10	6155	6165	0.16	66	67	0.00249
	USA_Pennsylvania	42	6084	6126	0.69	70	73	0.00249
	USA_Rhode-Island	171	2651	2822	6.06	69	73	0.00249
	USA_Tennessee	27	14731	14758	0.18	74	76	0.00249
	USA_Texas	64	28825	28889	0.22	78	78	0.00249
	USA_Utah	16	14886	14902	0.11	69	83	0.00249
	USA_Vermont	53	5227	5280	1.00	60	65	0.00249
	USA_Washington	65	15476	15541	0.42	77	81	0.00249
	USA_Wisconsin	22	8107	8129	0.27	75	79	0.00249

The table lists *all* US states where BA.2 has been sampled ≥10 times in GISAID. The low p-values indicate BA.2 is significantly increasing in sampling frequency relative to other Omicron variants, in 27/27 states, but not in the Northern Mariana Islands. The plot shows an example from Massachusetts, the state where BA.2 has been most sampled*. In the plot, each point represents samples from a given day. The size of the dot indicates the number sampled that day, with dates along the x axis, the proportion of BA.2 lineage sequences relative to other Omicron variants on the y axis.



Amplicon Dropout and SARS-CoV-2 sequences

As new variants arise among SARS-CoV-2 sequences, the mutations they carry can disrupt primer interactions and leave some parts of the full genome sequences unresolved; this situation will persist until new primers with superior performance can be adopted and incorporated into sequencing protocols. These adjustments will naturally have different paces in different laboratories.

The unresolved sections are generally filled in with N's.

To enable reporting sequences in the critical weeks just as a new highly transmissible variant is detected and found to be expanding, in the earliest days of a new expansion many sequences will have long stretches of N's. As people adapt their primer sets and incorporate sequencing strategies that better capture these regions, more complete forms of the variants become increasingly available over time.

The next two slides (6 and 7) illustrate how this scenario has played out with Omicron sequences.

Omicron carried so many mutations relative to earlier variants that the issue of amplicon dropout was exacerbated. However, this problem is not unique to omicron, and most new variants have presented similar initial sequencing challenges. We include a second example using Delta on slides 8 and 9.

Implications of Amplicon Dropout for interpreting sequences:

1) For experimentalist scientists ordering reagents to explore the immunological and virological impact of mutational patterns carried by new variants of interest and concern, care must be taken to make sure that the sequence they are using actually reflects the common circulating form of the variant. Some bioinformatics sites have treated "N"s as ancestral states, and this can lead to an under-representation of amino acid changes in common amplicon dropout regions in newly emerging lineages. We have taken *great care* to avoid this problem in the variant representative sequences included in these LANL VOI/VOC representative sequence sets.
2) Amplicon dropout can result in artificial chimeric sequences, where a very rare variant in a sample (e.g. a low-level laboratory contaminant) is successfully amplified by primers that miss the targeted sequence. Such "chimeras" will include stretches of a distinct form embedded within the sequence. Examples of stretches of ancestral or Delta sequences within Omicron sequences can be readily identified, as the Omicron Spike is so distinctive; these are discussed in slides 10-14, and examples are included in our fasta files and spread sheet.

Amplicon Dropout Regions in Omicron sequences, full genome sequences grouped by sampling month

Each sequence is represented as a row of pixels.

- All 6878 sequences from November, when Omicron was first detected and beginning to expand, are included (lower panel).

- 10,000 Omicron sequences were randomly selected from the December and January GISAID sequences (middle panels).

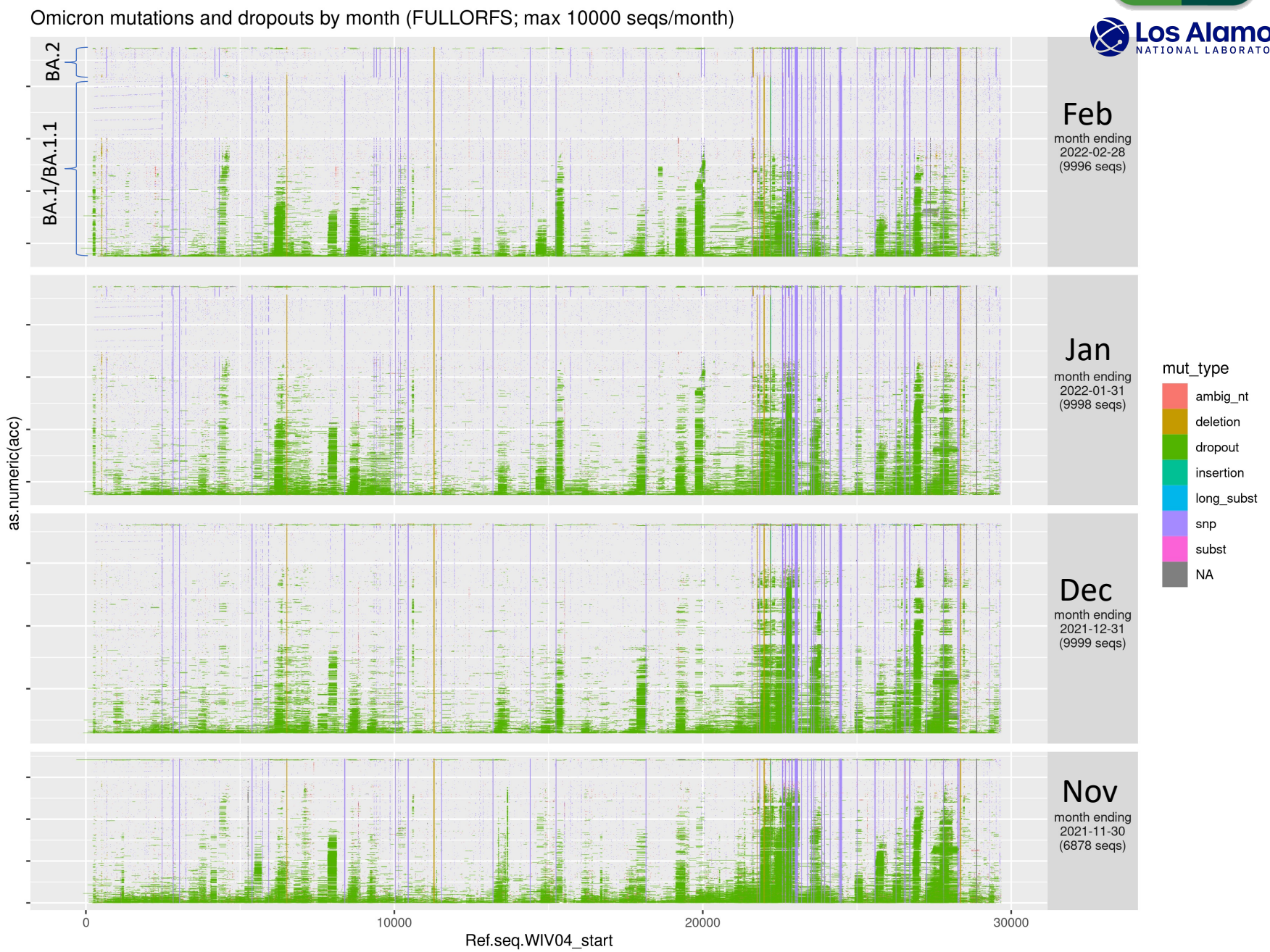
- All 9,996 sequences sampled in February and deposited in GISAID by February 19th are shown.

