Delta variants, C.1.2, Mu, P.1

Sept. 11, 2021 GISAID data feed provided to cov.lanl.gov

Bette Korber and James Theiler

“Delta variant” variants

We have highlighted 52 variants of Delta that seemed of particular interest as they had at least one or a combination of the following features:

- Common
- Consistently increasing in frequency compared to baseline Delta in multiple geographic regions
    - Slide 3 shows a mutation that shows such potential for positive selection
    - Slide 4 shows a counter example, a common mutation that is not consistently increasing, so its relatively high sampling is more likely to be founder effects or sampling artifacts.
    - Many of the variants listed are just found in a small number of countries, and will continue to monitor them.
- In an interesting region: the RBD, NTDss, or potentially related to furin cleavage

Note: Site Spike 142 in Delta. As a cautionary note: It has become apparent through various analyses that the G142 ancestral mutation of the spike in recent lineages, such as Delta, could be an artefact of sequencing protocols. Almost every new mutation that is common in Delta (found over 300 times in the 5/26/2021 – 8/24/2021 GISAID sample from 8/24/2021) comes in two flavors of Spike, some with the G142D mutation, and some with the ancestral G142 in this position. A slide illustrating examples of this is shown in slide 5. Given this, we are at least for the present, “setting aside” the G142 ancestral versions of Delta variants from our list, and unless a mutational variant is reported with only the ancestral form of G142 associated with it, we choose the G142D version of the variant Spike as the preferred form of the Delta variant.

In our current counts for a variant forms in Delta, we do not split the data by 142 status, and so do not risk undercounting new variant forms in our tallies. For example: in the last 60 days we find the form of Spike with this basic Delta backbone pattern: T19R,**G142D**,E156G,F157-,R158-,L452R,T478K,D614G,P681R,D950N exactly 63571 times, and the form of Spike without G142D: T19R,E156G,F157-,R158-,L452R,T478K,D614G,P681R,D950N exactly 7467 times, and sp the count of the basic form to be ~ 63,571 + 7,467 = 71,038 times; the G142D toggling is not considered.

With thanks to Elodie Ghedin for leading the NIH working group discussion on Spike site 142, and to Jim Davis for sharing his data and a very helpful talk.

Example of a mutational pattern suggesting possible increase compared to other Deltas: T250I

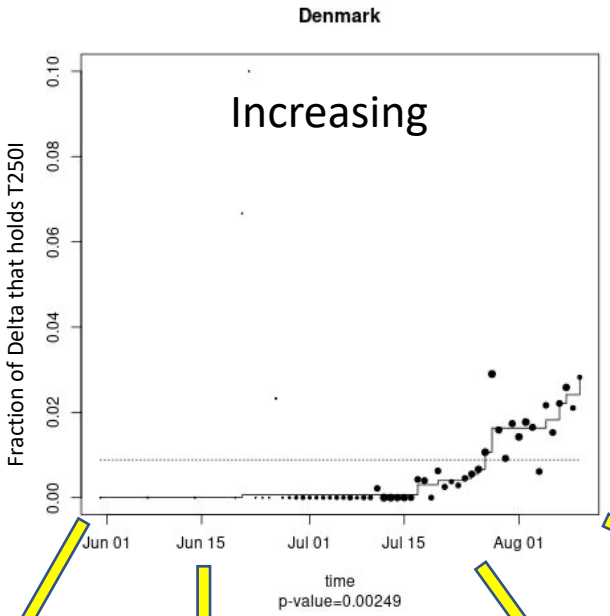
counts

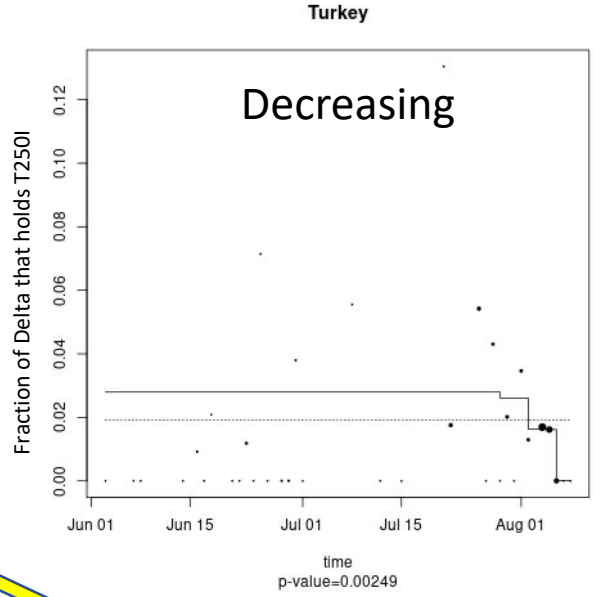
```
182        [T19R,T29A,      E156-,F157-,R158G,T250I,L452R,T478K,D614G,P681R,D950N]
380        [T19R,T29A,G142D,E156-,F157-,R158G,T250I,L452R,T478K,D614G,P681R,D950N]
```

T250I had a significant presence in 11 countries, it was increasing in 8, decreasing in just 1, relative to other Delta's.

