

## Week 12

---

1. Create covid4.md and use file  
metadata\_snpeff\_tidy\_1000K\_downsampled.rds

2. dNdS in different lineages

Lineages	Delta	BA.1	BA.2	BA.2.12.1	BA.4	BA.5	BF.7	BQ.1.1
Count_N slope	0.02831	0.04935	0.01101	0.01899	0.009368	0.009981	0.006653	0.0141
Count_S slope	0.01775	0.01363	0.01311	0.01414	0.01871	0.01036	0.009699	0.02404
dNdS	1.59493	3.62069	0.839817	1.342999	0.500695	0.963417	0.685947	0.586522

**Method as below**

3. Slope of Count\_N/Count\_S == 0 doesn't mean of neutral selection

Obs:

① Guessing slope of  $\frac{\text{Ant-N}}{\text{Ant-S}}$  as Neutral.  
Wrong.

Obs:

$$t=0 \quad \frac{1}{1} \rightarrow t=1 \quad \frac{2}{2}$$

$$k=0 \quad \checkmark$$

---


$$t=0 \quad \frac{3}{4} \rightarrow t=1 \quad \frac{4}{5}$$

$$k = \frac{1}{20} > 0 \quad \times$$

Proof:

① Guessing slope of  $\frac{C_{nt-N}}{C_{nt-S}}$  as Neutral.  
Wrong.

Proof:

$$\frac{d\left(\frac{C_N}{C_S}\right)}{dt} = \frac{C'_N C_S - C_N C'_S}{C_S^2}$$

$$\text{When } D \rightarrow \frac{C'_N}{C_N} = \frac{C'_S}{C_S}$$

∴ Slope is controlled by:

I. Increment:  $C'_N$  &  $C'_S$

II. Aggregate:  $C_N$  &  $C_S$ .

How to count dNdS

$$\textcircled{2} \quad \frac{C'_N}{C'_S} = = \text{dwd}s.$$

$$\frac{\Delta C_N}{\Delta C_S} = \frac{(C_N + \delta) - C_N}{(C_S + \delta) - C_S}, \quad \delta: \text{infinitesimal.}$$

$$\frac{\Delta C_N}{\Delta C_S} = \frac{(C_N + \delta) - C_N}{\delta} \bigg/ \frac{(C_S + \delta) - C_S}{\delta}.$$

$$= \frac{C'_N}{C'_S}$$

$$\therefore \frac{C'_N}{C'_S} \begin{cases} > 1 & \text{positive} \\ = 1 & \text{neutral} \\ < 1 & \text{negative} \end{cases}$$

