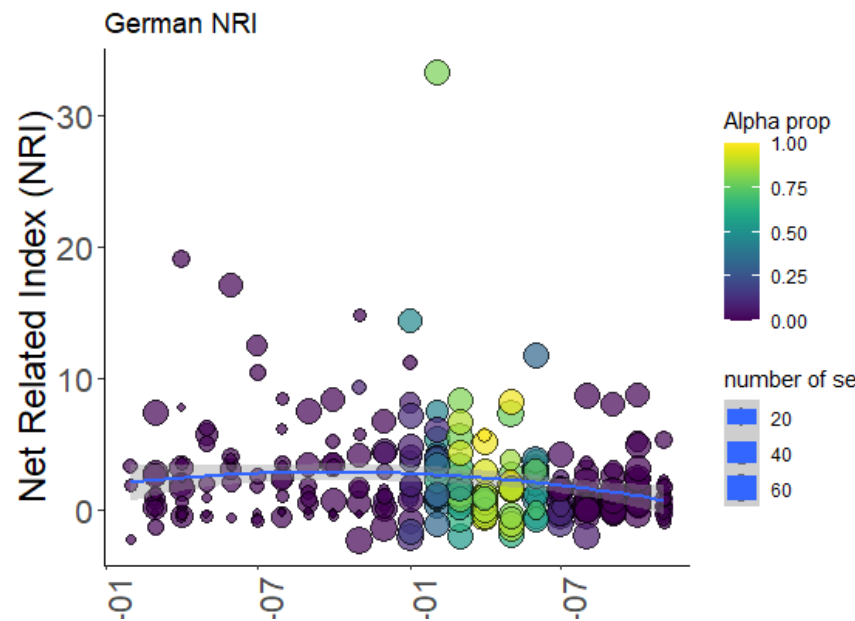


Intro - doc not in any particular order

- NTI (SESmntd) is more reliant on the close relatives so might expect stronger signals than NRI (SESmpd)
 - Clustered:
 - +ve NTI and +ve NRI
 - Overdispersed
 - -ve NTI and NRI
- D 'compares observed independent contrasts with those under a brownian or random shuffling null model'
 - Has low power in small clades - perhaps consider excluding smaller?
 - $D = 1$ = Random with respect to phylogeny
 - $D = 0$ = Distribution expected by brownian motion
 - $D > 1$ = More overdispersed than expected by random
 - $D < 0$ = More clustered than expected under brownian
- For updated
 - [EcoCovid_Notes](#)
- For more stats stuff see
 - [Eco_Covid_Stats_Notes](#)

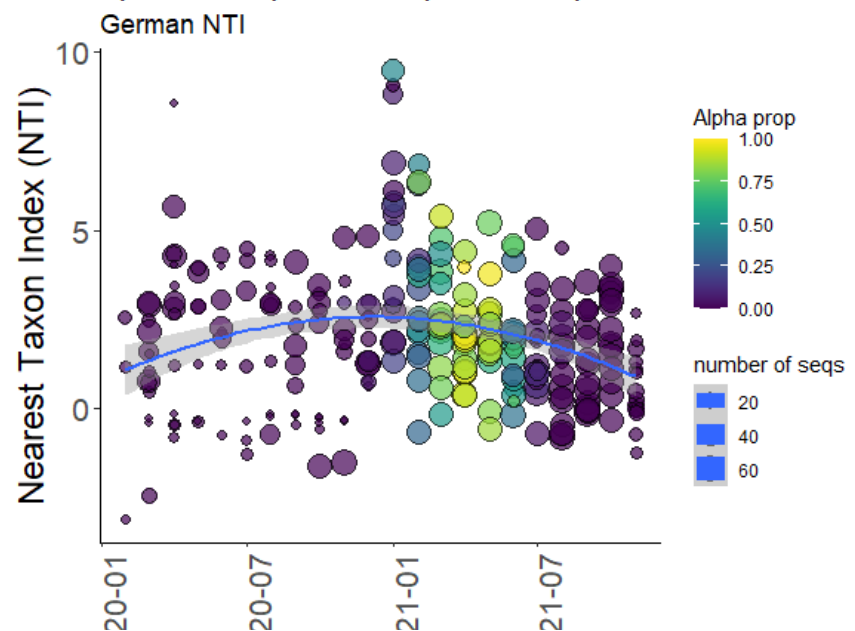
Germany NRI (updated)

This makes it look like the NRI decreases when the new variant spreads. More overdispersion as the new variant is spreading, other VOCs emergence as alpha decreases again which might be why we don't see a return to clustering after the alpha decr?



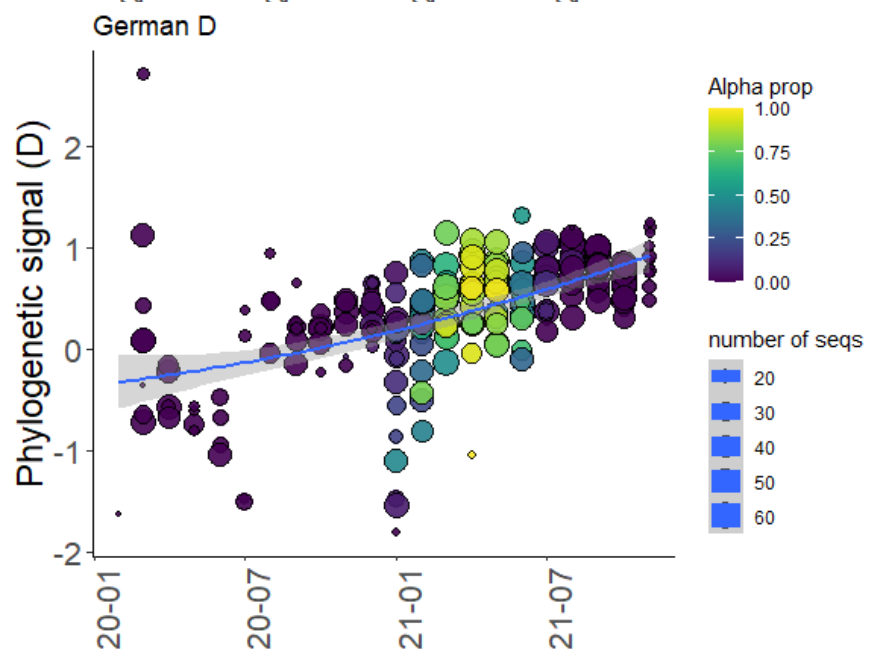
Germany NTI (updated)

Decreasing clustering as alpha spreads, then maybe fluctuations as different strains become more prevalent



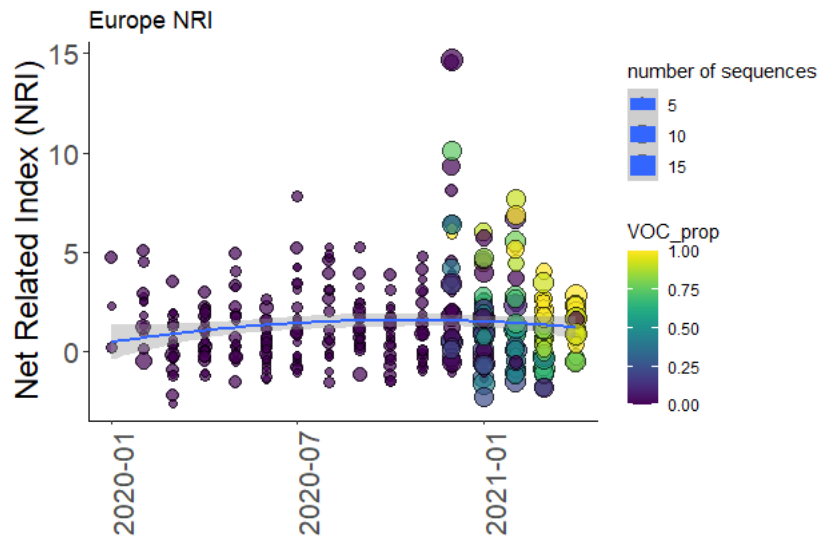
Germany D (updated) (>1 = over, 1 = rand, 0 = brown, < 0 = clust)

I removed any regions that had fewer than 5 sequences. Early in pandemic has variable values that then narrow to between 0-1 when the new variant emerges - to do with sampling or more eco-evo reasons? Does it then narrow again when delta emerges



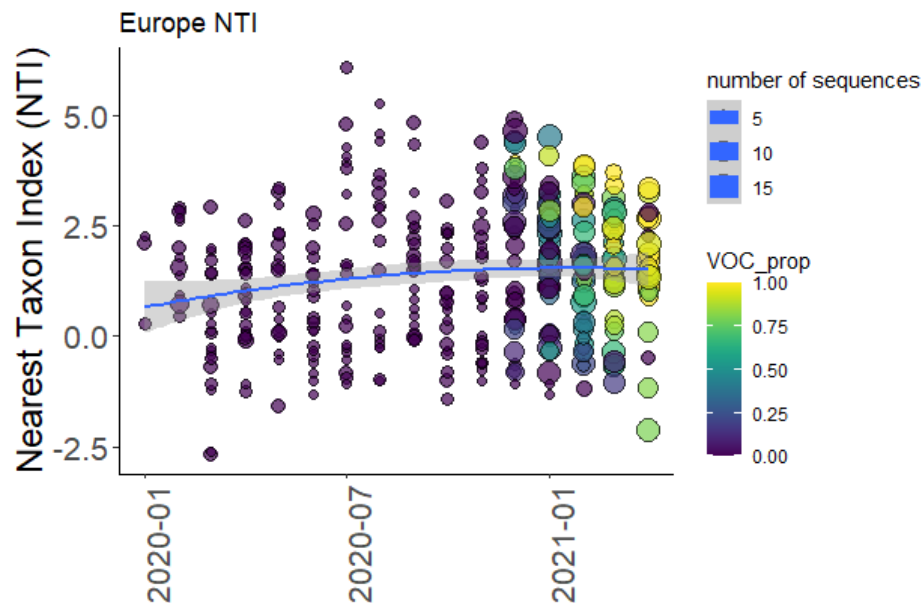
EU NRI

When the new strain is emerging it looks like it decreases the NRI values (less clustered). I think that actually this variant is decreasing the NRI values of areas as it makes regions more overdispersed when the variant is in mid proportions. Then when it spreads the NRI increases again, which you can see in the higher VOC prop regions



