

UK lineages summary report

This report gives summaries of lineages sampled in Northern_Ireland for week 2020-05-29. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-04-22. The analysis (eg time since last sample) is therefore undertaken from this date. 274 sequences from Northern_Ireland have been included in this analysis. 126 lineages have been recorded, 101 of which only contain one sequence.

A few notes: the size of a lineage may be due to a low amount of transmission of this lineage, but it is likely also that it just hasn't been sampled as frequently, especially for newer lineages. It's also important to realise that these lineages are *estimates* of how we think the virus is spreading in the UK after being introduced from abroad, as the low evolutionary rate of the virus makes it difficult to separate lineages with certainty.

The minimum number of introductions is 126 and the maximum is 146

Sequences which were replicates or too error-prone were removed from this analysis.

121 are lineages which were sampled less than five times in Northern_Ireland, and so have been left out of visualisation in the interests of clarity

Furthermore, those sequences which haven't been sampled in the last month are not shown.

Of the 5 that remain: 1 has reactivated. 4 lineages have been continuously circulating.

The following table contains information about the ten largest lineages and the number of sequences in the dataset. Information about other lineages is found in the appendix, along with the raw data for all of the other figures.

Each entry is the count of sequences from each lineage in each country, with the percentage of the total sequences from that lineage that this count represents.

"Activity score" is calculated by taking the average gap between sampling for each lineage, and dividing it by the number of days since the lineage was last sampled. Therefore the higher the number, the more active the lineage is. If the score is above 1, then it has been sampled *more* recently than expected given its average gap size. We might interpret this as an increase in activity. If the score is below 1, it has been sampled *less* recently than expected given its average gap size, so we might interpret this as a decrease in activity.

The global lineages are correct as of the data release on 2020-05-19

It is written to "summary_files" as "lineage_summary.tsv" for further use, and the full list of lineages is available in the same directory as "all_lineages.csv"

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK72	Mar-11, Apr-21	52	B.10	1	-0.0592
UK760	Mar-21, Apr-22	43	B.1.1	0	active today
UK295	Mar-11, Apr-22	10	B	0	active today
UK701	Mar-20, Mar-29	8	B.1	24	-0.0179
UK2904	Apr-06, Apr-22	8	B.1.p11	0	active today

These data is represented in the figure one. Note that the number of sequences is likely to be due more to

differing sampling efforts in different regions, rather than genuine differences in numbers of cases.

The raw data for this bar chart are in the table above.