Implement

$\Rightarrow$  Im on Cont-N & Cont-S

$k_1, k_2$  as slope

$$dn ds = \frac{k_1}{k_2}.$$

#### 4. Some Concern Lineages

BQ.1.1 / BQ.1

BA.5 / BA.2 / BA.4

BN.1 — 美国

BA.2.75

B.1.1.529

XBB. 来自 BA.2.75 和 BA.5

B.2.3.20

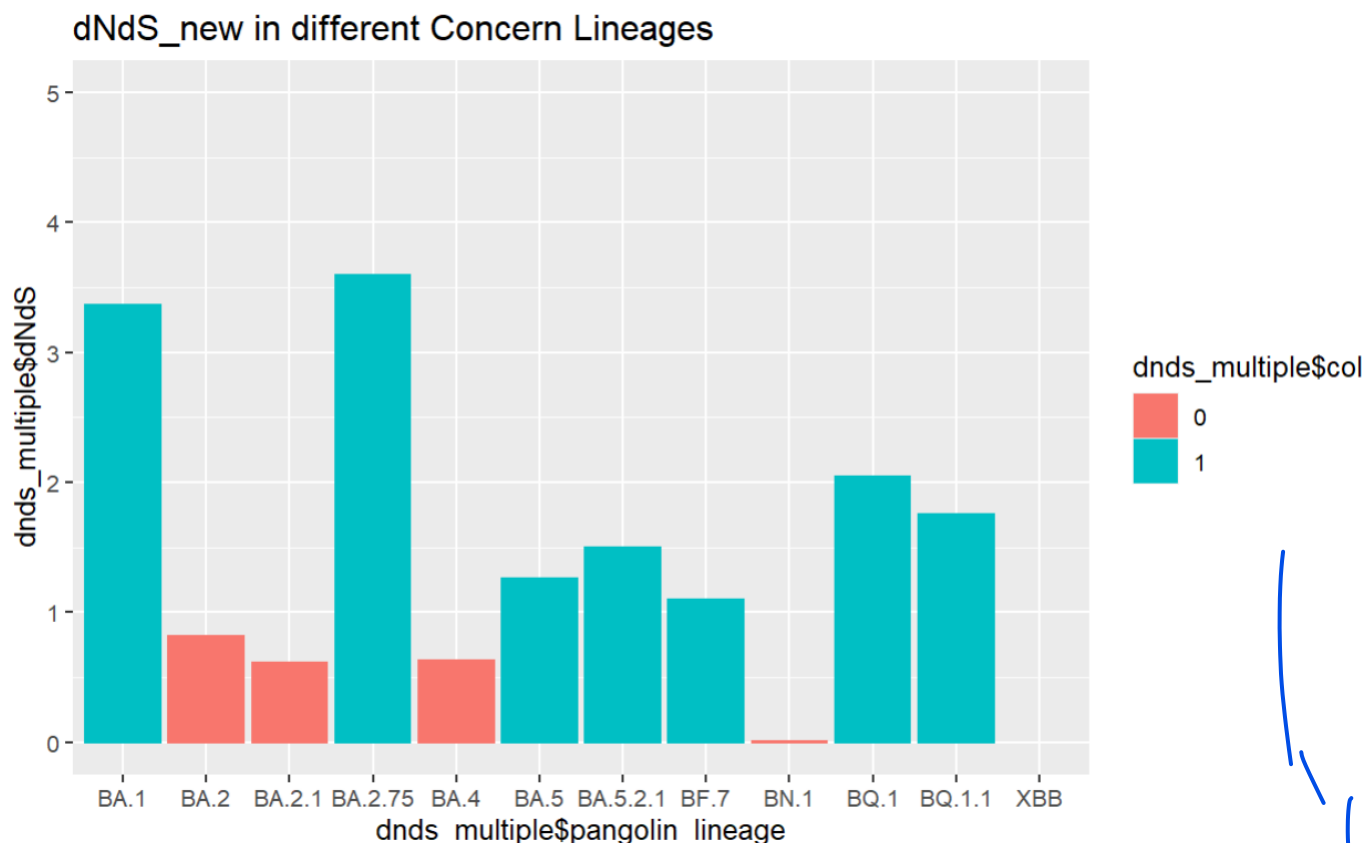
BJ.1

—— 22年11月16号

reference: <https://www.ecdc.europa.eu/en/covid-19/variants-concern>