Position: SPIKE 250 I, SPIKE 29 A				
	# IA	**# Others**	**p increasing**	**p decreasing**
Belgium	47	2947	0.00249	0.98507
Denmark	166	18564	0.00249	0.94527
France	40	7184	0.00498	0.99254
Germany	105	4477	0.00249	0.93781
Netherlands	58	7331	0.00498	0.37811
Norway	10	881	0.00249	0.99502
Sweden	66	4639	0.00249	0.99751
Switzerland	20	3732	0.39552	0.37811
Turkey	104	5309	0.26617	0.00249
USA	23	57228	0.55721	0.20647
United-Kingc	31	150563	0.00249	0.91045

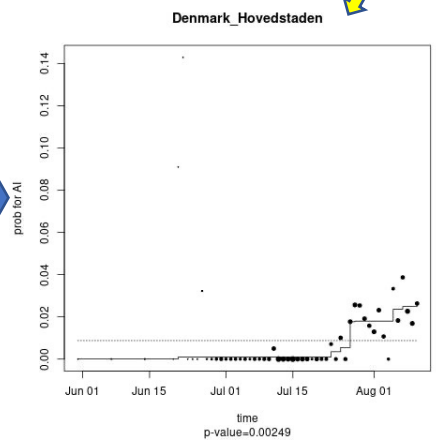
Examples

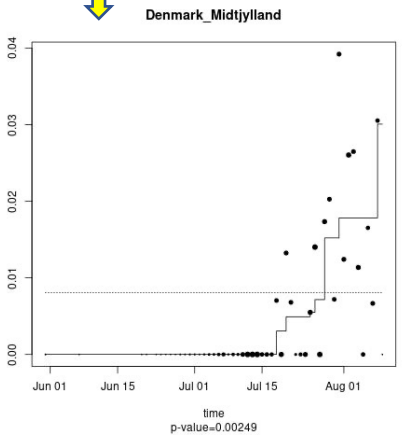


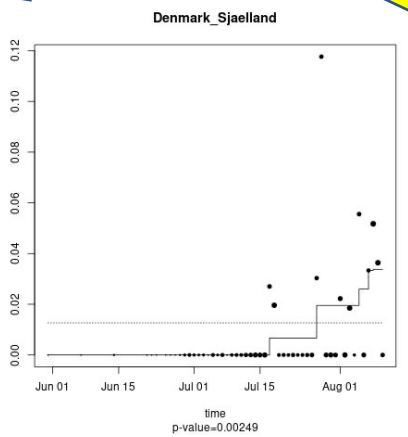


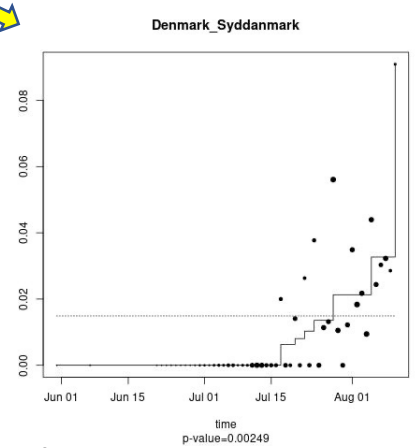
0 to ~4% mid-July to to mid-August

State level Position: SPIKE 250 I, SPIKE 29 A			
	# AI	**# Others**	**p-val**
Denmark_Hovedstaden	69	7805	0.00249
Denmark_Midtjylland	41	5045	0.00249
Denmark_Sjaelland	17	1324	0.00249
Denmark_Syddanmark	37	2448	0.00249









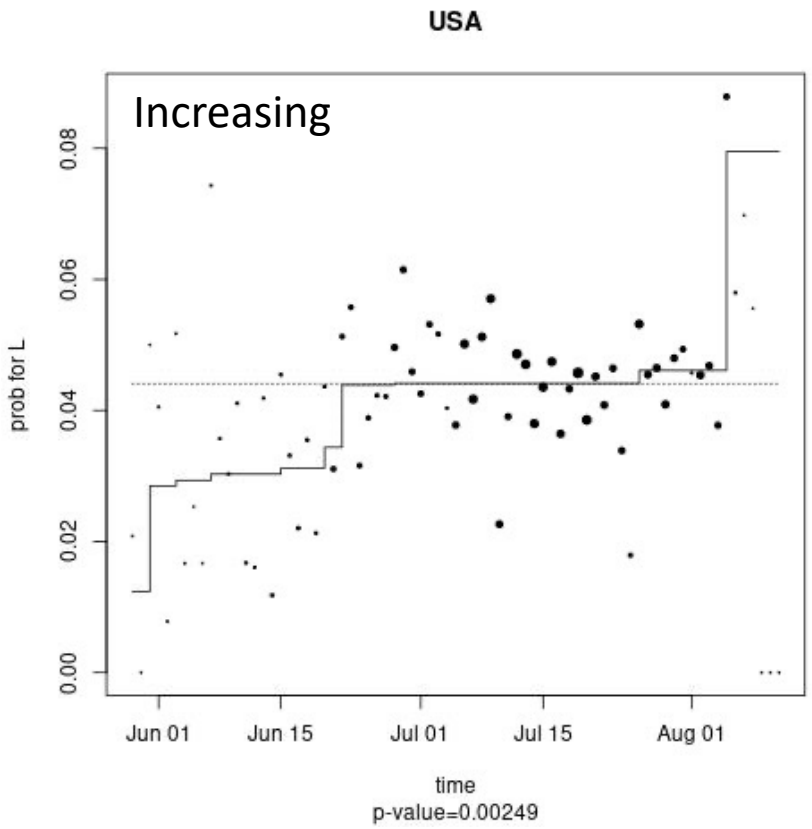
Hypothesis raising that T250I might be interesting, more data is needed, but this is a good candidate for exploratory testing