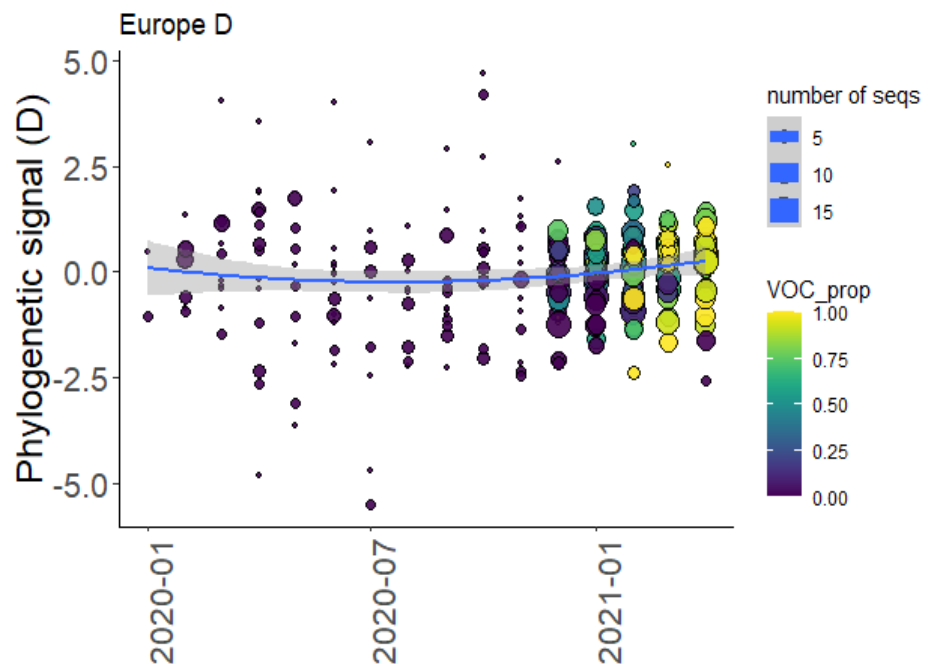
EU NTI

Less strong decrease in clustering during emergence as with NRI but still see the regions with higher VOC are more clustered than those with intermediate VOC proportions



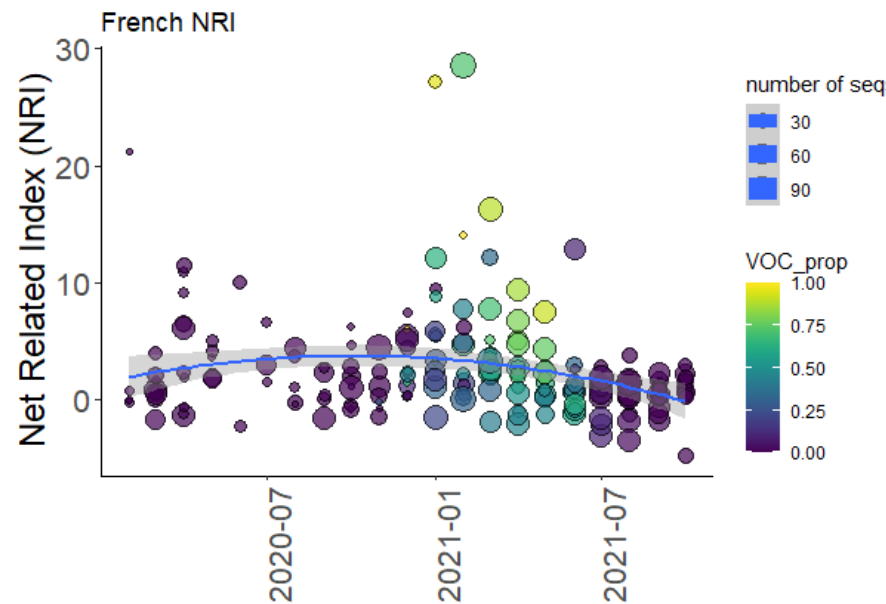
EU D

I have excluded regions with less than 2 samples for now. Similar to Germany it looks as though it was fairly random before emergence and then narrows down. However I suspect this may just be the effect of increasing sample size (CHECKED WITH PLOT AND NO CLEAR TREND WITH INCR SAMPLE SIZE). Not such a pattern with the VOC as with NTI/NRI



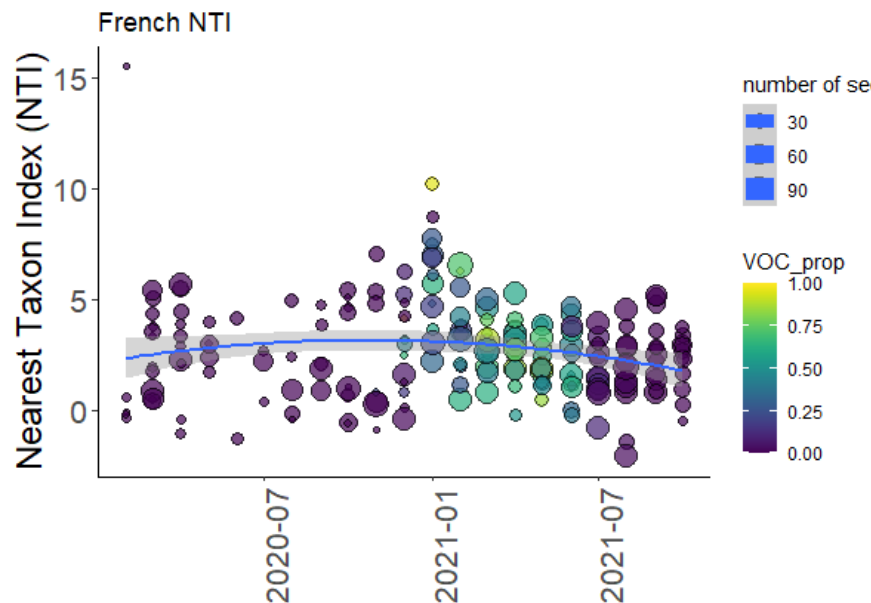
FRENCH NRI

With the French data, the VOC B.1.1.7 starts decreasing in proportion again later in 2021, however the variants replacing it I suspect are still from the same lineage and have evolved from B.1.1.7. See a slightly similar pattern to EU in which regions with higher VOC are more clustered than those regions with intermediate proportions.



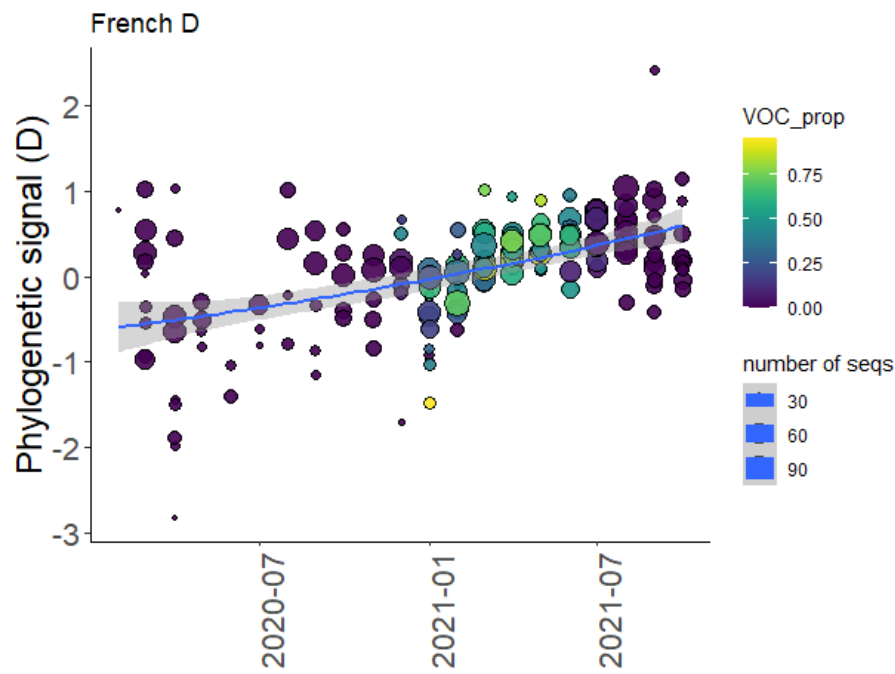
French NTI

A similar downward trend after emergence as clustering decreases but not such an obvious pattern with regions of high VOC being more clustered.



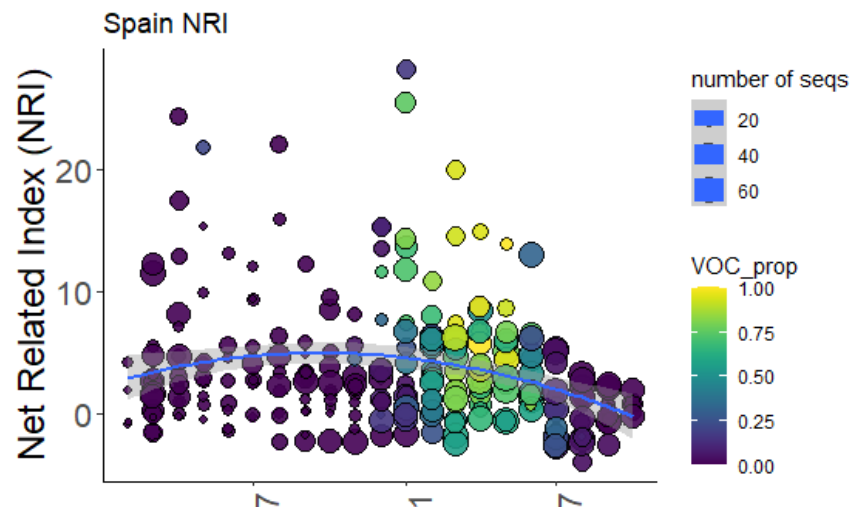
French D

Quite an interesting narrowing down of values, this could reflect better sampling later in pandemic or maybe the emergence brings the values closer to between 0-1 which is between expected for random/brownian cos for D -ve is clustered - emergence makes it less clustered. But you might expect high VOC regions to be more -ve again?