Green indicates stretches of N's in each sequence, the common vertical patterns are shared amplicon dropouts across many labs.

Purple stripes indicate actual SNPs relative to the reference sequence, pink contiguous SNPs, brown deletions, and sea green insertions.

Complete sequences are sorted at the top of each month's sample. Note: A higher proportion of complete sequences are available each month

- Note: Even the early complete sequences from November accurately captured the pattern of SNPS/indels common to the Omicron lineage.

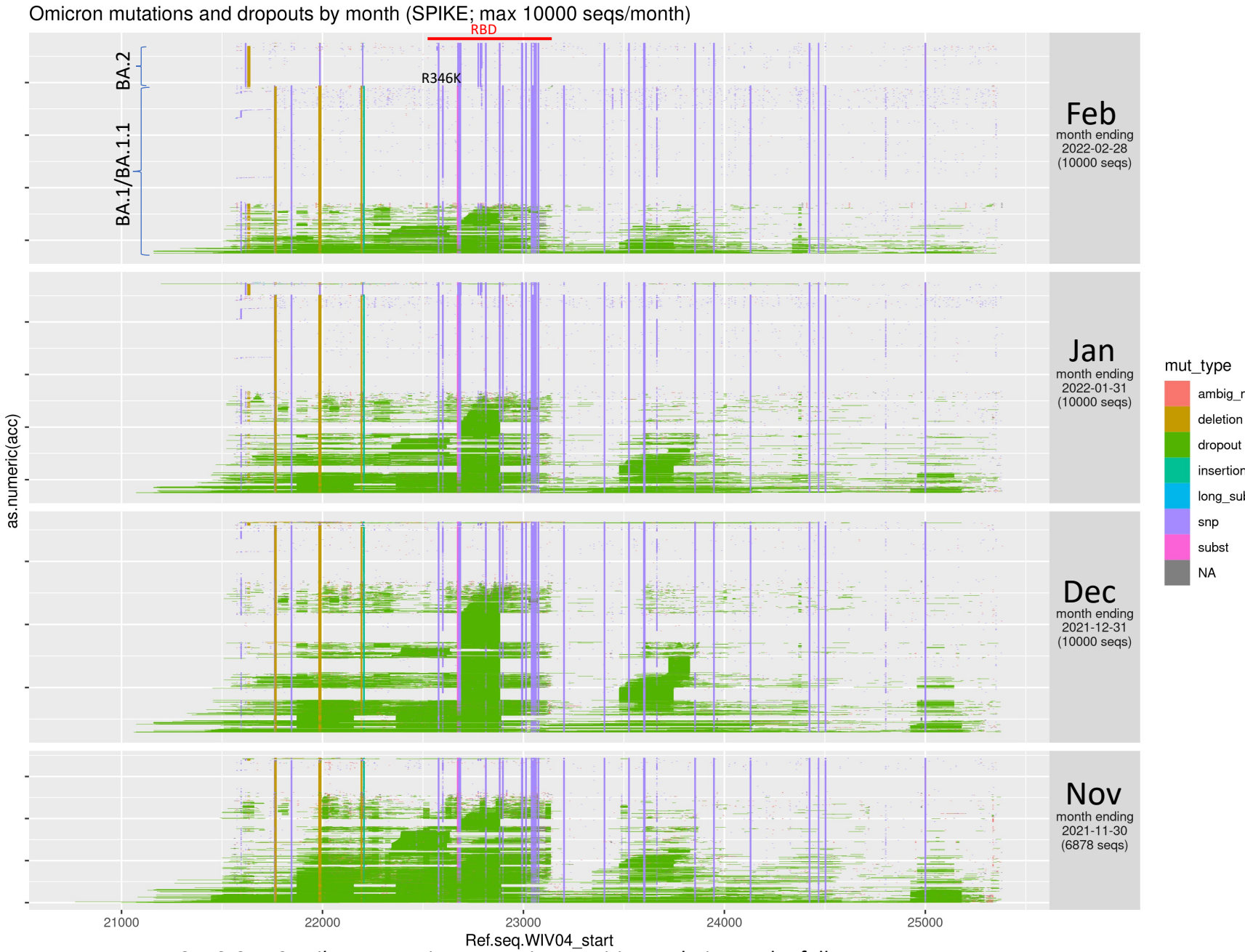


Amplicon Dropout Regions in Omicron, Spike gene expansion by sampling month

See slide 6 for a description.

Note the frequent Amplicon dropout regions in the RBD, and how the coverage frequency improves over time.

The drop out regions are also often associated with a chimeric stretches (detailed on slide 10). The chimeras are sequences with a strings of either ancestral base calls, or delta mutations.