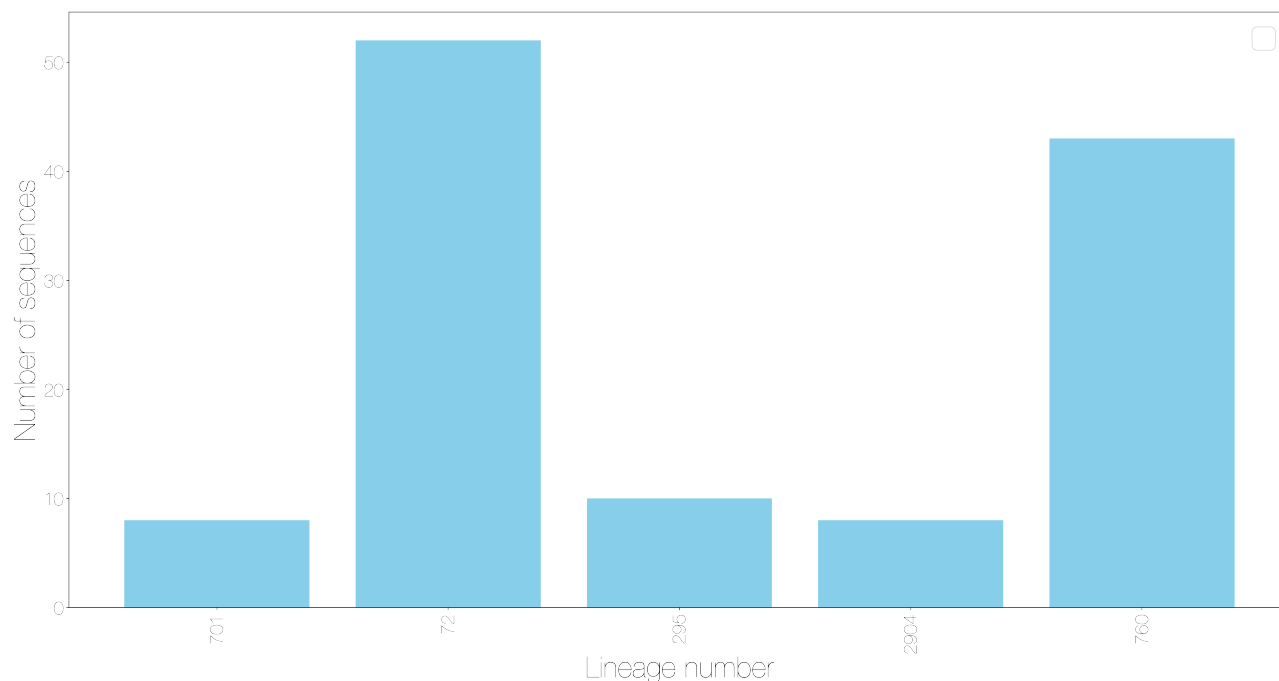


Figure 1: Number of sequences sampled in a lineage by country

Different sequencing centres have different delays in turn around from receipt of samples to submission of sequence data. This will affect all of the figures shown after this if lineages have geographical variation, as some regions have less up to date data.

The relative growth and decline of the ten most sampled lineages in terms of number of counties they are present in is shown in figure three.

These lineages are shown on the timeline. Each line represents the length of the cluster, from oldest to most recent sampling date. The dots are sized by the number of sequences taken on that date, and again are colour coded by country. The raw data has been written to a summary file.

The date of first sequence in the cluster is shown in figure five for every cluster with date information.

NB the lineage may have started anywhere in the UK, but has been recorded at least once in Northern_Ireland

For comparison, here is a plot of the day that every sequence was taken, coloured by country. Note that sequences without dates were not included.

The map shows the number of sequences sampled in each admin2 region in the UK. The colour scale is the same for all four countries, but with different underlying base colours.

There are some sequences with locations that are not matched to real Admin2 regions, some manual curation required.

Other results modules for UK lineage analysis can be added in here if required.