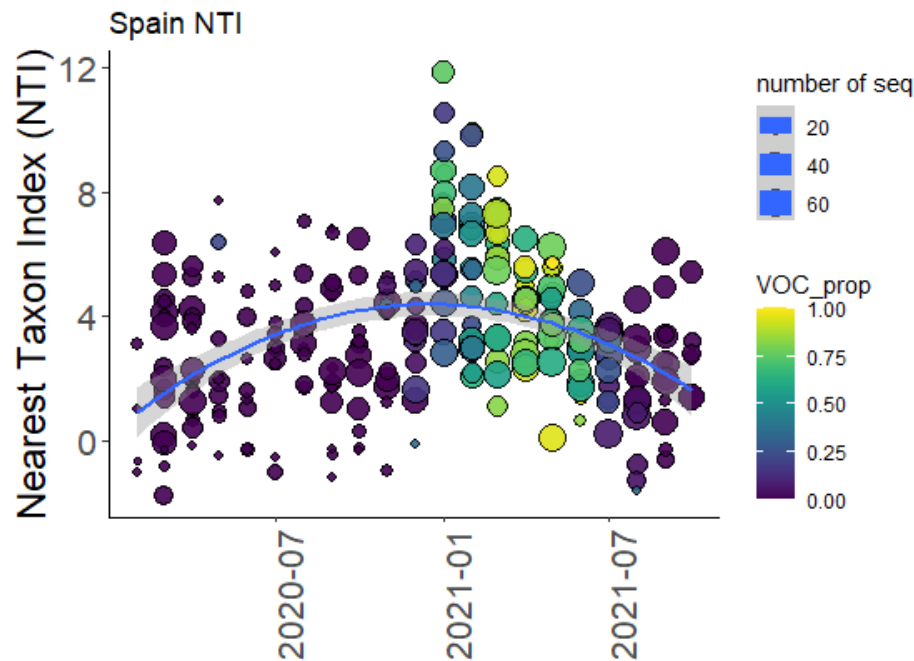
Spanish NRI

Only coloured by the alpha variant



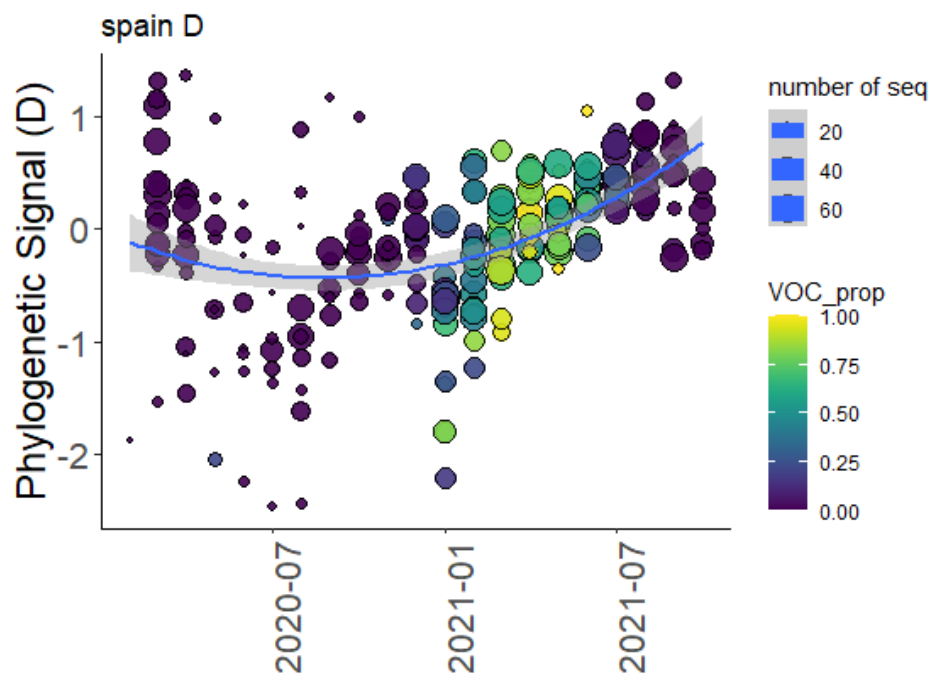
Spanish NTI

Can see the narrowing maybe caused by the emergence of variants which then begins to increase again



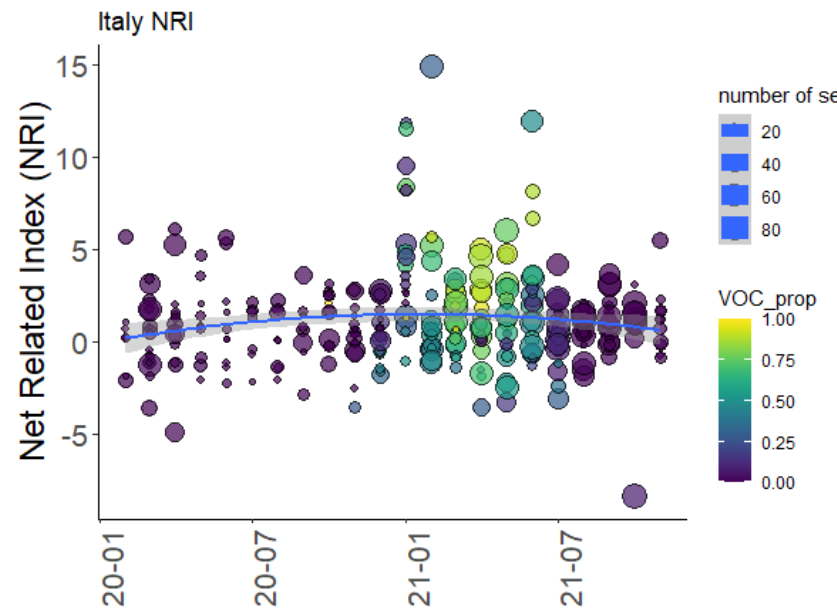
Spanish D

Similar to german, there is a shift from strongly clustered then can maybe see it start to shift back?



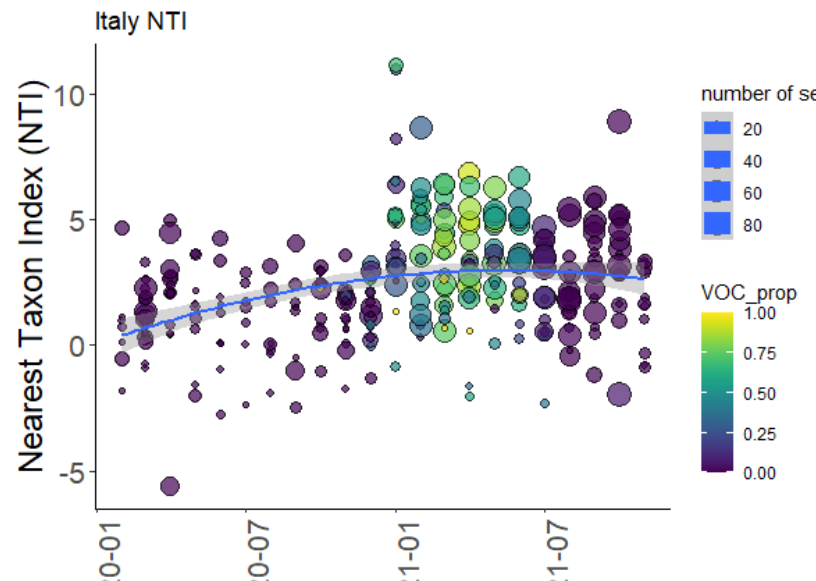
Italy NRI

Looks like a less clear pattern in the Italy data, some regions with considerably higher NRI while VOC emerges and then one at like -5?



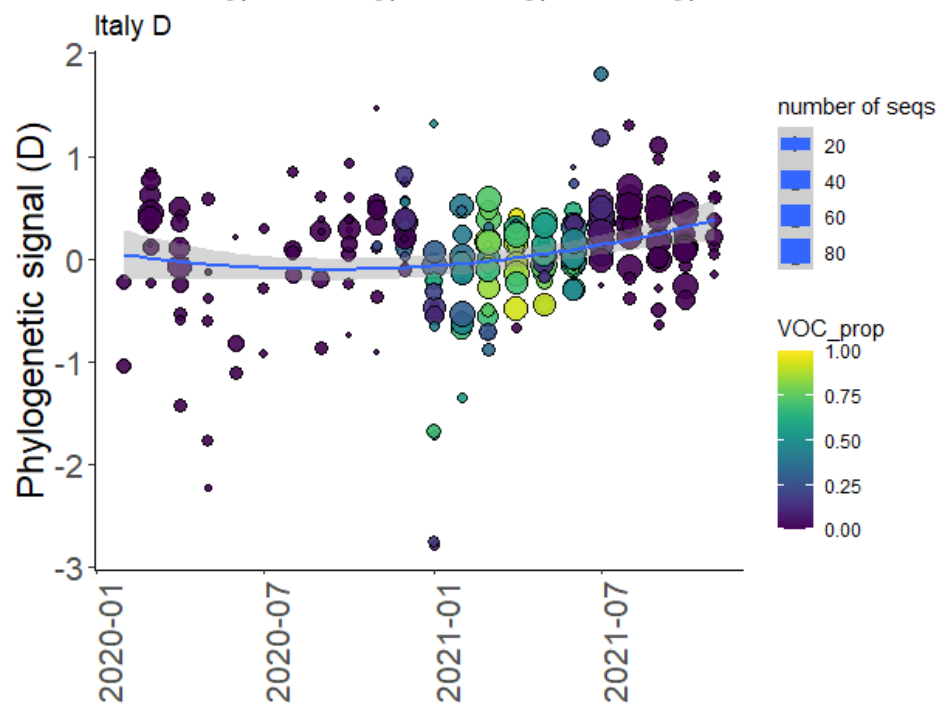
Italy NTI

Less clear pattern than spanish its more like german/french



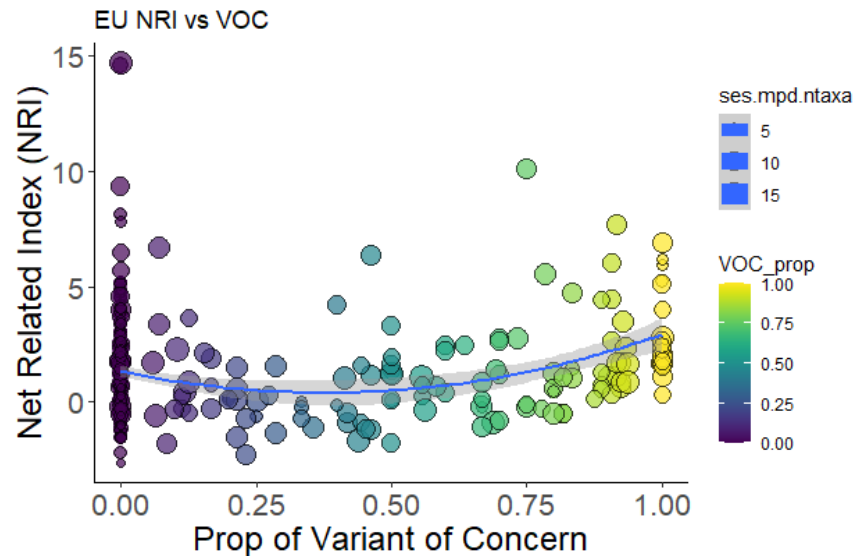
Italy D

Similar patterns in Italian with shift from very strongly clustered and then values clump together when VOC emerges and similarly to spanish you start to see the values dropping (becoming more clustered again) in more recent months



EU VOC vs NRI

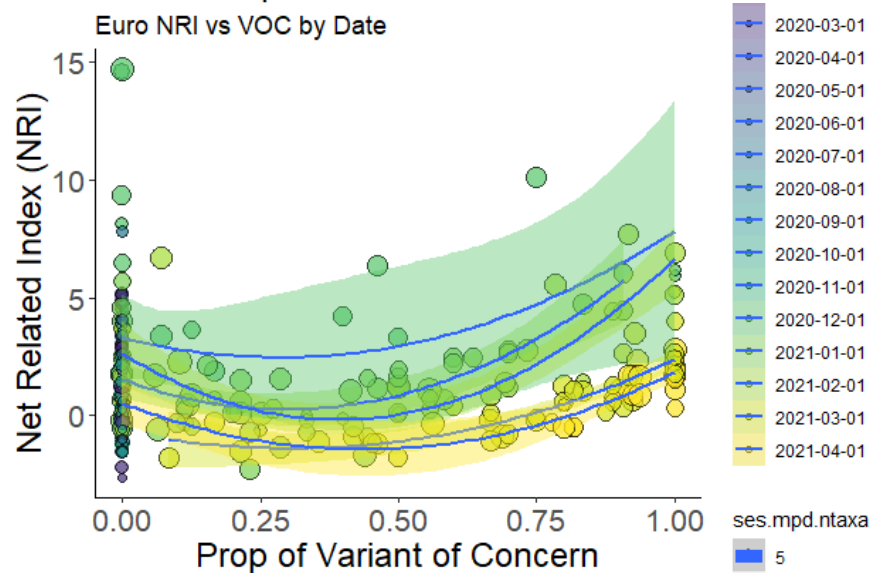
This supports the idea of NRI being highest before the VOC, with an increasing trend through time (seen on prev graph) then lowest at the emergence of the VOC followed by an increase as the VOC spreads to 100% (fixation?) which would be expected to increase clustering (more spp per genus than expected). The relatedness initially is lower as it spreads but then as it dominates there is increased clustering