## 5. dNdS\_new of Concern Lineages

I did linear regression on all 455 Omicron lineages, and picked a subset of them to plot. Those plotted lineages are lineages of concern.



**Blue: >1, positive**

**Red: <1, negative**

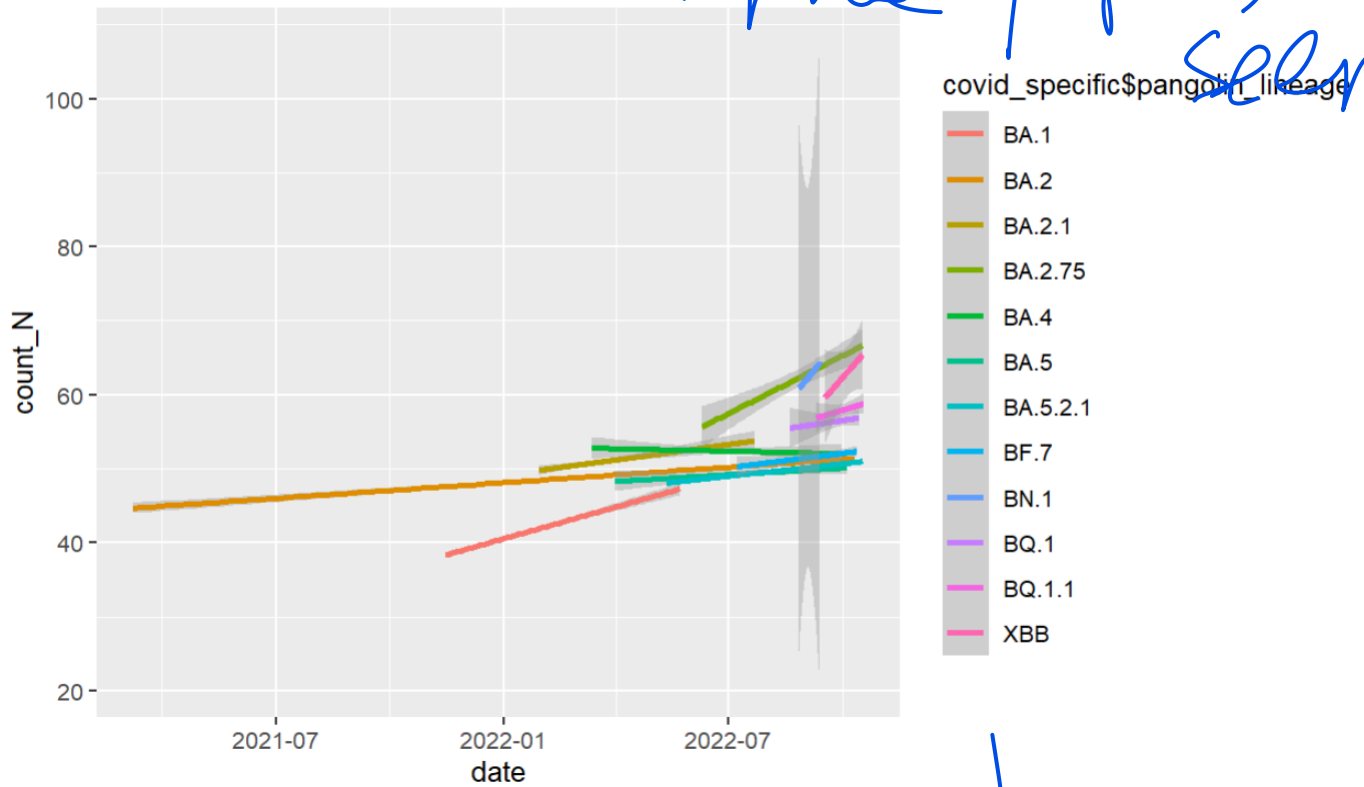
XBB has not many data, very noisy.