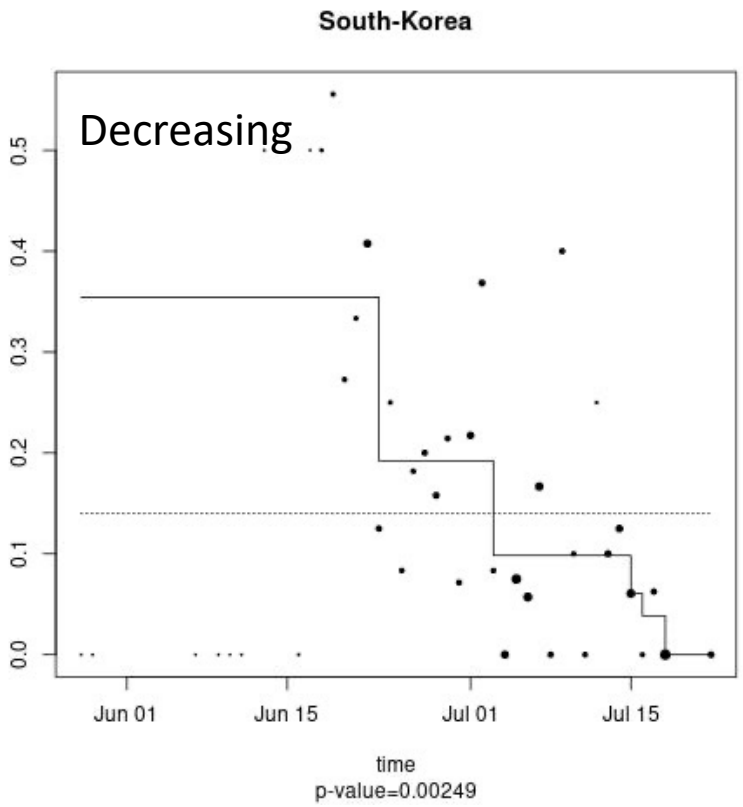
V1264 is a common mutation, but is likely founder/sampling effects, as it sometimes increases, sometimes decreases, and often is just maintained relative to other Deltas in the 19 countries it was found more than 10 times.

```
counts
2215 [T19R,G142D,E156-,F157-,R158G,      L452R,T478K,D614G,P681R,D950N,       V1264L]
 213 [T19R,      E156-,F157-,R158G,      L452R,T478K,D614G,P681R,D950N,       V1264L]

744  [T19R,G142D,E156-,F157-,R158G,A222V,L452R,T478K,D614G,P681R,D950N,       V1264L]
346  [T19R,      E156-,F157-,R158G,A222V,L452R,T478K,D614G,P681R,D950N,       V1264L]

569   [T19R,G142D,E156-,F157-,R158G,A222V,L452R,T478K,D614G,P681R,D950N,N1074S,V1264L]
318   [T19R,      E156-,F157-,R158G,A222V,L452R,T478K,D614G,P681R,D950N,N1074S,V1264L]
```

Countries

	p increasing	p-decreasing
Australia	0.00249	0.80348
Belgium	0.25622	0.92289
Denmark	0.05224	0.06219
France	0.84826	0.00249
Germany	0.00746	0.68657
Hong-Kong	0.93781	0.00249
India	0.57463	0.16169
Indonesia	0.0199	0.90796
Israel	0.16418	0.16667
Japan	0.00249	0.02488
Mexico	0.99751	0.00249
Netherlands	0.43781	0.6393
Singapore	0.00249	0.99502
South-Korea	0.6194	0.00249
Sweden	0.02239	0.53234
Switzerland	0.83085	0.00498
Turkey	0.00249	0.00498
USA	0.00249	0.92786
United-Kingdom	0.00249	0.88308





This suggests that V1264 is a less interesting candidate for exploratory testing

Small alignment of the most variable positions in Spike, position numbers are written vertically.

To represent these variants, we would choose the version with G142D present

Few NTDss and RBD *common* changes

NTDss RBD

XSpike run on just Deltas: the most common forms with I95T

The most common forms of Spike, those found > 300 time in the past 3 months.