EU VOC vs NRI coloured by date

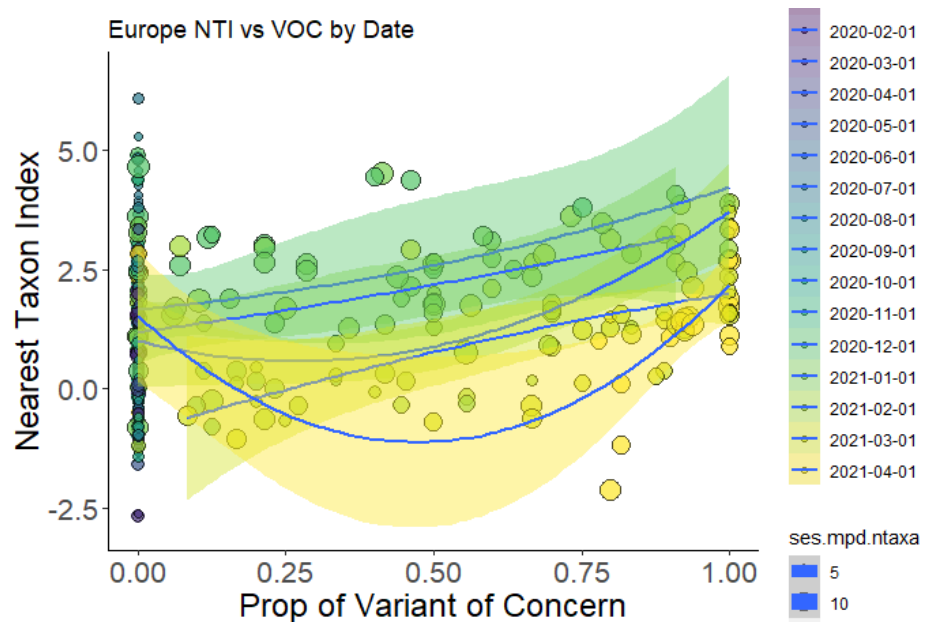
I like how this looks, so the lines are now linking dates and you can see that the more recent months have notably lower NRI values indicating phylogenetic overdispersion, perhaps as phylogenetic divergence increases due to evolution in all lineages?

But there is still a positive relationship between the proportion of VOC and the NRI, indicating that as the VOC dominated communities it leads to clustering



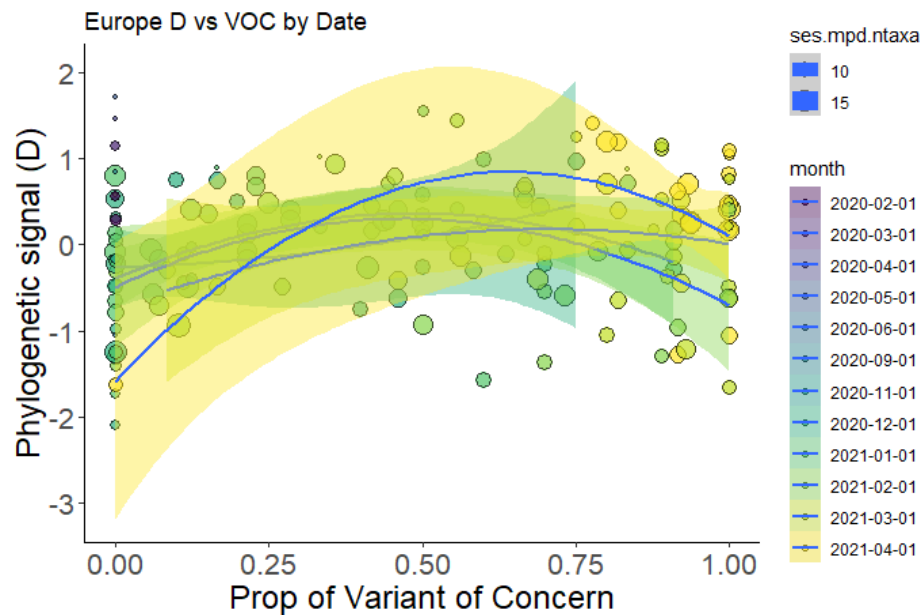
EU VOC vs NTI coloured by date

This shows a similar pattern to above but the most recent month is showing a big ol U bend where middle proportion regions have lower NTI which then increases as prop does. NTI is a metric which detects the patterns near the tips of tree (Keck & Kahlert)



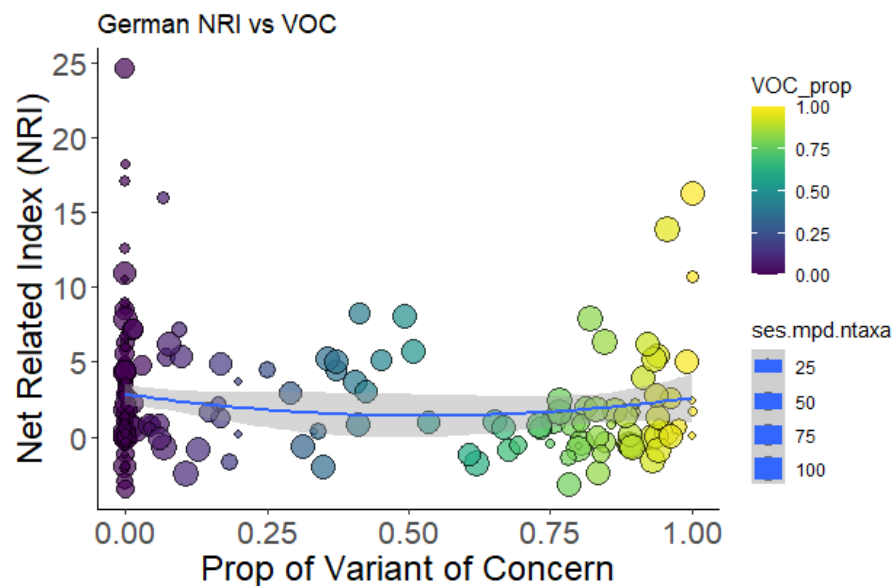
EU VOC vs D by Date (>1 = over, 1 = rand, 0 = brown, < 0 = clust)

Kind of the faint pattern that lower and higher VOC proportions are more clustered but quite vague



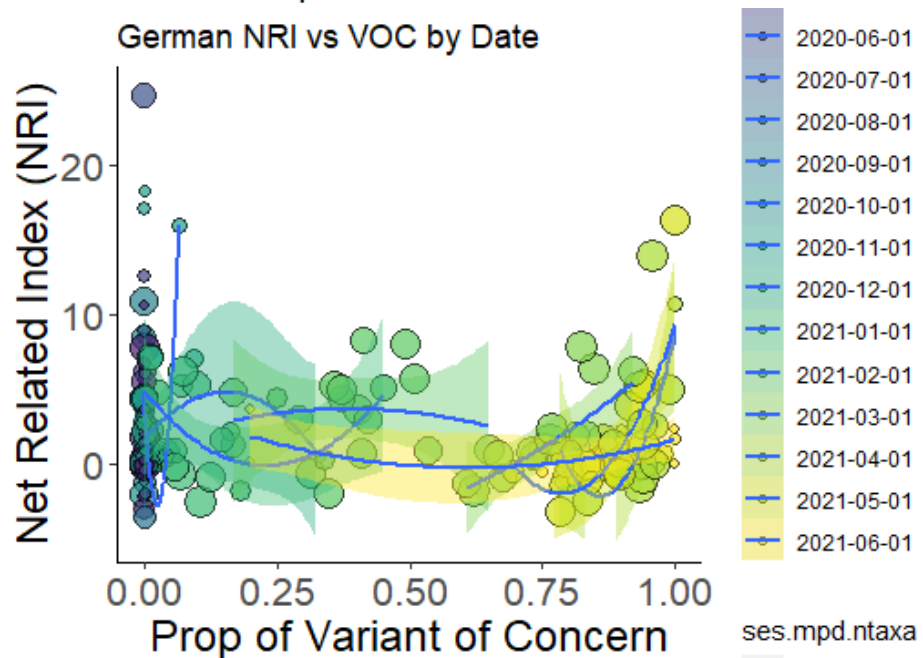
German VOC vs NRI

Less of a pattern than EU but still a slight trend for the values to narrow and only be low NRI values at mid proportions of VOC



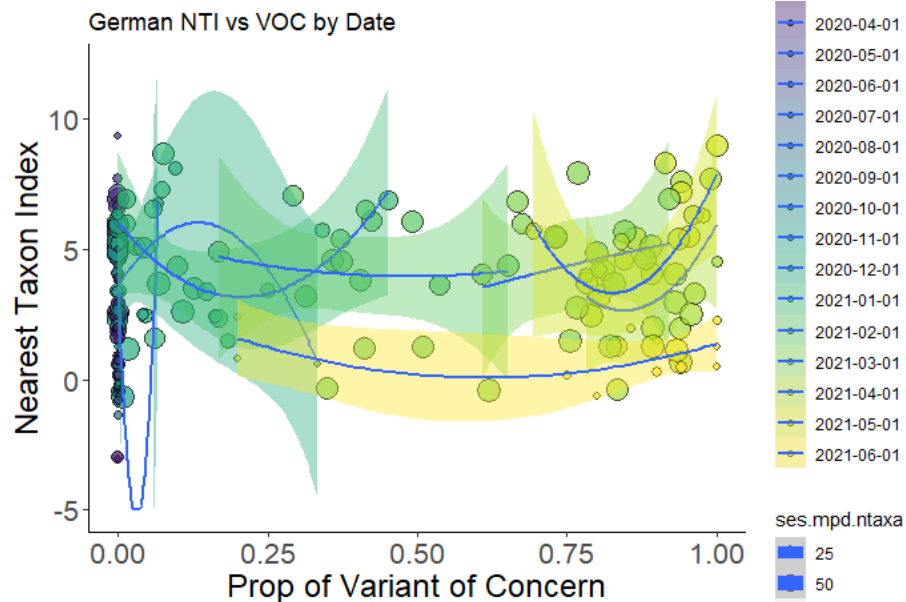
German VOC vs NRI by Date

Harder to see on here than on EU data as it spread through germany quicker so all the regions with the highest proportions are also the most recent dates. But there is a different trajectory at the highest props vs the middle