SARS-CoV-2 *spike* gene region expansion, positions relative to the full genome

GISAID

Los Alamos NATIONAL LABORATORY

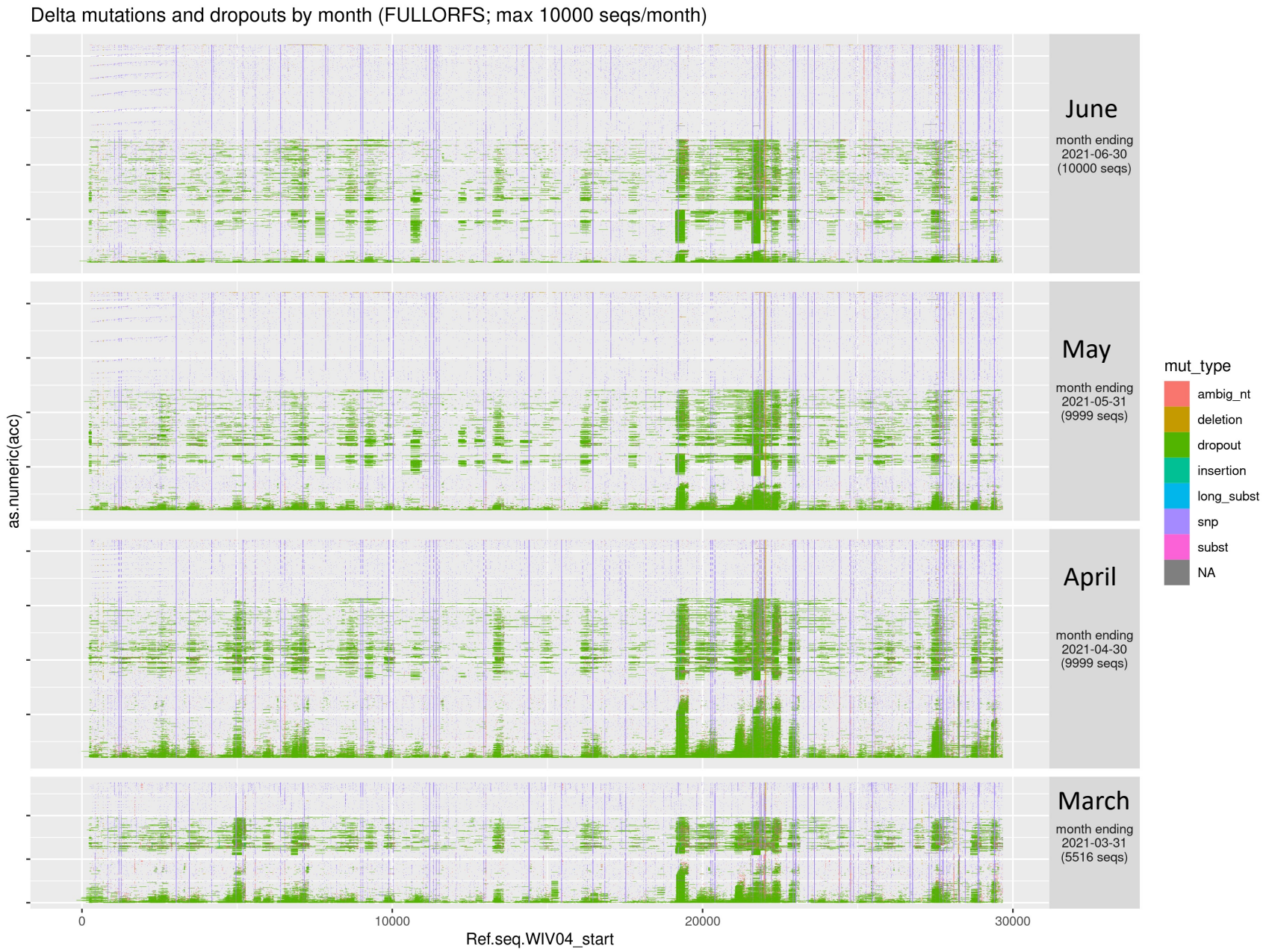
Amplicon dropout Regions in Delta sequences, full genome by sampling month March – July 2021

See slide 6 for a key.

Note:

- As with Omicron, a higher proportion of complete sequences are available each month, as the global transition to Delta was underway

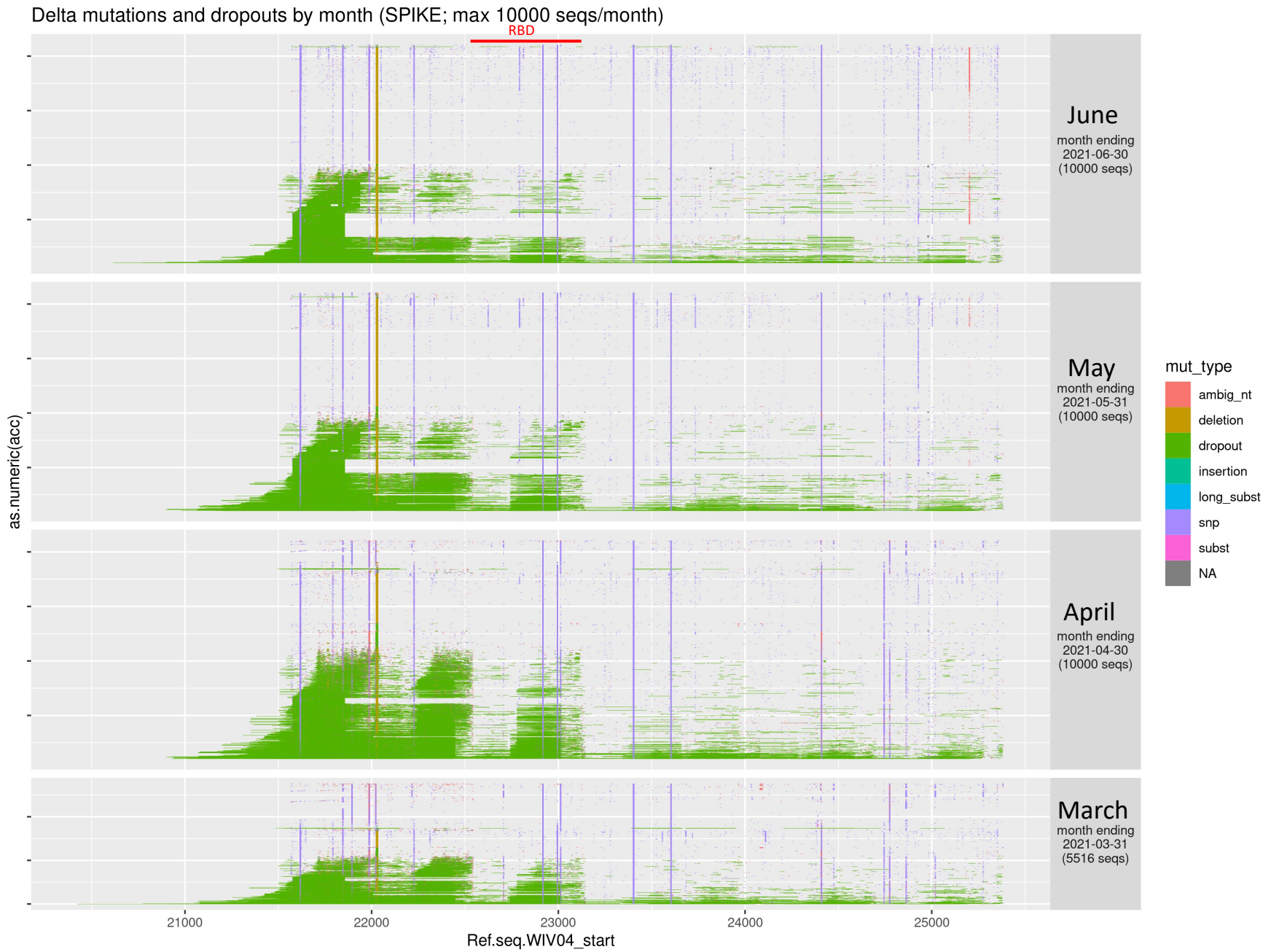
- Again, even the earliest *complete* sequences from March accurately captured the pattern of SNPS/indels common to the Delta lineage.