G142D toggles in association with almost all of them… only one exception, D138H.

```
                                                                                11111111
                 11111111222222344455566667778890111222 2
 222566779913445558245558014504717880195757026135 6
5179167075728256781280139876217237082910904442972 4
Global cseqs= 246849
LRATTHAVKTKSDGYEFRGAYTPDVVKGLNTTQQSAETTIADNVGPGMSV  Global     UK  Eu-UK  NAmer   Asia Africa  SAmer  Ocean  Local  Exact  Pct [Context]
                                                    246849  98889  78696  53631   9404    441   4724   1064 201341 <---------- Total
001 .........I...D..--G.........R............N........  83021  69888   8841   3093    727      1     59    412  82988  69985  84% [T19R,T95I,G142D,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N]
002 .........I.....--G.........R............N........   6973    706   4275   1298    183      0     10    501   6964   5747  82% [T19R,T95I,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N]

003 .........I..HD..--G.........R............N........   1102   1043     31     26      2      0      0      0   1102    825  74% [T19R,T95I,D138H,G142D,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N] *

004 .........I...D..--G.........R............*........     96      1     82      5      6      0      0      2     73     57  78% [T19R,T95I,G142D,E156-,F157-,R158G,L452R,T478K,D614G,P681R]
005 .........I.....--G.........R............*........    435     41    248    129      6      0      2      9    423    344  81% [T19R,T95I,E156-,F157-,R158G,L452R,T478K,D614G,P681R]

006 .........I...D..--G.........R........I...N........   1077     19   1034     15      1      0      8      0   1075   1002  93% [T19R,T95I,G142D,E156-,F157-,R158G,L452R,T478K,D614G,P681R,T719I,D950N]
007 .........I.....--G.........R........I...N........    627     12    610      5      0      0      0      0    627    591  94% [T19R,T95I,E156-,F157-,R158G,L452R,T478K,D614G,P681R,T719I,D950N]

008 .........I...D..--G.........R............N...S....    771    758      9      1      3      0      0      0    771    719  93% [T19R,T95I,G142D,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N,P1162S]
009 .........I.....--G.........R............N...S....     14      7      6      0      1      0      0      0     14     14 100% [T19R,T95I,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N,P1162S]

010 ....I....I...D..--G.........R............N........    618    609      9      0      0      0      0      0    618    574  92% [T19R,T51I,T95I,G142D,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N]
011 ....I....I.....--G.........R............N........      7      6      1      0      0      0      0      0      7      7 100% [T19R,T51I,T95I,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N]

012 .......F.I...D..--G.........R............N........    544    536      7      0      1      0      0      0    544    492  90% [T19R,V70F,T95I,G142D,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N]
013 .......F.I.....--G.........R............N........      9      3      5      0      1      0      0      0      9      9 100% [T19R,V70F,T95I,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N]

014 .........I...D..--G..H......R............N........    474    399     74      0      1      0      0      0    474    421  88% [T19R,T95I,G142D,E156-,F157-,R158G,Y248H,L452R,T478K,D614G,P681R,D950N]
015 .........I.....--G..H......R............N........     14      4     10      0      0      0      0      0     14     10  71% [T19R,T95I,E156-,F157-,R158G,Y248H,L452R,T478K,D614G,P681R,D950N]

016 F........I...D..--G.........R............N........    458    397     38     20      3      0      0      0    458    395  86% [L5F,T19R,T95I,G142D,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N]
017 F........I.....--G.........R............N........     27      1     16      8      2      0      0      0     27     18  66% [L5F,T19R,T95I,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N]

018 .........I...D..--G.........R............N.I......    457     86     76    274     19      0      2      0    457    366  80% [T19R,T95I,G142D,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N,V1104L]
019 .........I.....--G.........R............N.I......    386     19    102    261      3      0      0      1    386    265  68% [T19R,T95I,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N,V1104L]

020 .....Y...I...D..--G.........R............N........    342      7    323      1      0     11      0      0    342    302  88% [T19R,H66Y,T95I,G142D,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N]
021 .....Y...I.....--G.........R............N........     68     11     57      0      0      0      0      0     68     58  85% [T19R,H66Y,T95I,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N]

022 .........I...D..--G.........R..........L.N........    339      0    318     16      5      0      0      0    339    276  81% [T19R,T95I,G142D,E156-,F157-,R158G,L452R,T478K,D614G,P681R,I850L,D950N]
023 .........I.....--G.........R..........L.N........    188      0    180      6      2      0      0      0    188     93  49% [T19R,T95I,E156-,F157-,R158G,L452R,T478K,D614G,P681R,I850L,D950N]

024 .........I...D..--G.........R....H.......N........    301     62    199     31      7      0      2      0    284    208  73% [T19R,T95I,G142D,E156-,F157-,R158G,L452R,T478K,D614G,Q677H,P681R,D950N]
025 .........I.....--G.........R....H.......N........    231      4    218      4      0      0      5      0    231    173  74% [T19R,T95I,E156-,F157-,R158G,L452R,T478K,D614G,Q677H,P681R,D950N]

026 .............D....................R............N........     63      0     20      6     37      0      0      0     61     50  81% [T19R,G142D,L452R,T478K,D614G,P681R,D950N] **
027 ...............................R............N........    365      1    196     32    136      0      0      0    296    219  73% [T19R,L452R,T478K,D614G,P681R,D950N]
```

How to read this

```
Local   Exact Pct [Context]
6964    5747  82%  [T19R,T95I,E156-,F157-,R158G,L452R,T478K,D614G,P681R,D950N]
```

This mutation list refers to an exact from of Spike that carries this pattern of mutations. It was found exactly 5747 times, and accompanied by additional mutations 6964 times.

Range of dates: 2021-05-26 - 2021-08-24

Pango linage designation in GISAID (version: 2021-07-28): AY.1, AY.10, AY.11, AY.12, AY.2, AY.3, AY.3.1, AY.4, AY.5, AY.6, AY.7, AY.8, AY.9, B.1.617.2

*Only form with > 300 samples to *not* have an ancestral G142 toggle.

C.1.2
Which forms to use?

It is still quite rare and over half the Spike sequence have large gaps.