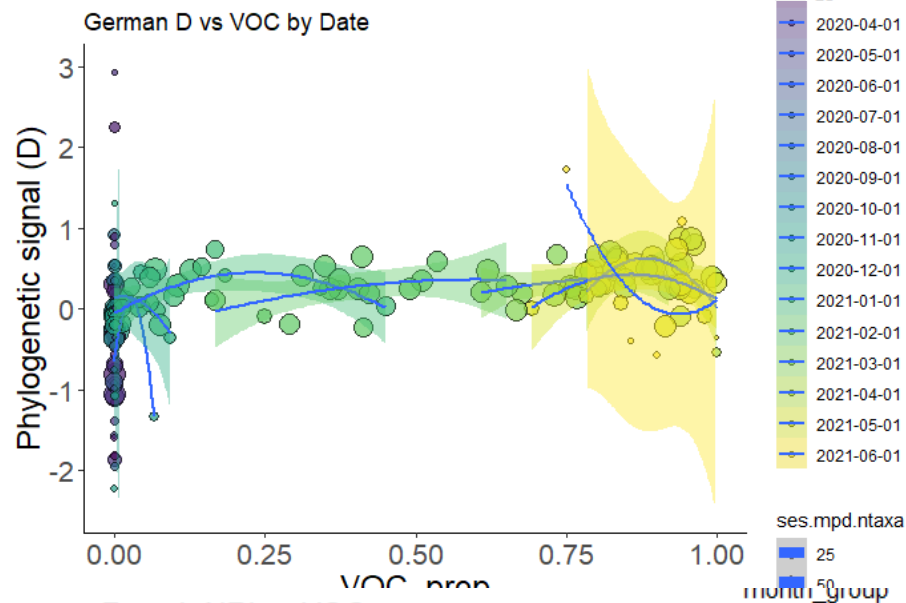
German VOC vs NTI by date

Very similar pattern to NRI



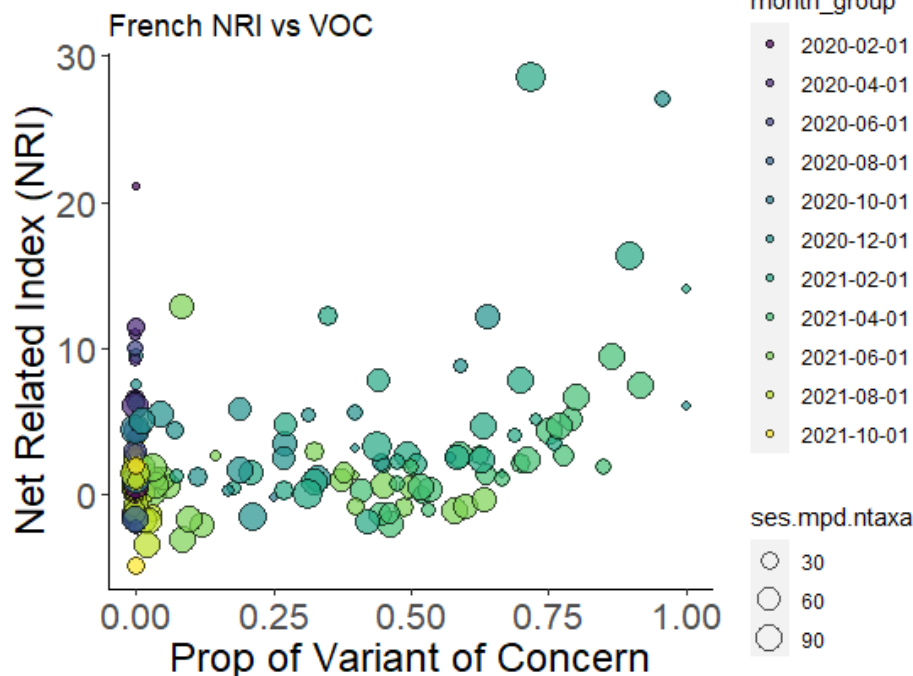
German VOC vs D by date (>1 = over, 1 = rand, 0 = brown, < 0 = clust)

Pretty hard to see any patterns in here but I think the upwards trend as regions become less clustered due to the spreading VOC. but then in the highest VOC regions not really seeing the return to clustering



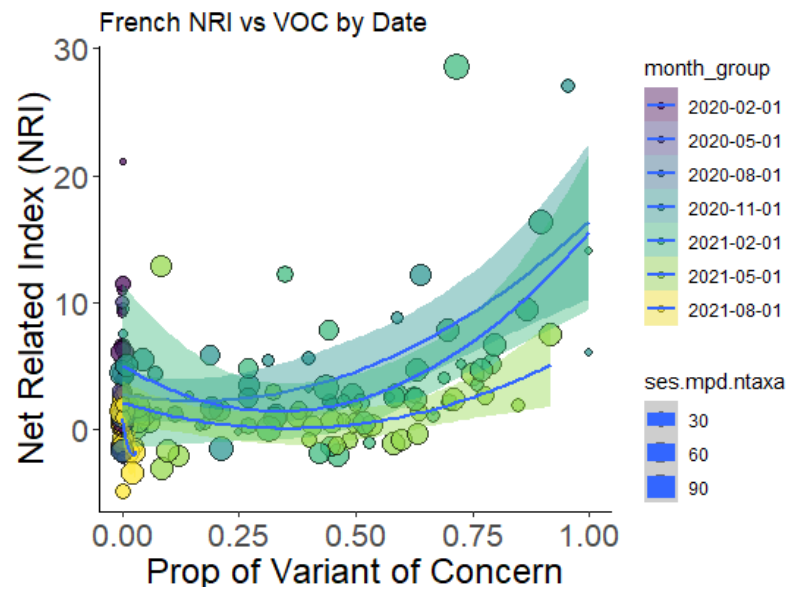
French VOC vs NRI

Trend for VOC to increase with VOC proportion, not seeing same spread of NRI values in the low VOC regions as was in EU or german



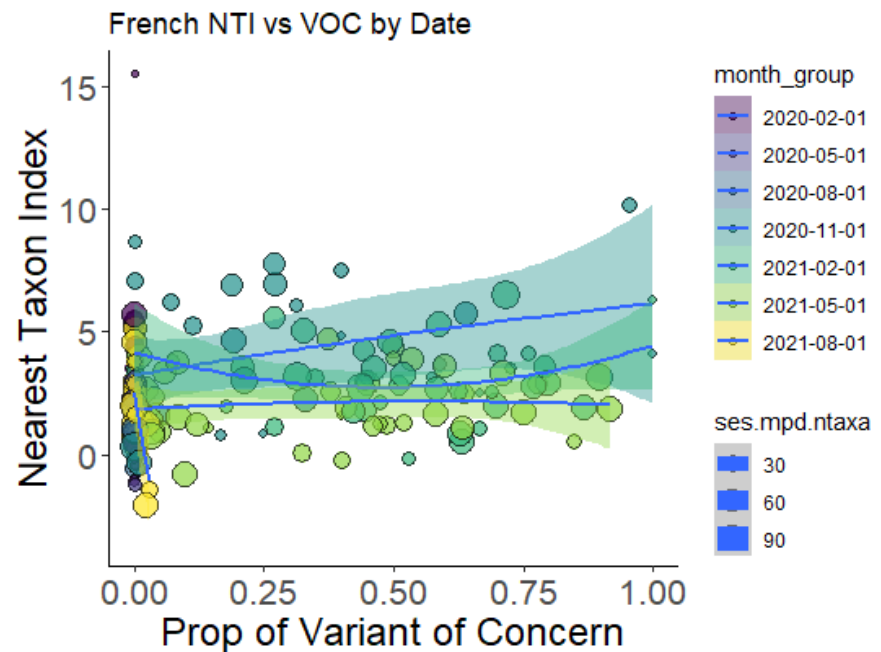
French NRI vs VOC by Date

So the most recent dates have low levels of the B.1.1.7 but are probably descended from the same lineage. Slight trend for most recent dates have lower clustering.



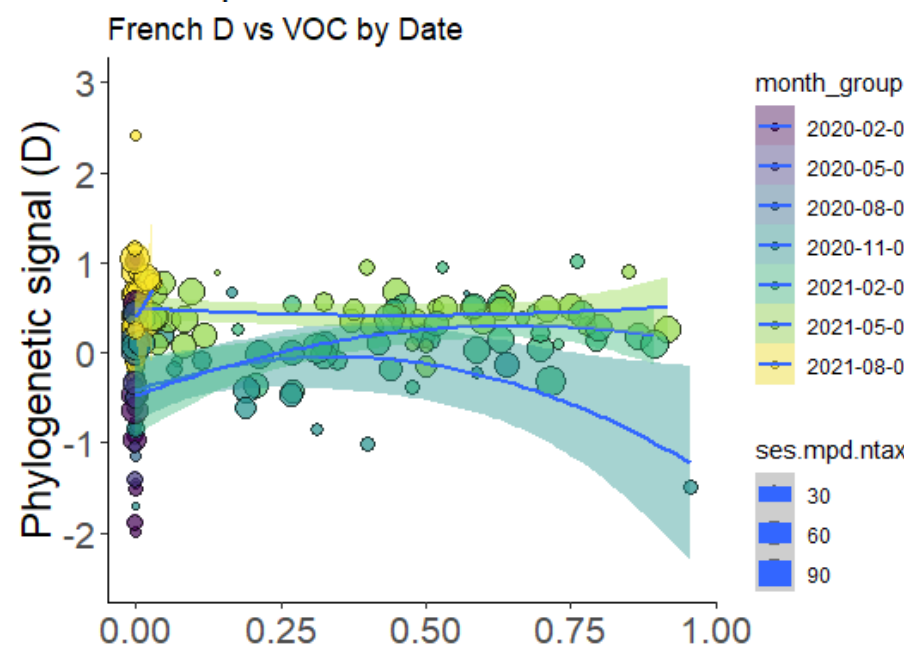
French NTI vs VOC by Date

Less pattern with higher VOC not really having any higher NRI values but there does seem to be an effect of date with most recent dates having lower NTI values



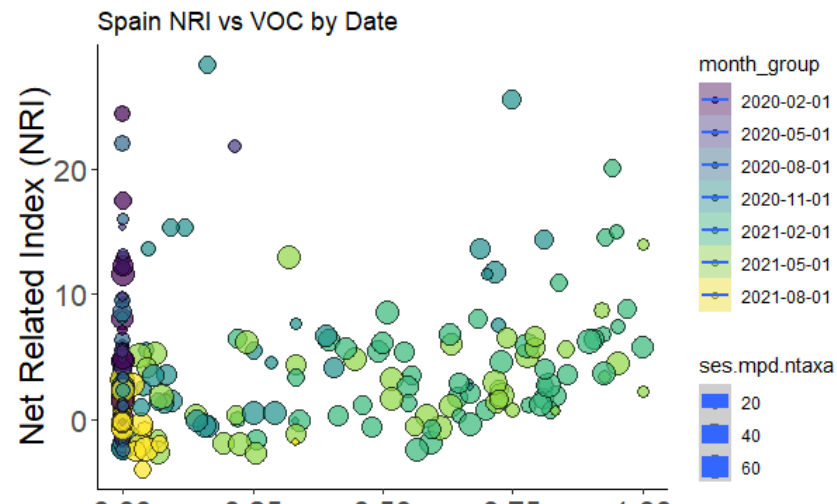
French D vs VOC by Date

More recent dates have higher D values = less clustered.