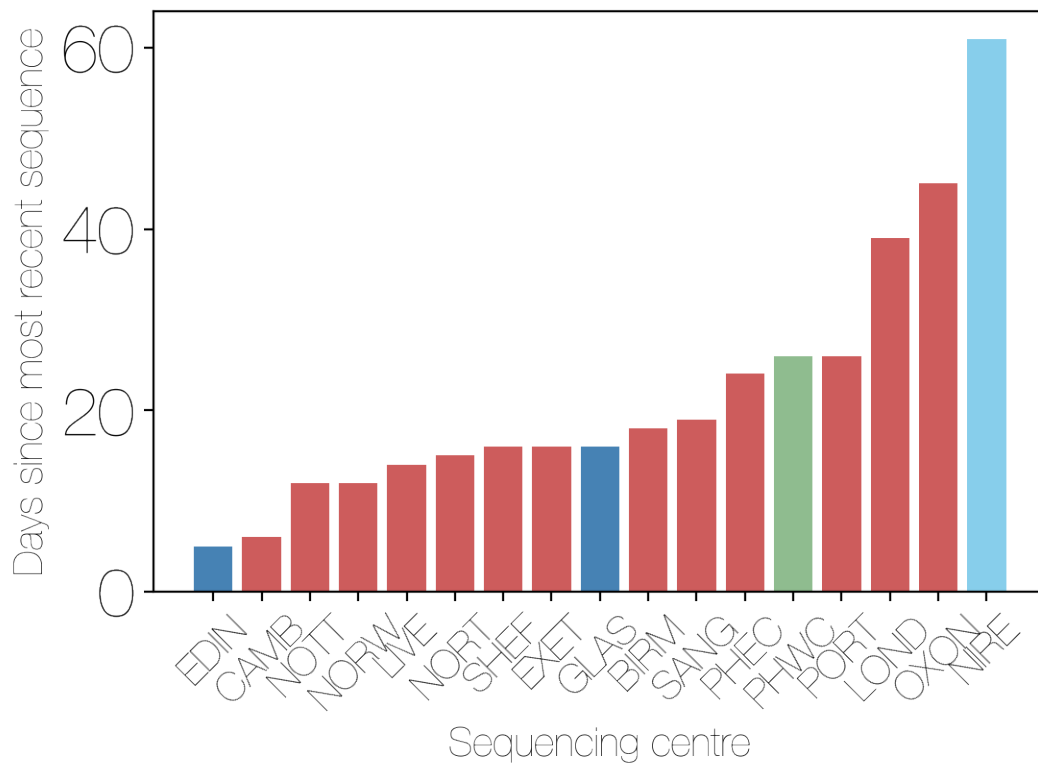


Figure 2: Lag since the most recent sequence from each sequencing centre to most current date

Appendix

Below are the raw data tables for each of the figures in the report.

Table S1 Description of all lineages that have been circulating in the last month, and have more than 5 sequences.

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK72	Mar-11, Apr-21	52	B.10	1	-0.0592
UK760	Mar-21, Apr-22	43	B.1.1	0	active today
UK295	Mar-11, Apr-22	10	B	0	active today
UK701	Mar-20, Mar-29	8	B.1	24	-0.0179
UK2904	Apr-06, Apr-22	8	B.1.p11	0	active today