Bette Korber & James Theiler

Sept 1 GISAID data

Please see the new paper from the South African group regarding their discovery of this variant:

Scheepers et al. The continuous evolution of SARS-CoV-2 in South Africa: a new lineage with rapid accumulation of mutations of concern and global detection. https://www.medrxiv.org/content/10.1101/2021.08.20.21262342v2

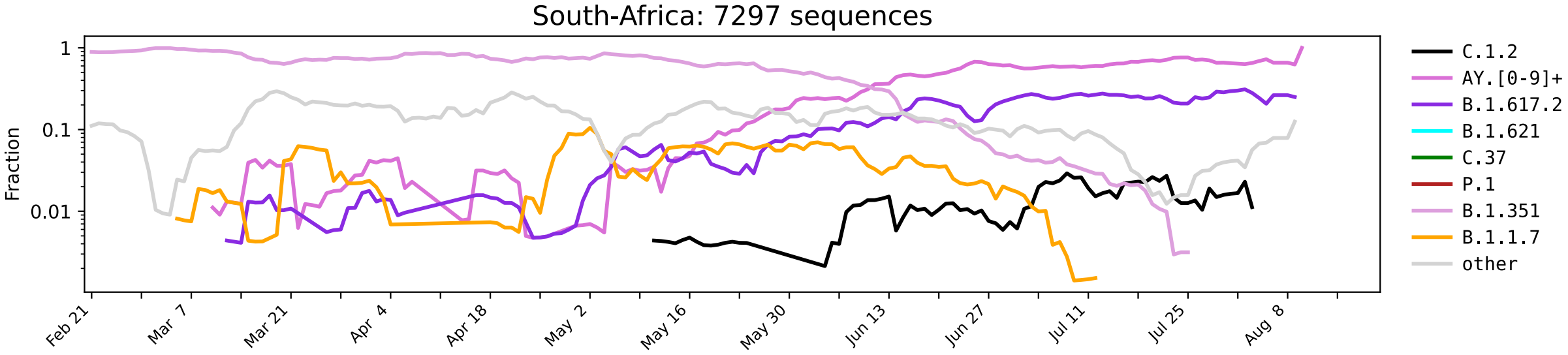
The mutations included in the listing to represent C.1.2 were based on very early sampling, as people were requesting them. It is very likely the Spike mutational listing will need updating as more data accrues. We were asked to select two for the NIH variant working group, so we picked the most common form in South Africa, and a still very rare variant found in the UK and Europe that had interesting amino acid substitutions.

This lineage is only sampled much in S. Africa, and even there still quite rare.

But it has also been detected in the UK, Portugal, Switzerland, Mauritius, Zimbabwe, Botswana, the DRC, Shenzhen.

In South Africa, C.1.2 sampling has come up very slightly. These plots of the dynamics are based on the Pango associated names in GISAID and inclusive of all data as of Sept. 1. 2021.

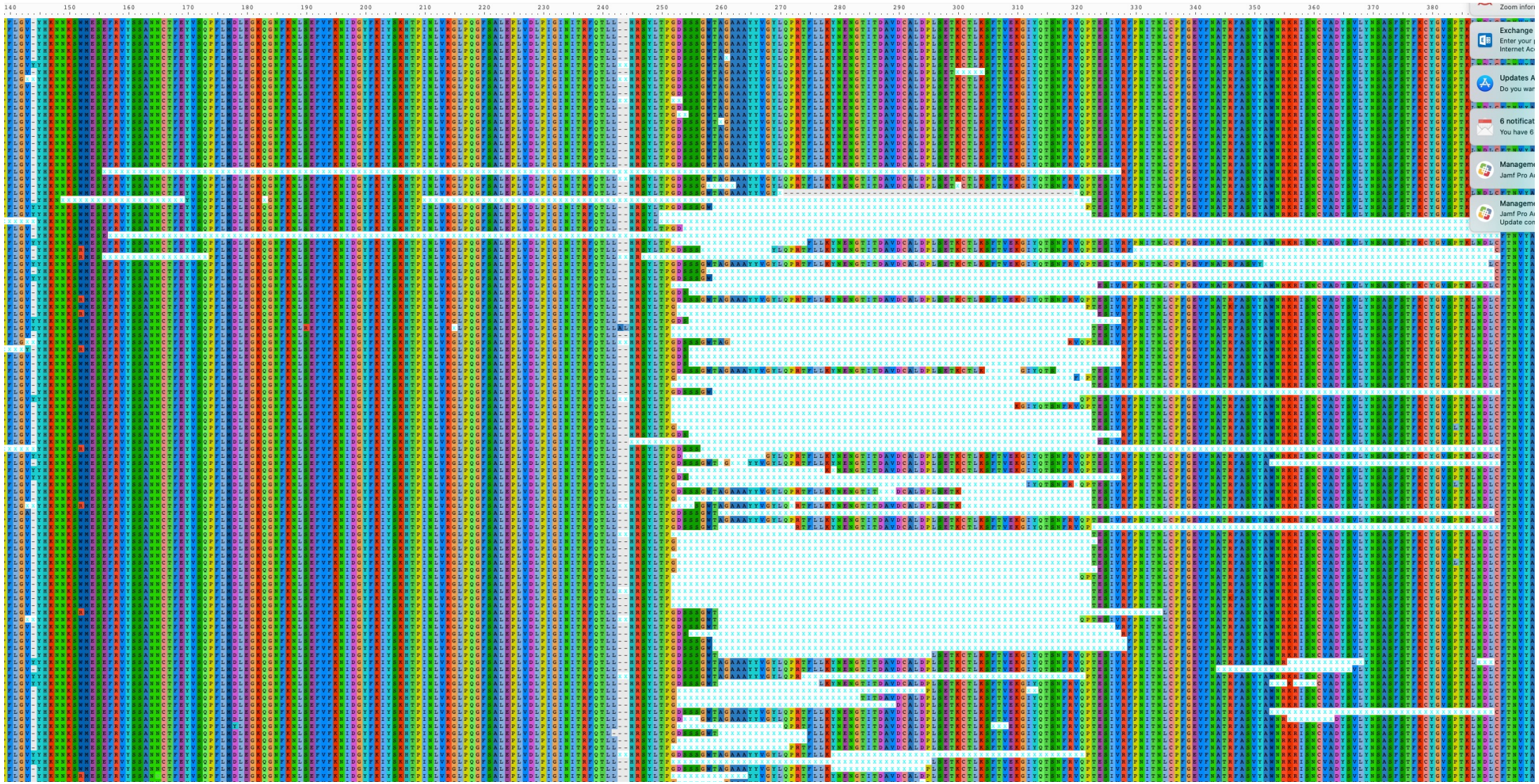
Beta (B.1.351) prevalence in S. Africa transitioned to Delta prevalent (B.1.617.2 and AY's) in early June.
C.1.2 (black) has become more frequent over the summer, but is still quite rare (black). The plot is weekly average of sampling frequency, log scale, and highlights some forms we are tracking. We just split Delta on B.1.617.2 and AY's for convenience.