GISAID

Los Alamos NATIONAL LABORATORY

Amplicon Dropout Regions in Delta, Spike gene expansion by sampling month



GISAID

Los Alamos NATIONAL LABORATORY

Recombination and chimeric sequences

I. Recombination is an evolutionary mechanism used by coronaviruses

- It is possible that recombination between two distinct SARS-CoV-2 variants (e.g. Omicron and Delta) may occur within a host that is naturally co-infected. Thus, a natural recombinant form with selective advantage may be able expand in SARS-CoV-2.

II. There are two ways apparent recombination can arise in the laboratory

- If two variants are present in a sample, recombination can occur during PCR amplification, giving rise to recombinant sequences generated in vitro.
- If the dominant variant in a sample has a primer mismatch, a rare variant or low-level contaminant in the sample may be preferentially amplified, giving rise to a chimeric sequence that is an apparent recombinant.

III. It is important be aware of chimeric sequences in subsequent analyses.

- Chimeric sequences can impact conclusions based on phylogenetic analyses. Sometimes they will yield particularly long branches within a clade, sometimes they will form distinctive branches between the two clades representing the lineages from which the parents were derived. Either artifact can impact conclusions such as timing the origins of a lineage and tree-based estimates of positive selection.
- Common chimeric sequence forms could result in noisy estimates of mutational frequencies.
- If a natural recombination event was confirmed (for example, its sequence was confirmed and/or a recombinant lineage began to be transmitted and resampled in multiple geographic regions) it would be an interesting event in its own right, but could also impact phylogenetic analyses that assume no recombination.

Many chimeric sequences are evident among Omicron variants. These include (i) chimeric stretches of either ancestral or Delta sequence in an Omicron backbone, (ii) Omicron BA.1 in a BA.2 backbone, or (iii) BA.1 chimeric stretches in Delta backbones. While these chimeric sequences are likely to be *in vitro* artifacts, it is also possible that some may reflect an actual *in vivo* recombination event. We include in this update a small number of examples of (i) the most common forms of chimeras among Omicron samples, and (ii) several that would be particularly interesting *if* they were indeed found to be biological in origin. These examples are highlighted in this document, and also included in the LANL variant spreadsheet and fasta files, to alert people to their presence and provide illustrative examples.

BA.1 Chimeric sequences found >50 times among the 243,069 BA.1 sequences

Count	perc	HD	Mutation string relative to the ancestral form:
134941	55%	0	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F] **BA.1**
*5918	2.4%	3	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I, G339D,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
*1738	0.7%	3	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,S371L,S373P,S375F, S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
*1313	0.5%	5	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-, G339D,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
828	0.3%	4	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,S371L,S373P,S375F, S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H, D796Y,N856K,Q954H,N969K,L981F]
567	0.2%	20	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
395	0.2%	7	[A67V, T95I,G142D,V143-,Y144-,Y145-, G339D,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
330	0.1%	2	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-, +214EPE,G339D,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
298	0.1%	6	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-, G339D,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,A701V,N764K,D796Y,N856K,Q954H,N969K,L981F]
226	0.1%	9	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,S371L,S373P,S375F, N440K,G446S, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
172	0.1%	8	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,S371L,S373P,S375F,K417N,N440K,G446S, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
143	0.1%	5	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
95	0.0%	11	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,S371L,S373P,S375F, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
75	0.0%	19	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-, G339D, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
68	0.0%	8	[A67V, T95I,G142D,V143-,Y144-,Y145-, G339D,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,A701V,N764K,D796Y,N856K,Q954H,N969K,L981F]

BA.1 chimeric forms found >10 times carrying **Delta signature mutations.** We lowered the threshold for inclusion to 10 as Delta and BA.1 were cocirculating, so Delta chimeras have a greater potential to be biologically interesting.