*dNdS of XBB: 11.26744*

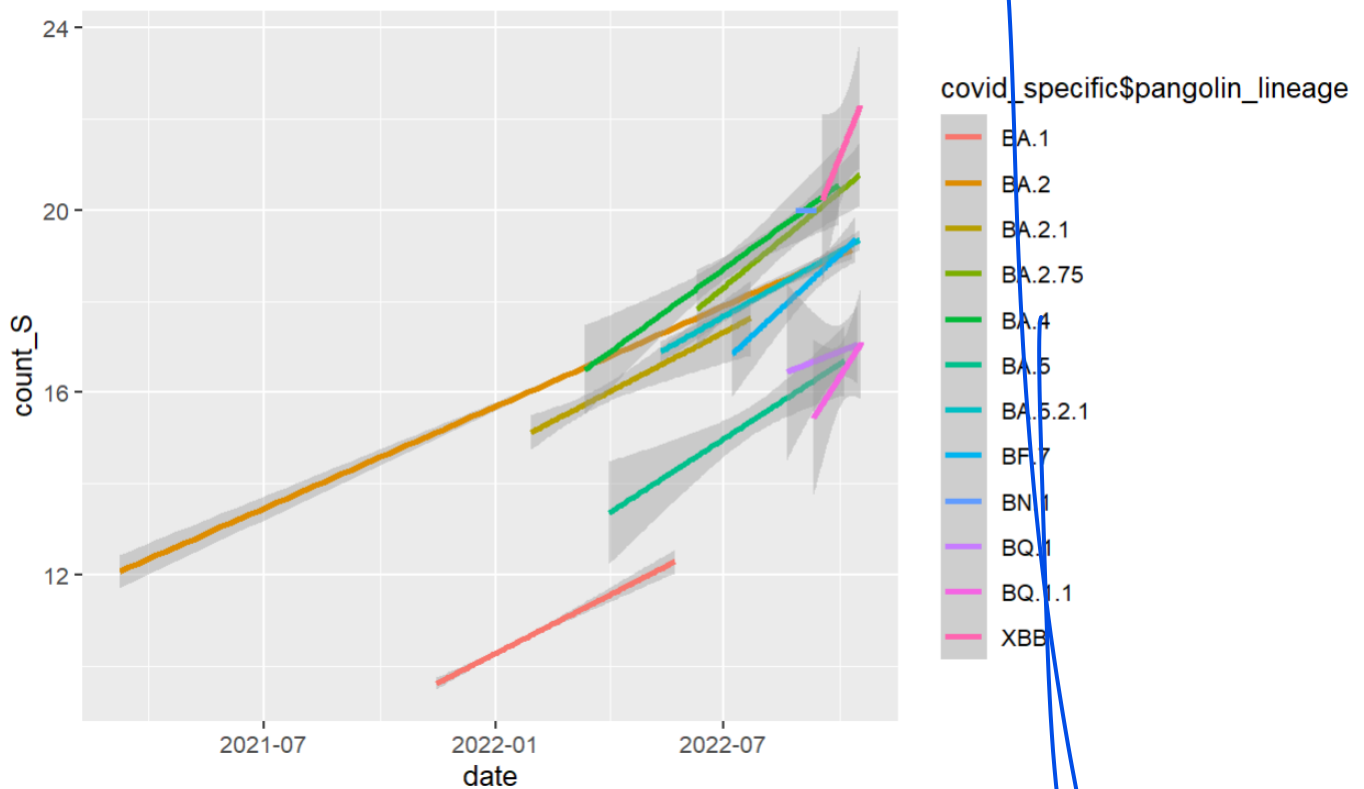
## 6. Distribution of these lineages

75%  
back - C].  
give only.  
per mutation.

一. Slope of Count\_N  
delta\_N(slope)

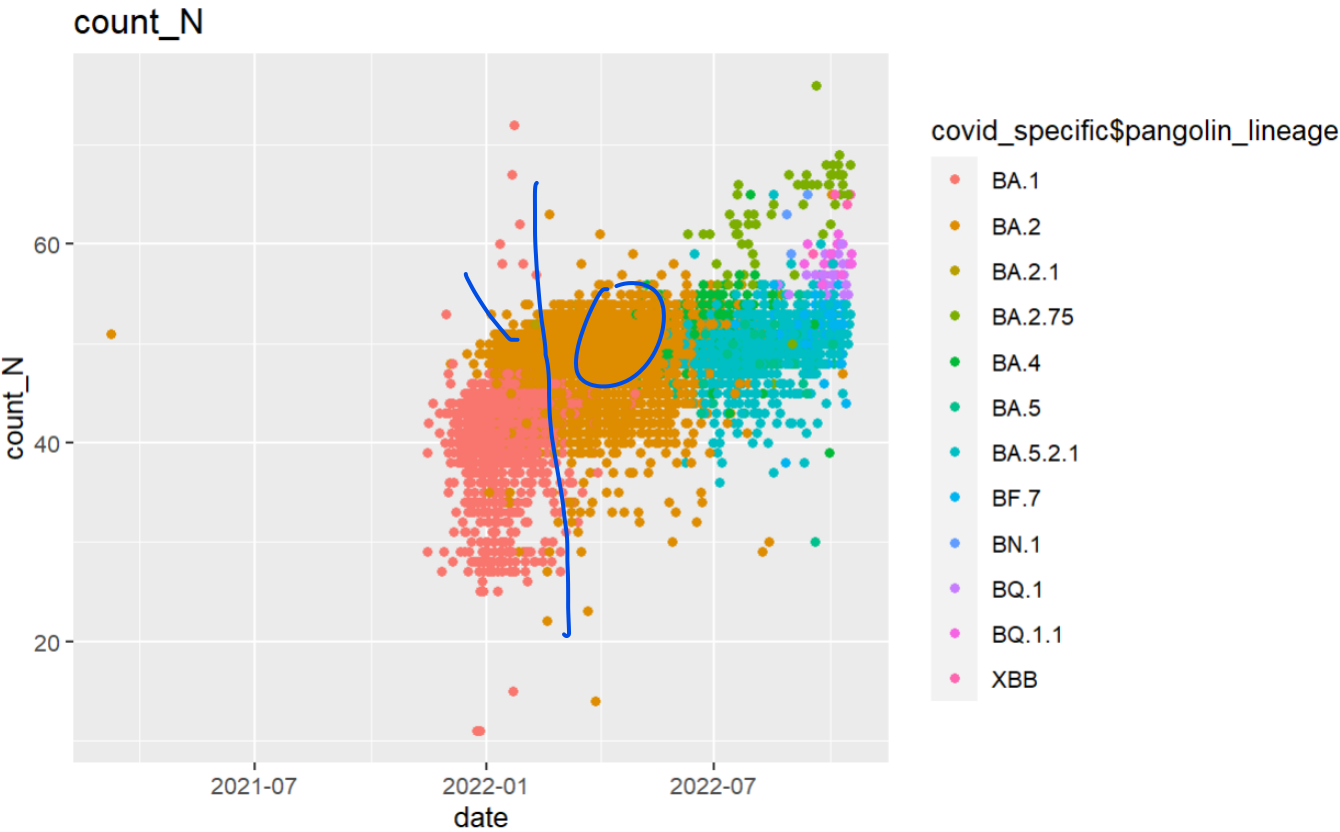


二. Slope of Count\_S  
delta\_S(slope)