The Spike region between 150 and 410 is often missing data in C.1.2 lineage sequences sampled from GISAID 9/1/2021. The light blue X's are unresolved sequence in this amino acid alignment. We factor in only called bases to arrive at the consensus form, and then picked a variant that matched the consensus to try to be representative. --bk.

Snapshot from C.1.2 Spike protein alignment

150 410



Only 7 of the 126 C.1.2 at the time sampling (9/1/2021) in GISAID passed LANL quality control in the complete genome, intact through the whole genome. Here are their Spikes. We choose the full genome from among these, picking the relevant Spike based on the complete alignment of all 126 C.1.2 available from GISAID on Sept 1.

The 7 QC passed C.1.2 have these 5 forms of Spike:

1. C.1.2	7	2	28.6%	0	[P9L,P25L,C136F,Y144-,R190S,D215G,A243-,L244-,N440K,Y449H,E484K,N501Y,D614G,H655Y,P681H,T716I,T859N,D936H]	Most interesting variant.
2. C.1.2	7	2	28.6%	5	[P9L,P25L,C136F,Y144-,R190S,D215G,A243-,L244-,Y449H,E484K,N501Y,L585F,D614G,H655Y,N679K,T716I,T859N]	
3. C.1.2	7	1	14.3%	5	[P9L,C136F,Y144-,R190S,D215G,A243-,L244-,Y449H,E484K,N501Y,D614G,H655Y,N679K,T716I,T859N]	Backbone of conserved sites.
4. C.1.2	7	1	14.3%	6	[P9L,C136F,Y144-,R190S,D215G,A243-,L244-,Y449H,T478K,E484K,N501Y,D614G,H655Y,N679K,T716I,T859N,D936N]	
5. C.1.2	7	1	14.3%	6	[P25L,C136F,Y144-,R190S,D215G,A243-,L244-,P384L,Y449H,E484K,N501Y,D614G,H655Y,N679K,T716I,T859N]	

Form 3 is the backbone of C.1.2 , both the Spike consensus and the most common form.
GISAID example: EPI_ISL_3132623