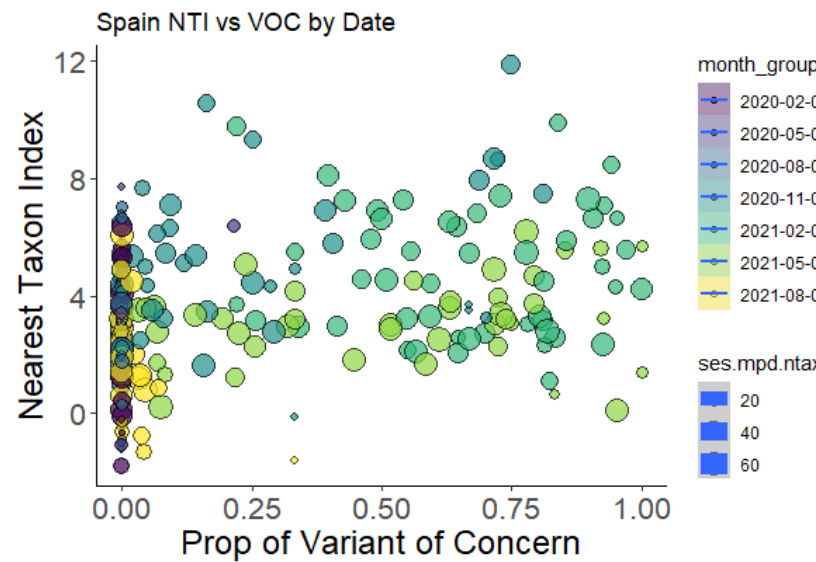


Spain NRI vs VOC by Date

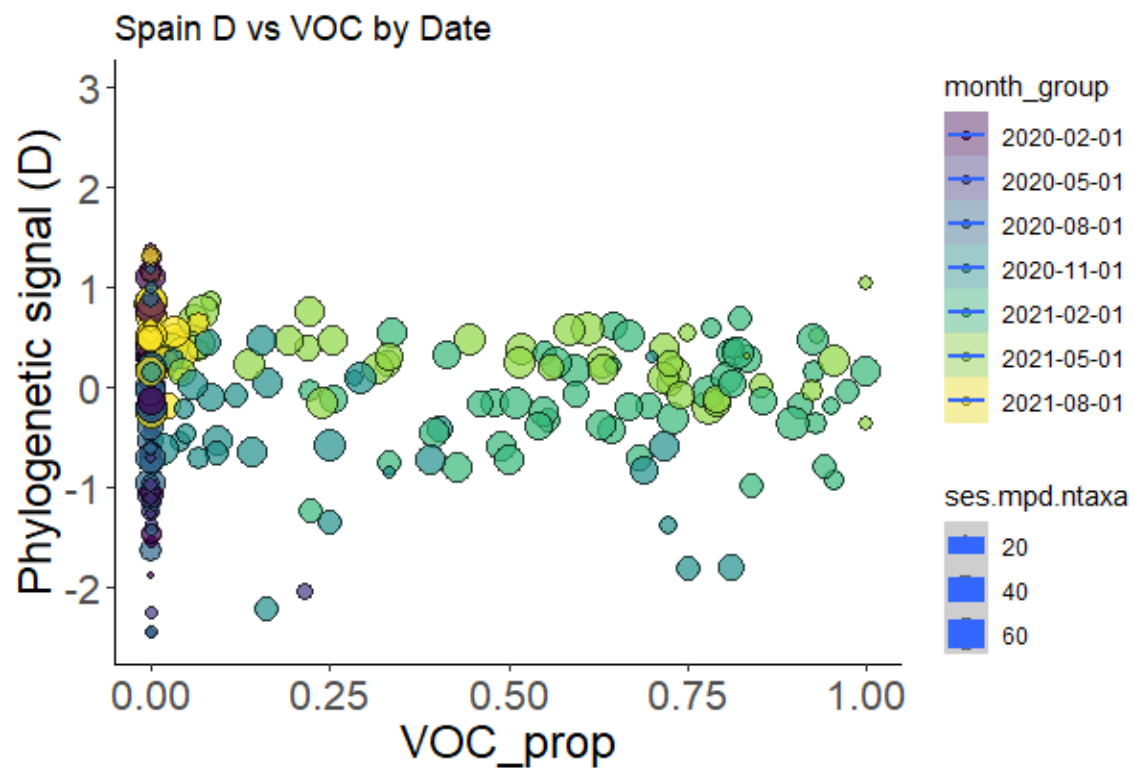
Cant compute the lines cos not enough data points



Spain NTI vs VOC by Date

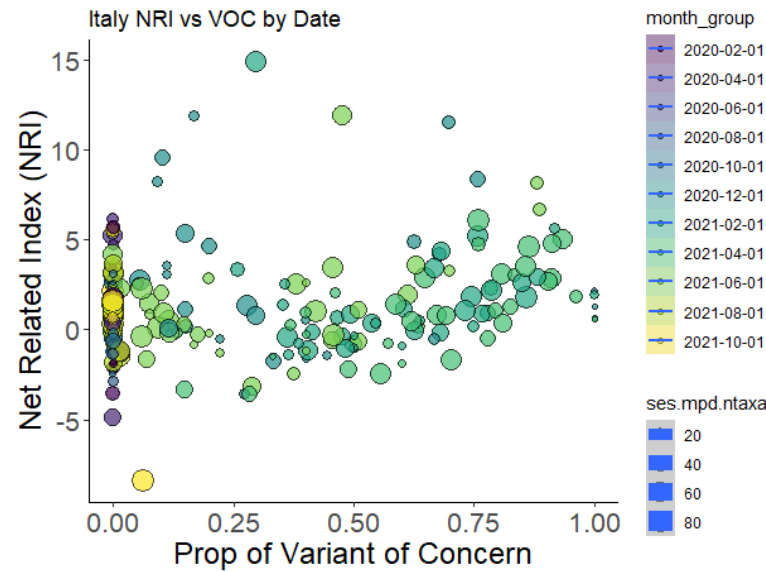


Spain D vs VOC by Date



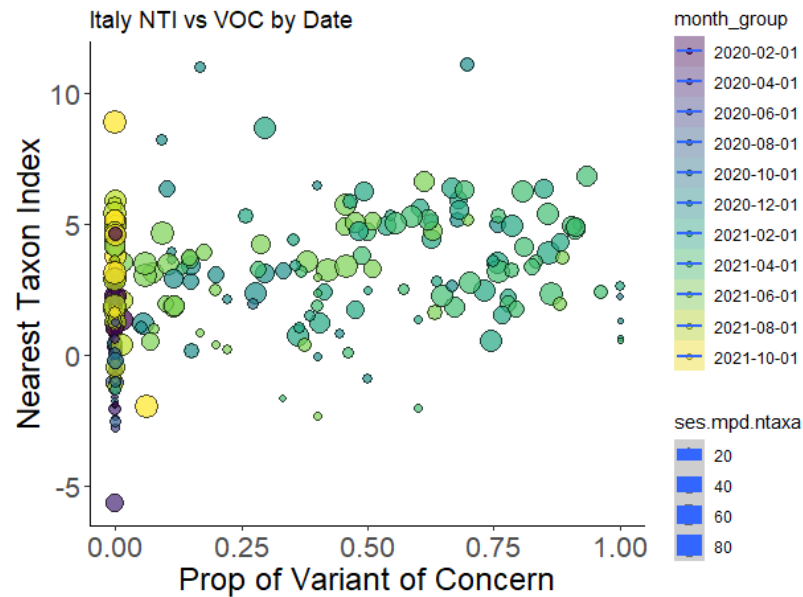
Italy NRI vs VOC by Date

Cant put lines on cos not enough data points for stat_smooth. Not a particular clear interaction with date for italian NRI