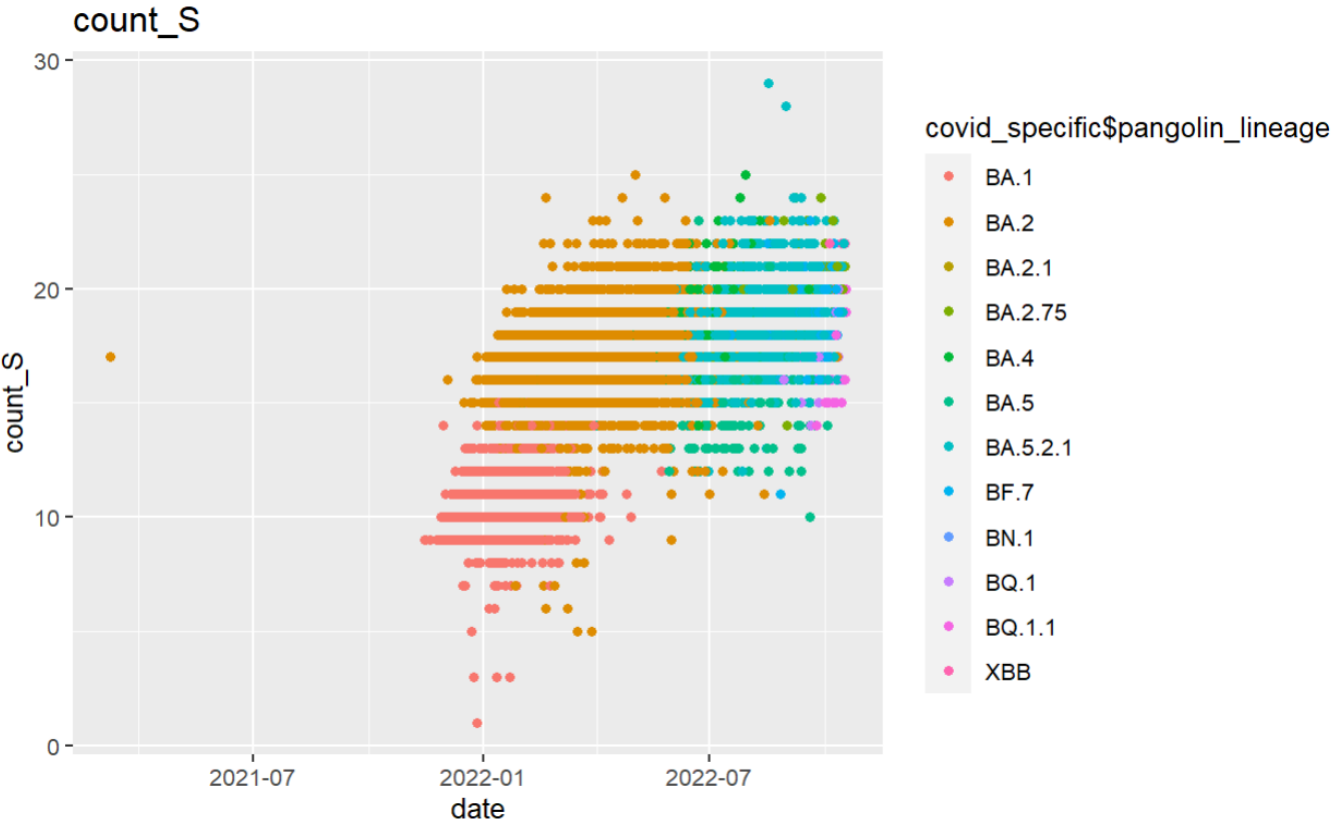


三. Point plot of Count\_N





四. Point plot of Count\_S



## 7.Conclusion

Those "popular" lineages were mostly positive selected by environment. Thus, for BQ.1.1, we should focus on if it will continue been positively selected. And for XBB, with more data, dNdS would get lower, but will it still be positive?

While result of BA.2 is "interesting". What happened to that? Non\_Syn increased fast at first, while got flat after. In other words, BA.2 was firstly positive selected, while later negatively.

What drives that?

## 8. Continue

I could take a look on the trend of dNdS on each lineages, that since when BA.2 changed from positive to negative?

And I think BA.1 has been positively selected till end, it could mean that all BA.1 have been mutated to another lineage(some aa\_change has been fixed.)