*196	0.1%	4	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,S371L,S373P,S375F, L452R,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
*152	0.1%	13	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-, L452R,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
*12	0.0%	19	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE, L452R, T478K, D614G, P681R,N764K,D796Y,N856K,Q954H,N969K,L981F]
*14	0.0%	23	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-, L452R, T478K, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
20	0.0%	14	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-, G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
*29	0.0%	20	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-, L452R, T478K, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]

BA.1.1 Chimeric sequences among 211797 BA.1.1 sequences

167320	79 %	0	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
3342	1.6%	3	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F, S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
1602	0.8%	4	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F, S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H, D796Y,N856K,Q954H,N969K,L981F]
356	0.2%	5	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-, G339D,R346K,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
328	0.2%	9	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F, N440K,G446S, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
280	0.1%	8	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F,K417N,N440K,G446S,T547K,D614G, H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
227	0.1%	5	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
129	0.1%	8	[A67V,H69-,V70-,T95I, N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F, S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H, D796Y,N856K,Q954H,N969K,L981F]
104	0.0%	11	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
94	0.0%	6	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
90	0.0%	7	[A67V, T95I,G142D,V143-,Y144-,Y145-, G339D,R346K,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
85	0.0%	4	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K, N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
78	0.0%	7	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F, Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H, D796Y,N856K,Q954H,N969K,L981F]
71	0.0%	6	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F, Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
70	0.0%	5	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F, S477N,T478K, Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H, D796Y,N856K,Q954H,N969K,L981F]
64	0.0%	5	[T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,V736G,N764K,D796Y,N856K,Q954H,N969K,L981F]
58	0.0%	7	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F, N440K,G446S,S477N,T478K, T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
56	0.0%	4	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F, S477N,T478K, Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]

BA.1.1 forms found > 10 times carrying **Delta signature mutations**:

140	0.1%	4	[A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,R346K,S371L,S373P,S375F, L452R,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]

BA.2 Chimeric sequences among 58780 BA.2 sequences

47336	81 %	0	[T19I,L24S,P25-,P26-,A27-,G142D,V213G,G339D,S371F,S373P,S375F,T376A,D405N,R408S,K417N,N440K,S477N,T478K,E484A,Q493R,Q498R,N501Y,Y505H,D614G,H655Y,N679K,P681H,N764K,D796Y,Q954H,N969K]
*4712	8.0%	4	[T19I, G142D,V213G,G339D,S371F,S373P,S375F,T376A,D405N,R408S,K417N,N440K,S477N,T478K,E484A,Q493R,Q498R,N501Y,Y505H,D614G,H655Y,N679K,P681H,N764K,D796Y,Q954H,N969K]

BA.2 forms found over 10 times carrying **BA.1 signature mutations**:

34	0.1%	2	[T19I,L24S,P25-,P26-,A27-,G142D,N211-,L212I,V213G,G339D,S371F,S373P,S375F,T376A,D405N,R408S,K417N,N440K,S477N,T478K,E484A,Q493R,Q498R,N501Y,Y505H,D614G,H655Y,N679K,P681H,N764K,D796Y,Q954H,N969K]
17	0.0%	3	[T19I,L24S,P25-,P26-,A27-,A67V,H69-,V70-,G142D,V213G,G339D,S371F,S373P,S375F,T376A,D405N,R408S,K417N,N440K,S477N,T478K,E484A,Q493R,Q498R,N501Y,Y505H,D614G,H655Y,N679K,P681H,N764K,D796Y,Q954H,N969K]

Legend. Mutational strings are aligned, and ancestral stretches within the Omicron-designated chimeras are indicated by spaces. *Indicates sequences that were included in the spreadsheet and reference alignments either because they were very common, or to provide interesting examples of BA.1/Delta chimeras. Regions which span Delta signature mutations in a BA.1 or BA.1.1 backbone are highlighted in purple. BA.1 signatures in the BA.2 backbone are highlighted in red. BA.1.1 is similar to BA.1 In Spike, but adds a R346K mutation. Sequences were sampled in a 60 day period between 2021-12-22 to 2022-02-14. Counts reflect counts of the exact form.

Omicron fragments in Delta backbones

Delta-lineage baseline Spike mutations:

T19R,T95I,G142D,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N (Note: A27S, found in several chimera's listed below is commonly found with the Delta lineage.)

Omicron BA.1 lineage Spike baseline mutations:

A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,N211-,L212I,+214EPE,G339D,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F

These are the forms of Delta with stretches of Omicron sequences that were found more than one time within a distinct Delta Pango lineage. Purple are stretches of Delta, red of Omicron.

Pango designation	N_Pango	N	perc	HD	[Spike mutation strings]
AY.43	9756	2	0.0%	7	[T19R,A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,L452R,T478K,D614G,P681R,D950N]
AY.103	4244	2	0.0%	7	[T19R,G142D,E156G,F157-,R158-,N211-,L212I,+214EPE,L452R,T478K,Q613H,D614G,P681R,S691F,D950N]
AY.4	6536	2	0.0%	29	[T19R,A27S*,T95I,G142D,E156G,F157-,R158-,N211-,L212I,G339D,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
B.1.617.2	2013	2	0.1%	4	[T19R,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,L452R,T478K,D614G,P681R,D950N]
AY.126	1382	2	0.1%	6	[T19R,A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,L452R,T478K,D614G,P681R,I850L,D950N]
B.1.617.2	2013	2	0.1%	12	[T19R,A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,L452R,T478K,E554Q,D614G,H655Y,N679K,P681H,N856K,D950N,Q954H]
AY.100	1350	2	0.1%	6	[T19R,A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,L452R,T478K,D614G,P681R,D950N]
AY.39	818	2	0.2%	6	[T19R,A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,L452R,T478K,D614G,P681R,D950N]
AY.25	1152	2	0.2%	8	[T19R,A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,L452R,T478K,D614G,P681R,D950N]
AY.122	5732	6	0.1%	7	[T19R,A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,L452R,T478K,D614G,P681R,D950N]
AY.3	1434	6	0.4%	7	[T19R,A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,L452R,T478K,D614G,P681R,D950N]
AY.4	6536	7	0.1%	6	[T19R,A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,L452R,T478K,D614G,P681R,D950N]
AY.103	4244	11	0.3%	7	[T19R,A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,L452R,T478K,D614G,P681R,D950N]
AY.1	15	2	13.3%	19	[T19R,T95I,G142D,E156G,F157-,R158-,G339D,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,D950N]
AY.98.1	791	3	0.4%	7	[T19R,A67V,H69-,V70-,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,P251L,L452R,T478K,D614G,P681R,D950N]
AY.4	6536	4	0.1%	32	[T19R,A27S*,T95I,G142D,E156G,F157-,R158-,N211-,L212I,+214EPE,G339D,S371L,S373P,S375F,K417N,N440K,G446S,S477N,T478K,E484A,Q493R,G496S,Q498R,N501Y,Y505H,T547K,D614G,H655Y,N679K,P681H,N764K,D796Y,N856K,Q954H,N969K,L981F]
B.1.617.2	2013	5	0.2%	3	[T19R,T95I,G142D,V143-,Y144-,Y145-,E156G,F157-,R158-,L452R,T478K,D614G,P681R,D950N] Pango designation N_Pango N prec HD

This exact from of Spike 25 times in 6 AY Delta sublineages

Legend: This table is based on sequences sampled in the last 60 days, ending 2/14/2022. Headings: Pango designation; N_Pango is the number of sequences in our QC'd Spike sequence alignment with the noted Pango designation; N is the number of sequences with exactly the specified mutational pattern found within the Pango lineage; "perc" is the percentage of the Pango lineage with the exact form that is specified; and HD is the Hamming distance from the sequence shown and the consensus form of the particular Pango lineage specified. Purple mutations are characteristic of the Delta lineage, red the BA.1 Omicron lineage.