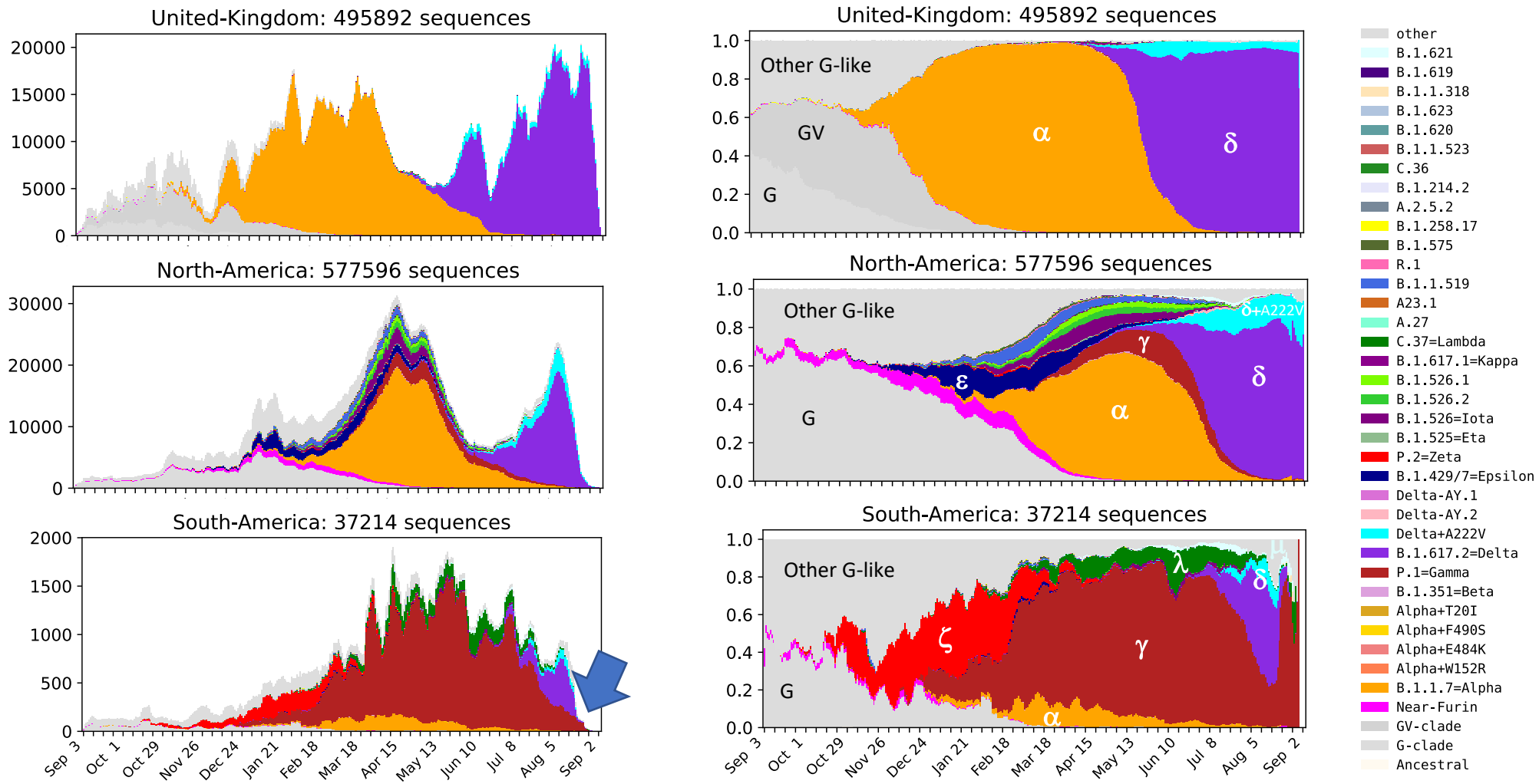
Form 1 has an additional mutation in the RBD, and is found internationally, UK, Portugal, Mauritius.
GISAID example: EPI_ISL_2989113
It also has N440K and has switched out N679K for the more common P681H, both of these mutations may augment Furin cleavage, and the N679K is the more common in C.1.2. All C.1.2 have one or the other, not both.

Notes on frequencies of particular Spike mutations in 126 C.1.2 that arise in addition to form 3 above:

P25LP L 49%
N440K K is rare 5%, but is international (UK, Portugal, as well as South Africa
T478K K is occasional, 20%
L585F is occasional, 19%
879T T is rare, 9%
D936H H is rare, 6%, only one N
P384L L is rare

P681H rare N679K is much more common, it is one or the other.

Mu and Gamma seem to be sampled with increasing frequency recently in South America.