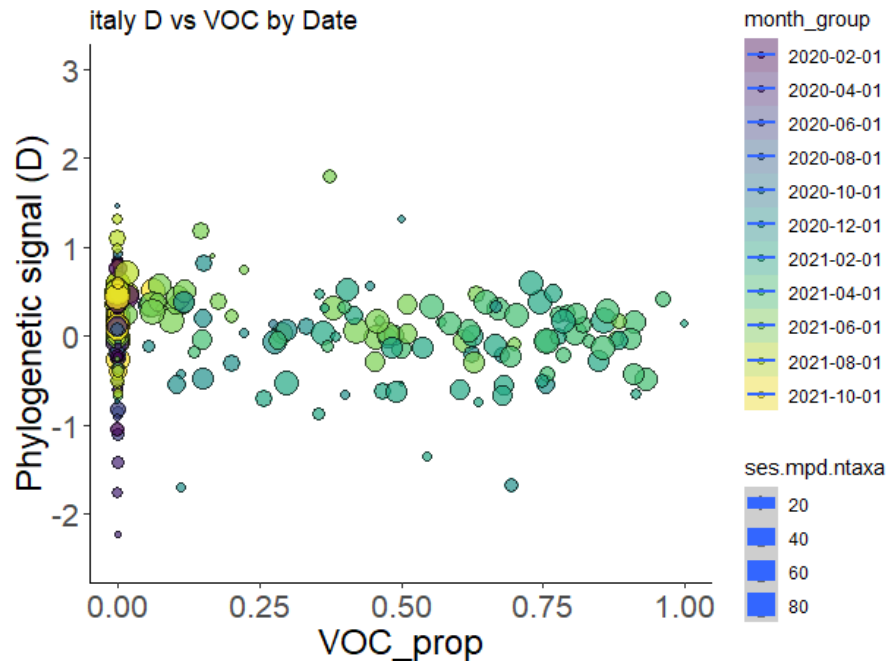


Italy NTI vs VOC by Date

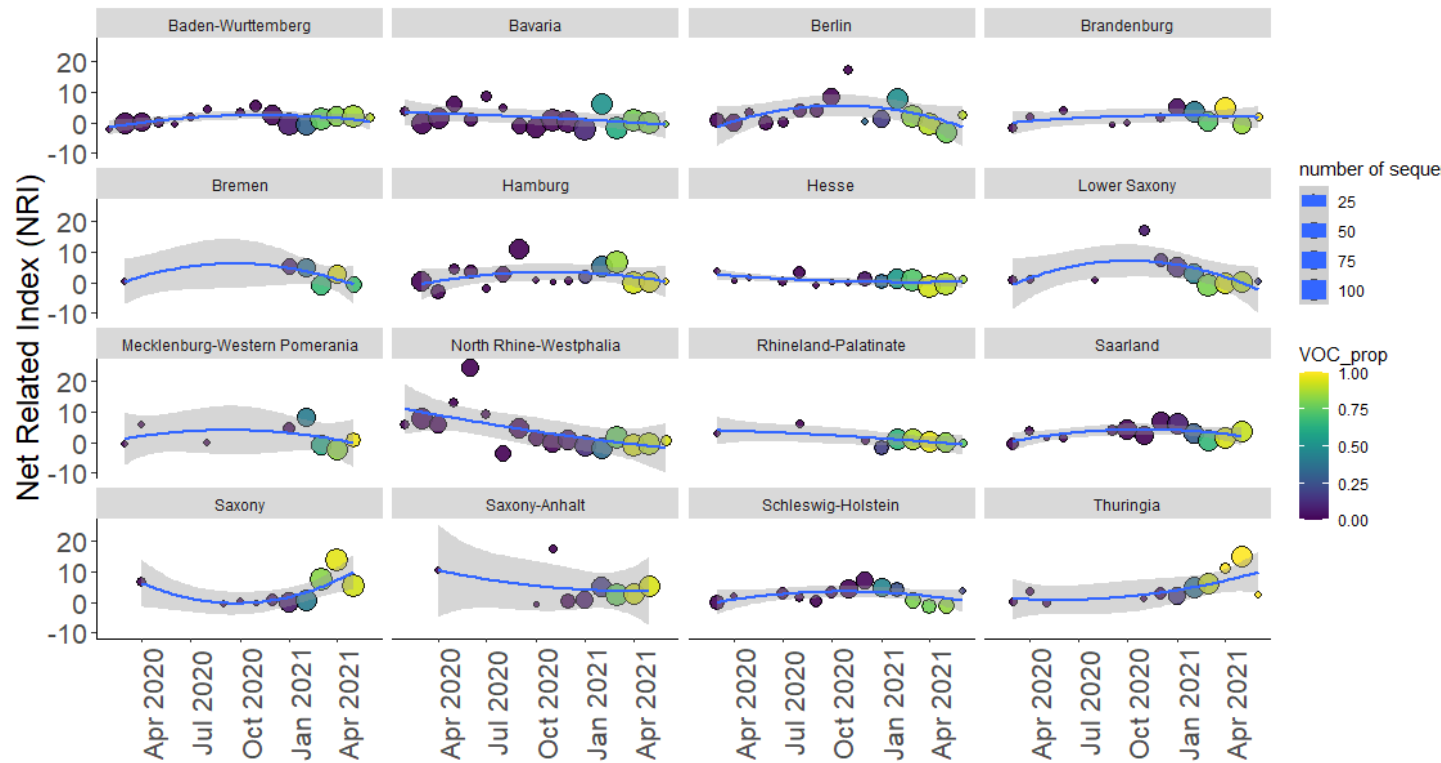
Again couldnt make lines and not a strong pattern



Italy D vs VOC by Date

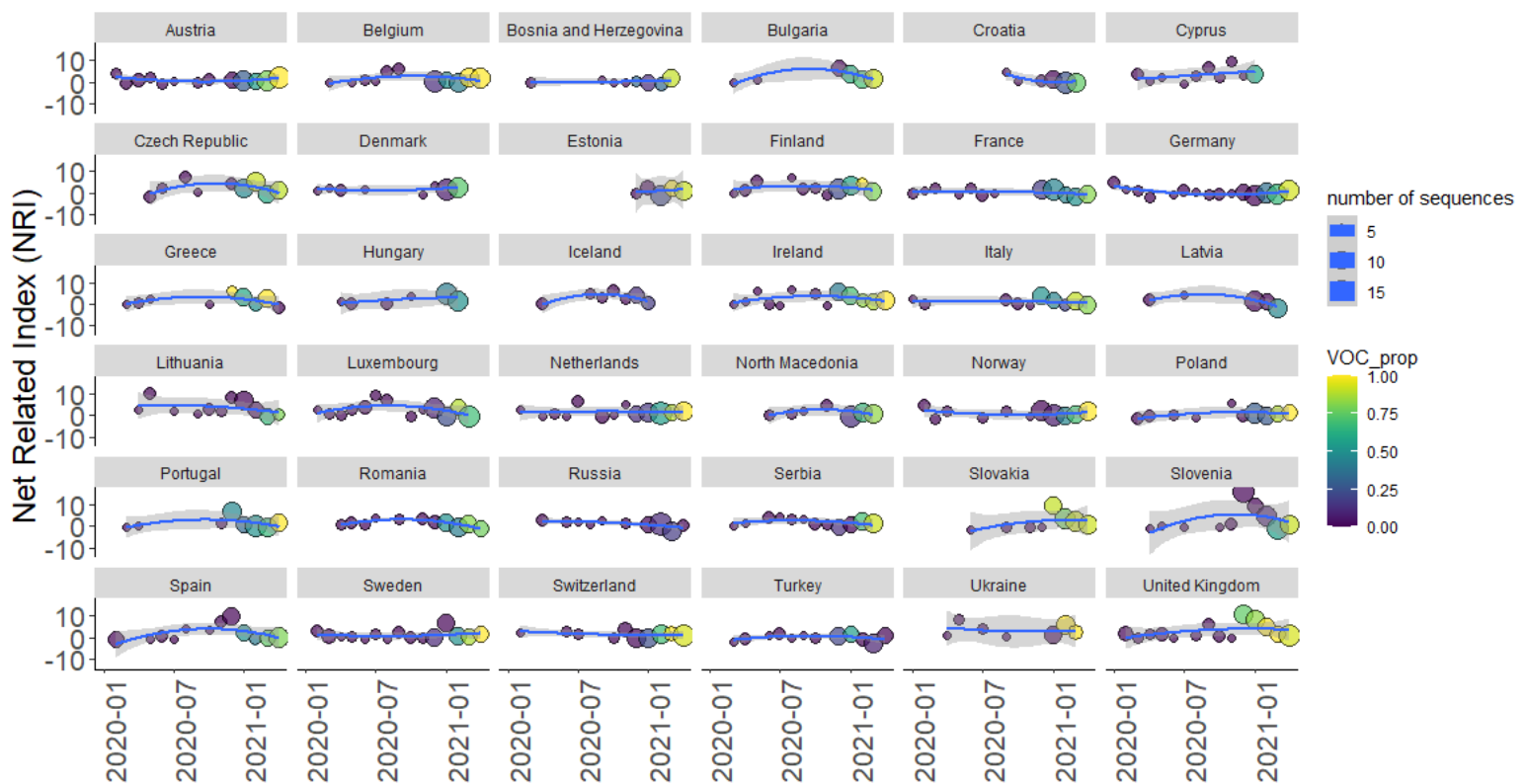


Regional german NRI



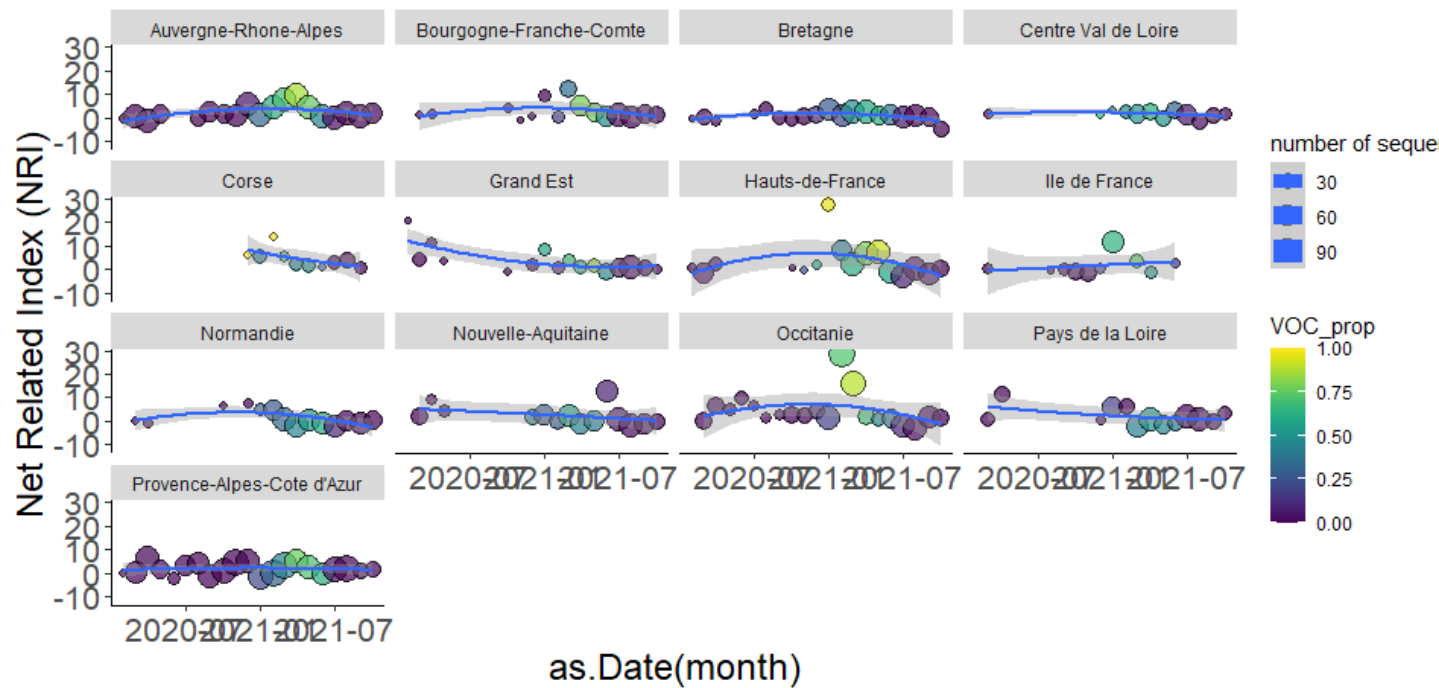
Each region kinda different. However regions like Saxony and Thuringia make a case for VOC to increase clustering at high proportions. But then it looks like in other regions the high VOC values arent correlated to higher NRI values

EU regions



Not really much of a pattern but what I can kind of see is it looks a bit like the middle prop VOC are relatively less clustered than pre emergence, not much evidence for NRI to ten increase when VOC at high props

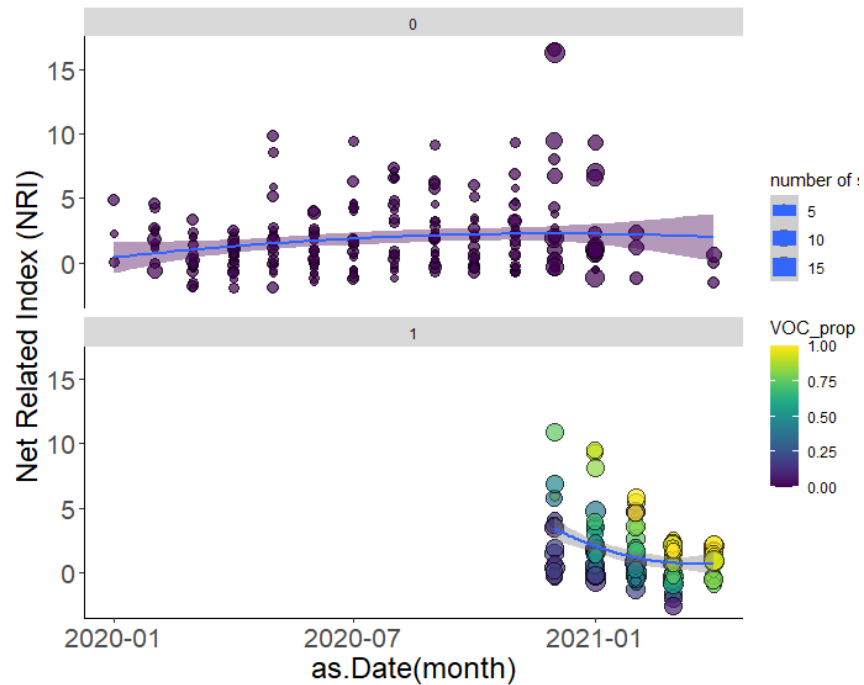
French regional NRI



With the french data you can kind of see that in some regions (ARA) the NRI values increase with the higher proportions of VOC (more clustered) which then decreases as the VOC decreases again. Would be useful to have more recent data for EU and German really