There were 162 *additional* distinct chimeric sequences with stretches of Omicron in Delta backgrounds that were only found once in a given Pango lineage set, in addition to the set of 17 that were found multiple times shown above.

There were no Omicron BA.2 specific mutations found in Delta backgrounds.

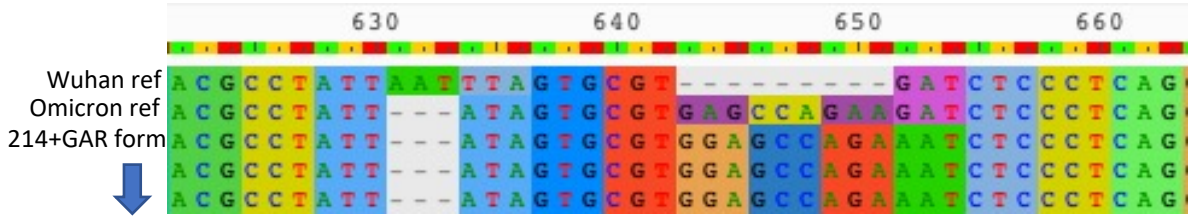
T19I,L24-,P25-,P26-,A27S,G142D,V213G,G339D,S371F,S373P,S375F,T376A,D405N,R408S,K417N,N440K,S477N,T478K,E484A,Q493R,Q498R,N501Y,Y505H,D614G,H655Y,N679K,P681H,N764K,D796Y,Q954H,N969K

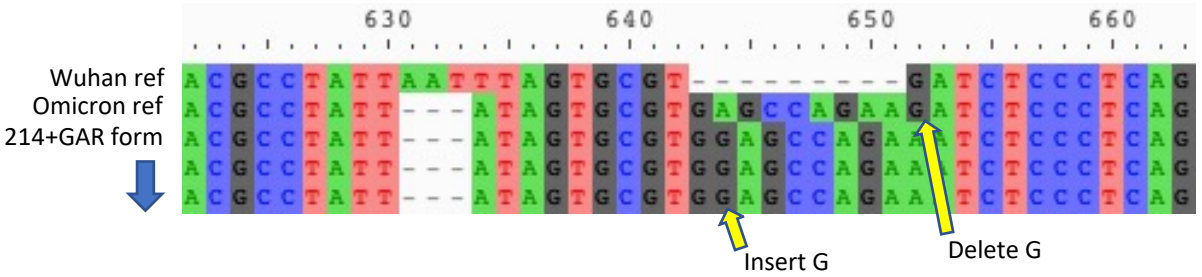
S 214 +GAR insertion form resulting from 2 mutations in BA.1 or BA.1.1, near 211-215. This distinctive form is common in many Southern Slavic nations

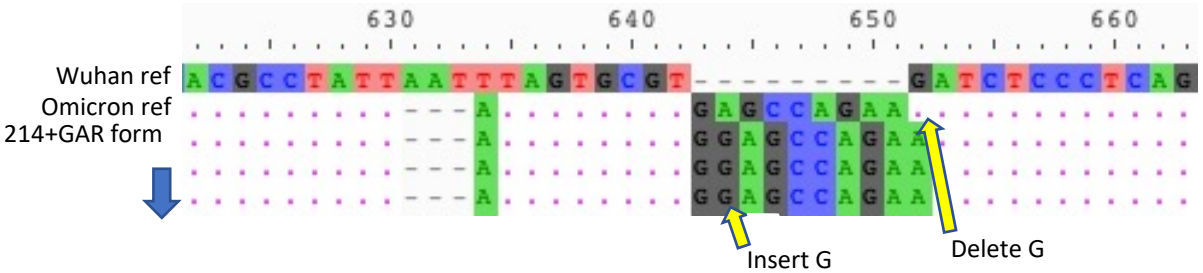
As of Feb. 24, 2022 there were 8,566 examples in GISAID.

This altered insertion is found in
BA.1 >2,000 and BA.1.1(with R346K) >2,000

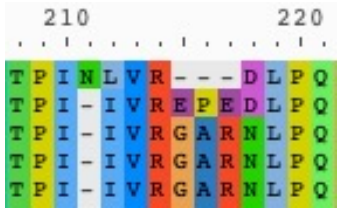
Three views of the codon aligned nucleotide region of Spike