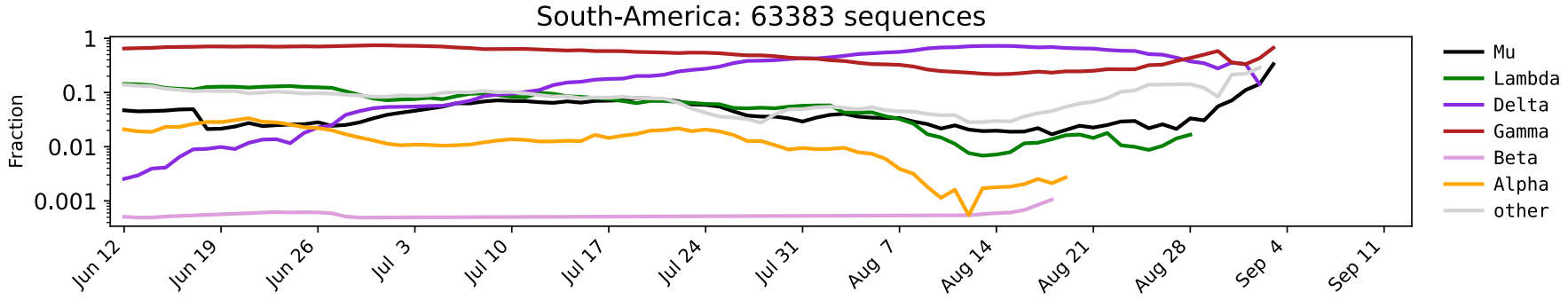


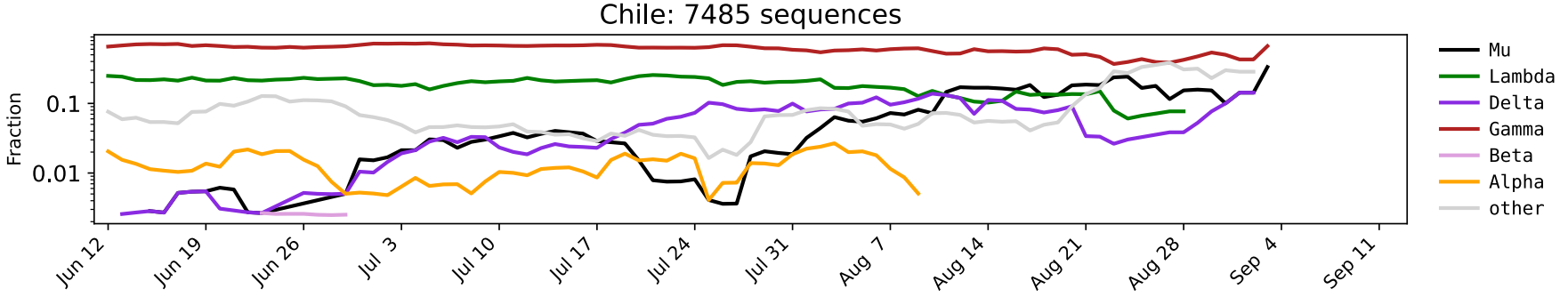
This may be an artifact due to very limited sampling the most recent weeks (see blue arrow), more time is needed to resolve this, but this situation merits continued monitoring to see if sampling resolves to increases in Delta.

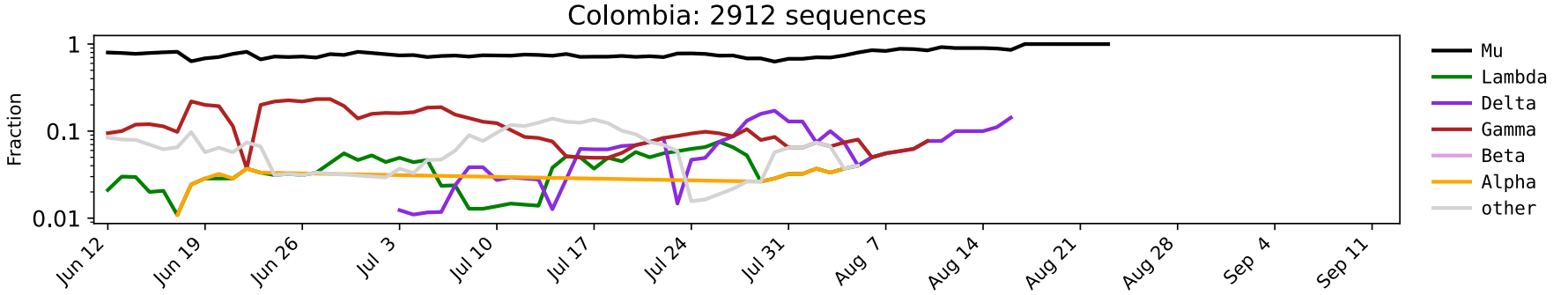
The next two slides take a closer look at the Mu and Gamma variants that are persisting in Chile, Columbia, and Brazil. Interesting variant forms of Mu and Gamma are available in the 9/11/2021 listing.

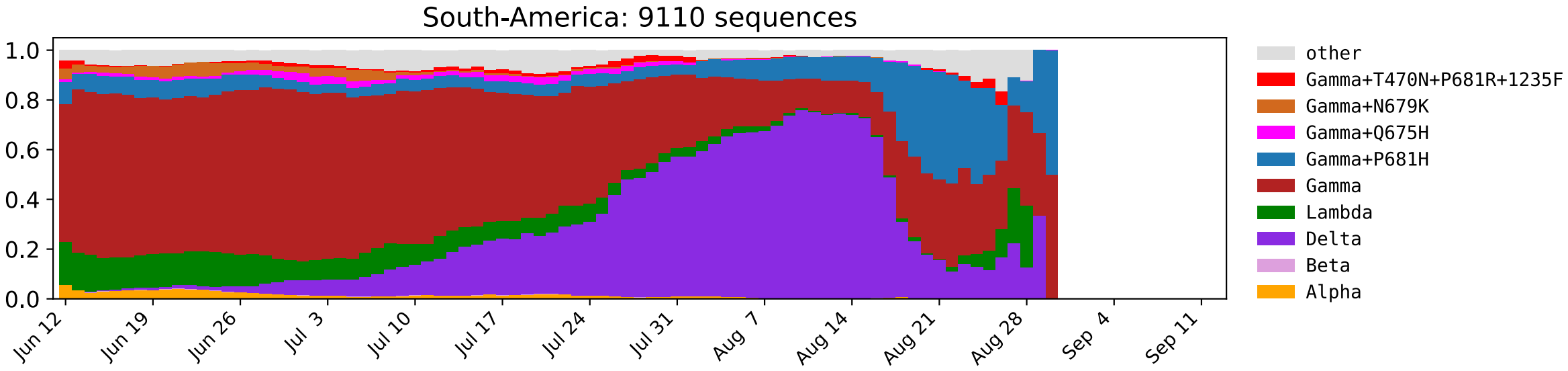
Mu: B.1.621

-- Mu is most common in the America's.
-- Mu was increasing in some parts of the United States through mid-July, but appears to be diminishing in sampling frequency as Delta has increased.
-- Mu and variants are still common in S. America, particularly in Chile and Columbia.







Variants of P.1 (Gamma) that carry increasing positive charge relative to base line P.1, particularly that have added P681H, are becoming increasing common among recent Brazilian samples, and persisting along with Delta. The numbers are still very small in recent sampling, and these Gamma mutations have some regional associations within Brazil, so the overall frequencies in each time window of each Gamma form are biased by local sampling issues.

This situation merits continued monitoring, to determine if the transition to Delta will continue in Brazil, or if novel P.1's may be competitive.