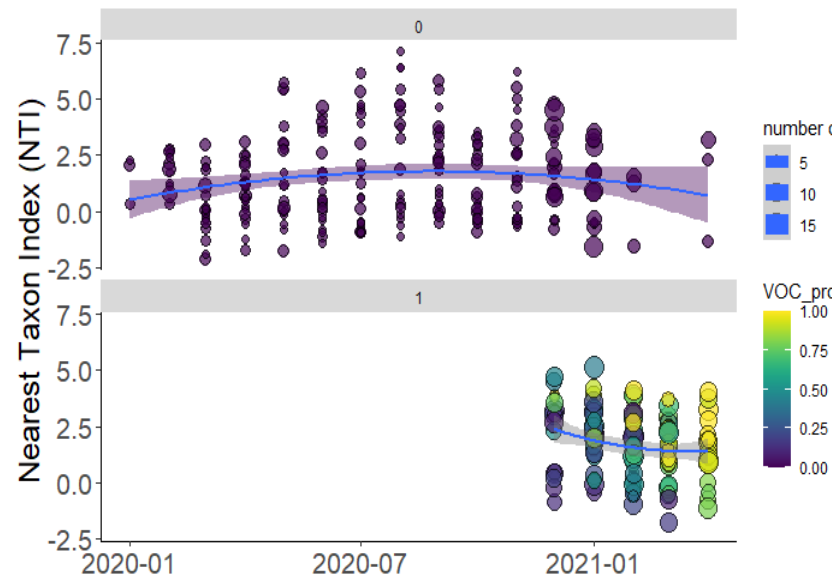
EU Background NRI dynamics

Want to know how the NRI changes through time without the emergence of VOC. with this then comparing it to how the NRI changes as the VOC spreads and afterwards. Top plot it regions with no VOC, kind of a flat/gradual upwards trend. Then the bottom plot shows the NRI once the VOC has emerged in regions which clearly shows a different trend



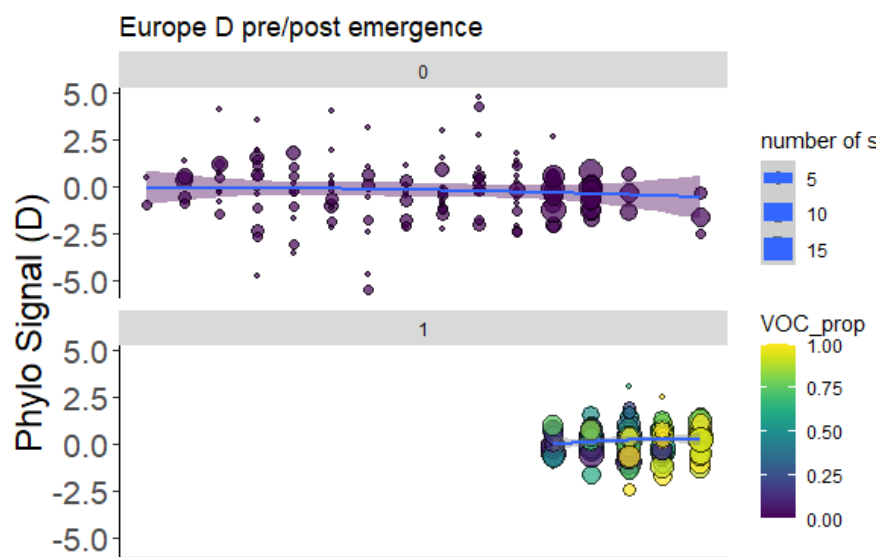
EU background NTI dynamics

NTI shows a similar pattern but the downwards trend can be seen in other regions (with no VOC recorded) after the emergence. Perhaps because regions dispersion would be affected pre detection. So the emergence of this new variant causes a decrease in clustering, I think this is driven by the regions that have middle proportions of the variant so are considerably more overdispersed, then as the variant rises to fixation we see an increase in clustering again



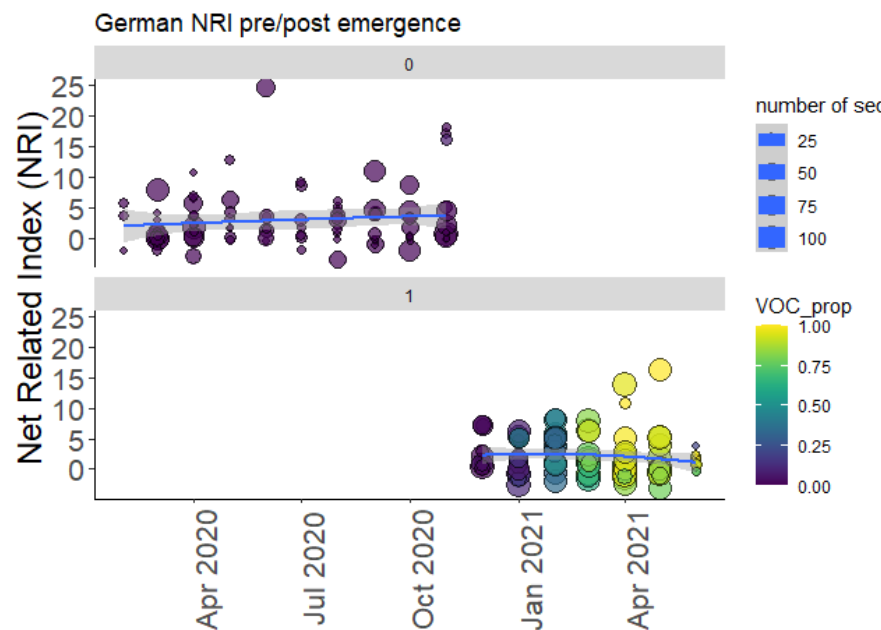
EU Background D (>1 = over, 1 = rand, 0 = brown, < 0 = clust)

D doesn't show the same VOC separation as NRI. although does display the pre emergence D, although variable is ~ 0 (consistent with brownian motion)



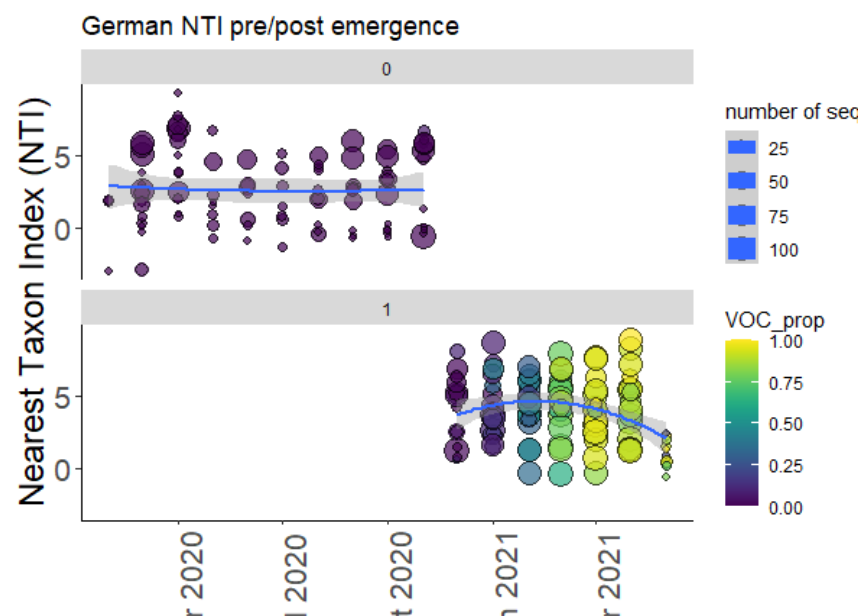
Ger background NRI dynamics

Similar pattern to the eu with pre emergence NRI trend fluctuating/flat compared to post emergence decrease in NRI but much more subtle



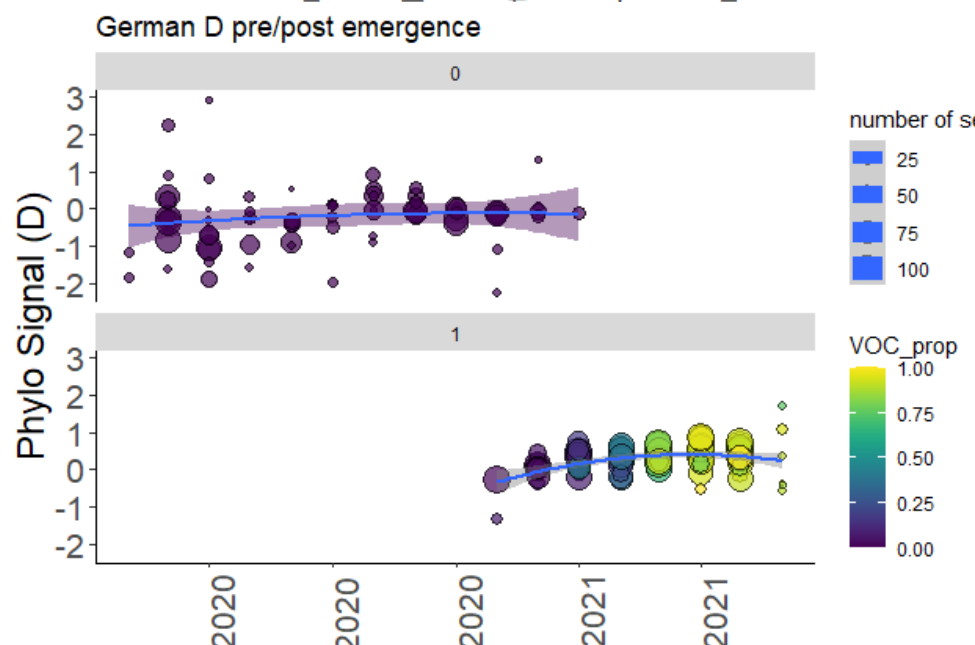
Ger background NTI dynamics

Within the plot of after the VOC emerges you can see the decrease in NTI for the middle proportions and then the increase again as the VOC dominates



German Background D

Trending at around 0 pre emergence then after emergence there is a slight increase towards overdispersion (but 0-1 = random). Don't see the decrease back towards clustering that you might expect in the higher VOC regions



Phylogenetic imbalance

Calculated colless' index for each month.
Seems almost entirely driven by the number of strains present, couldnt calculate it for europe data (more phylogeny quality?)
Fluctuations through time but mainly still dependent on the number of tips of the phylogeny