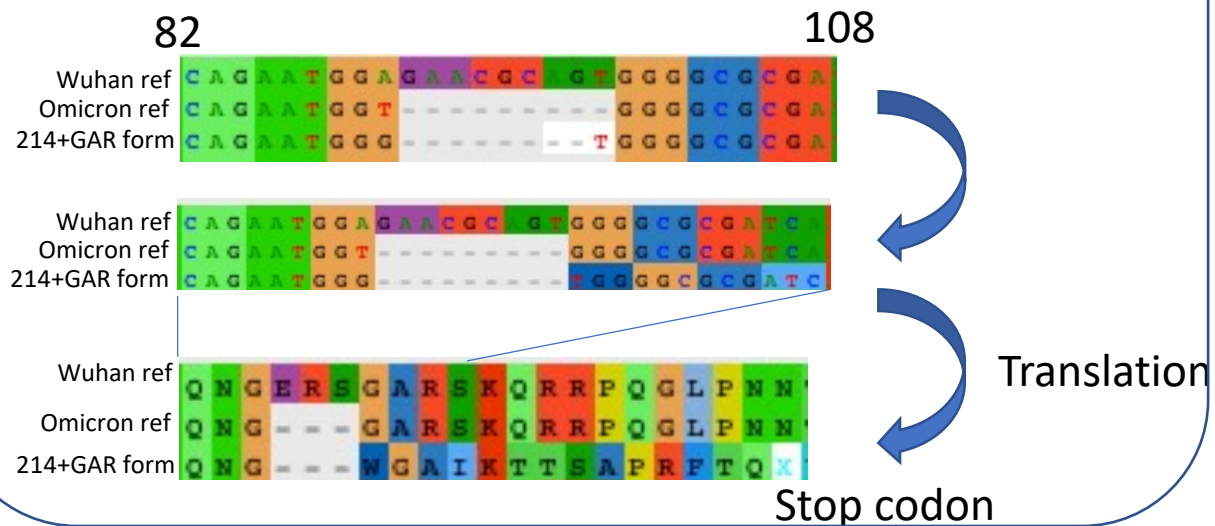
Translation



LANL Alignment, amino acid changes

N211-,L212I,+214GAR,D215N

This variant also introduces a frameshift in the nucleocapsid protein at amino acid 34, where there is usually a 3 amino acid deletion in Omicron:

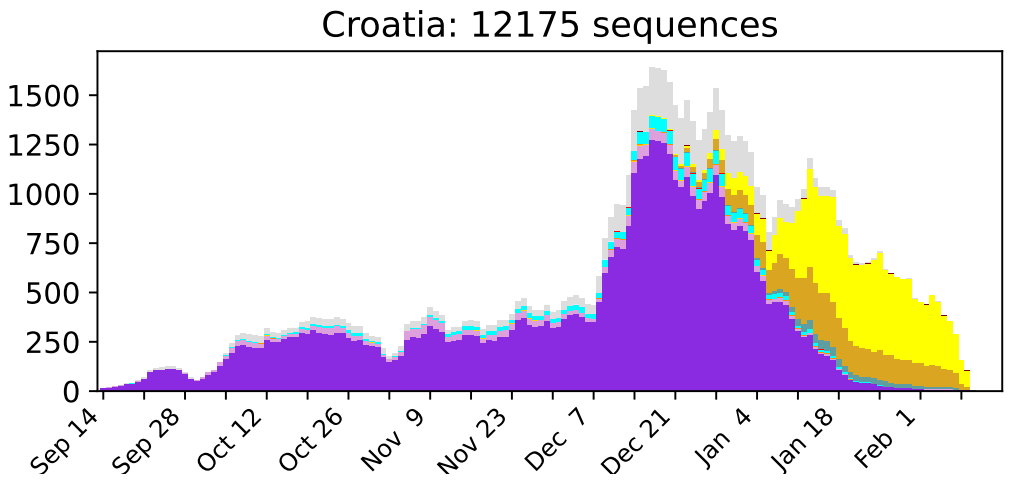


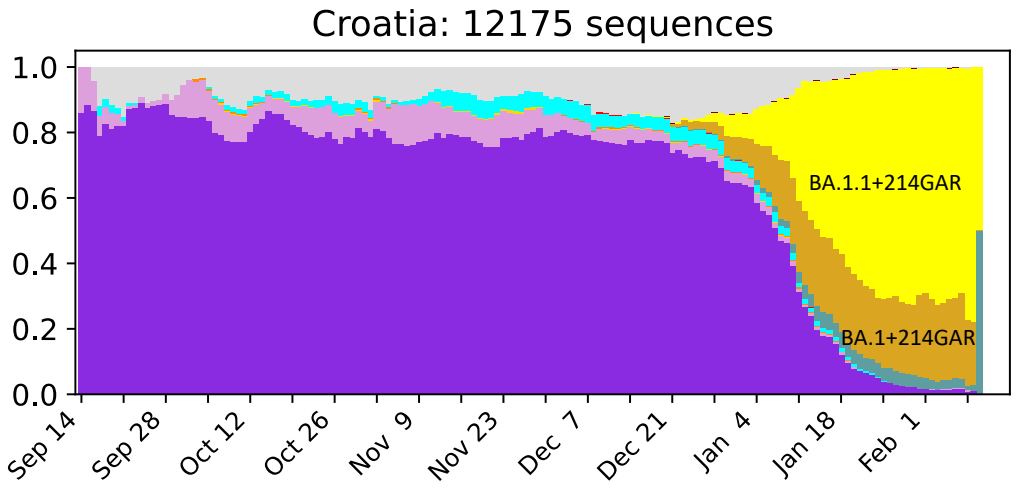
GISAID

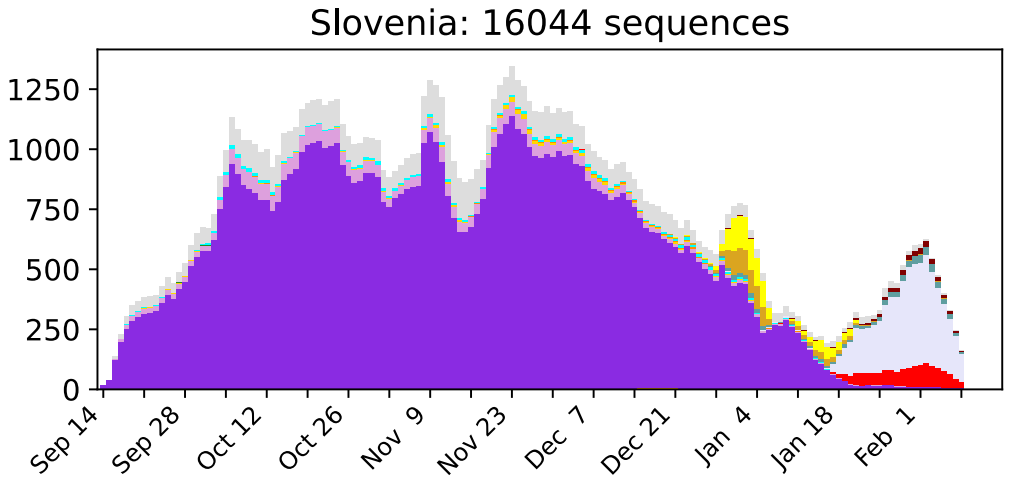
Los Alamos NATIONAL LABORATORY

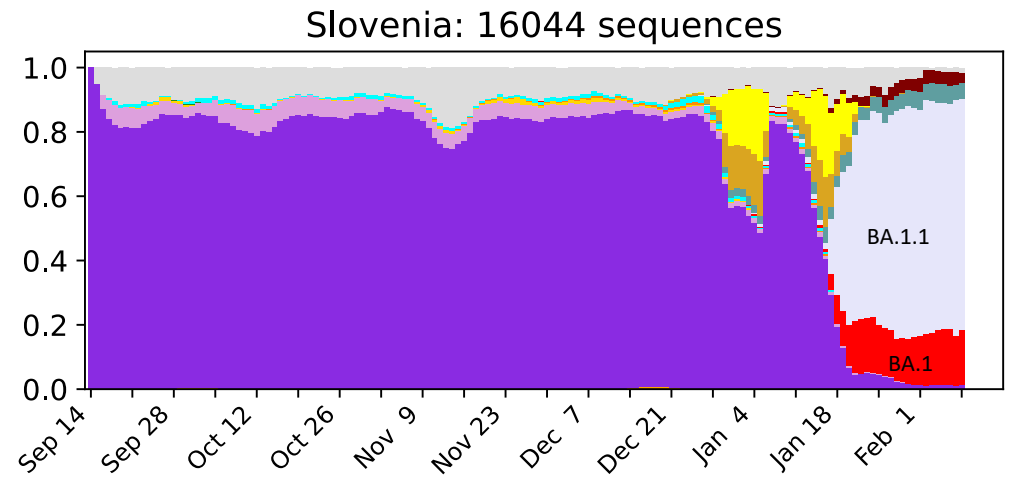
The high frequency of sampling of the S 214+GAR from in the Southern Slavic countries is consistent with founder effects or possible sequencing issue:

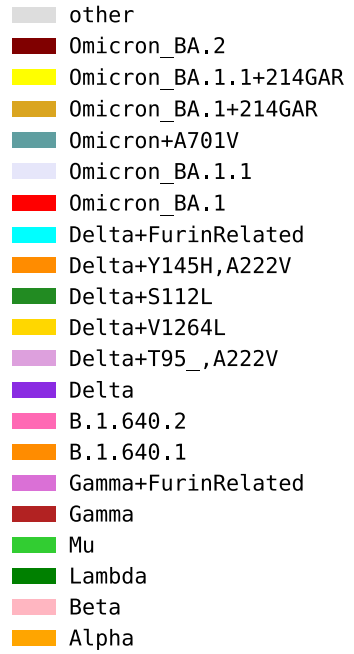
The S 214+GAR variant dominated the initial Omicron expansion in both Croatia and Slovenia, and was found both in BA.1.1 and BA.1 backgrounds. In Slovenia, these have been replaced by more conventional Omicron BA.1.1 and BA.1. forms.
It is possible that this form is a sequencing artifact, particularly given the frameshifted/truncated Nucleocapsid.
BA.2 has only recently been introduced into the geographic region.











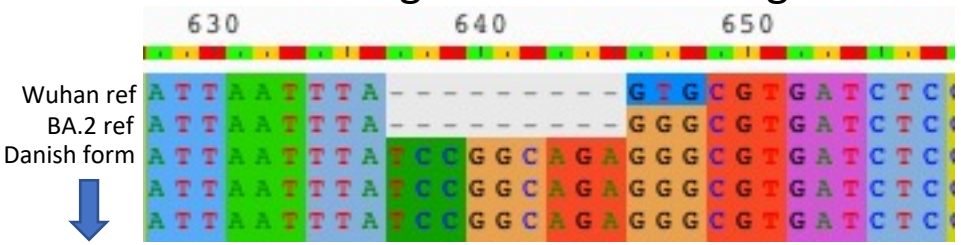
With thanks to the sequencing teams in Croatia and Slovenia for sharing their data.

GISAID

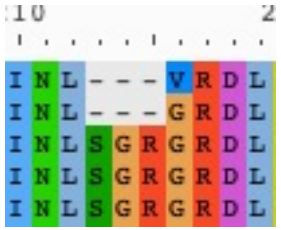
Los Alamos NATIONAL LABORATORY

Insertion S 212 +SGR in BA.2 is found in only Denmark, and though found at a very low level it is increasing locally

Codon aligned nucleotide region



Translation



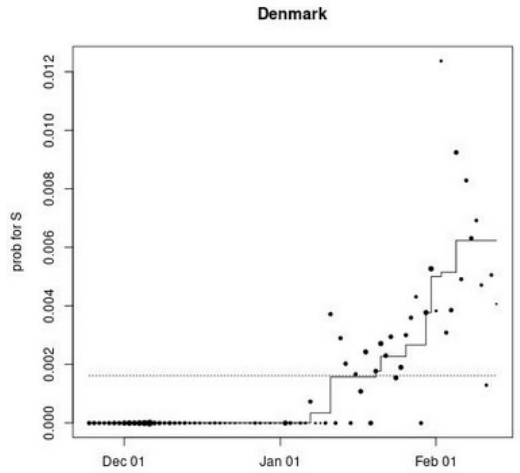
LANL Alignment, amino acid changes relative to BA.2

+212SGR

BA.2 + 212SGR: Found only in Denmark, at very low but increasing levels

Country level

	# S	# Others	Total	S/Total (%)	# days	Time window	p-val
Denmark	159	98439	98598	0.16	82	81	0.00249



With thanks to the sequencing teams in Denmark for sharing their data.

While this BA.2 insertion is still rare, it is interesting as Omicron is sampling indel variants in this region, and other variants have also carried three amino acid insertions in this region:

Summary of past insertions in this region

BA.1: +214 EPE	Global transition
BA.1: loss of 214 insertion	May be a sequencing artifact
BA.1: +214 EPE change to +214 GAR	~5000 times
BA.2: +212 SGR	~100 times

Other lineages with 212-215 insertions

B.1.214.2 +214 TDR	-- Belgium, France ~ a thousand
A.2.5. +214 AAG (+ D215Y)	-- US, central America, a few thousand