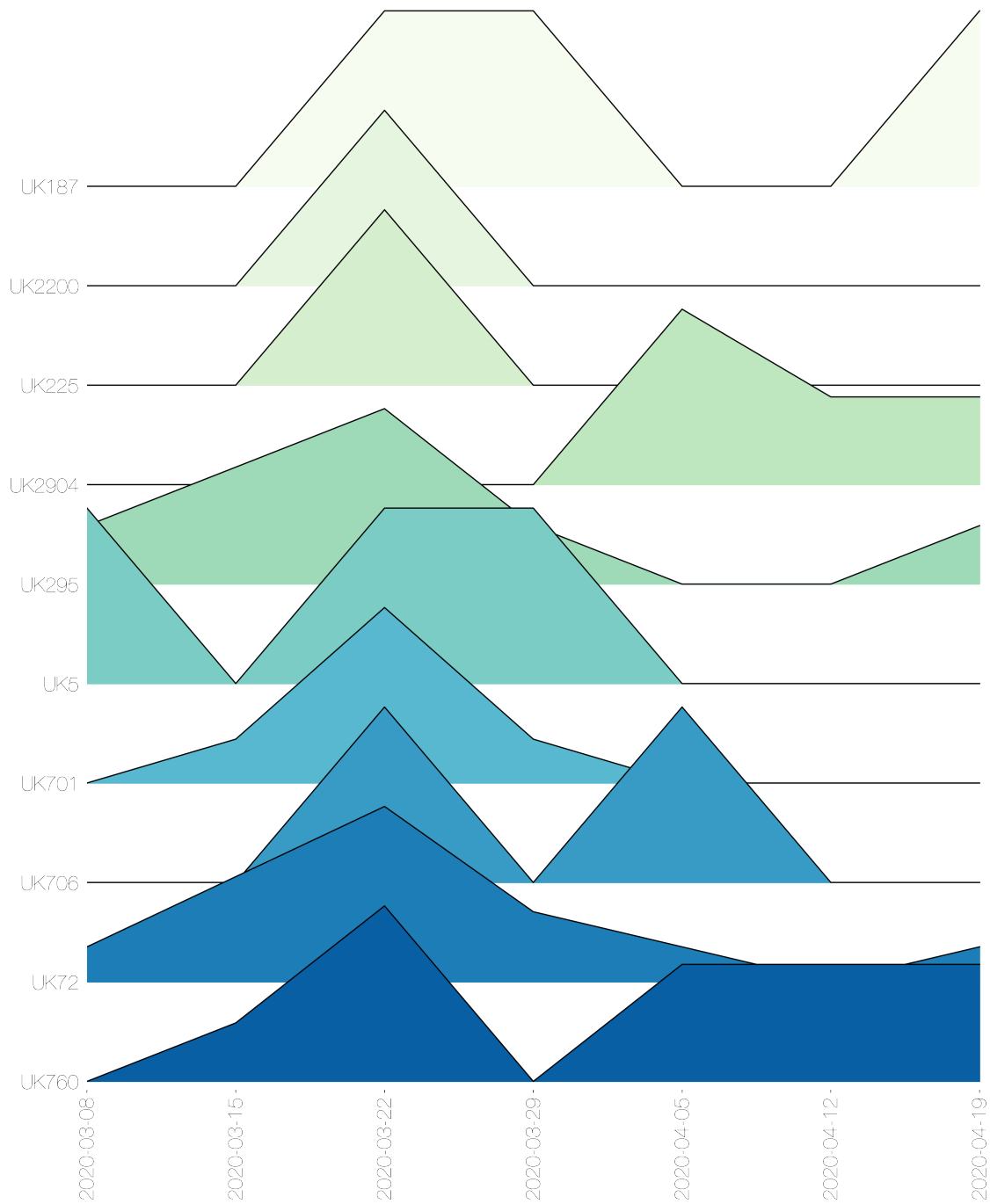


Figure 3: Lineages by number of adm2 regions present by epiweek

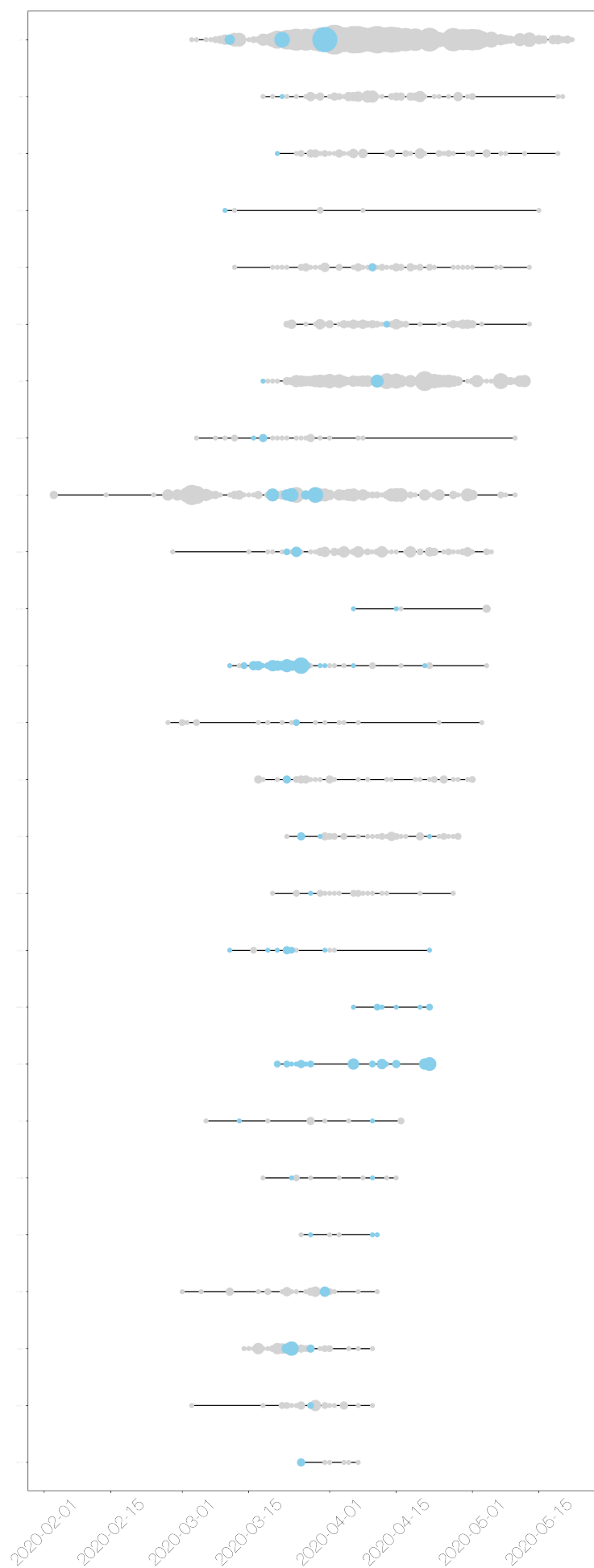
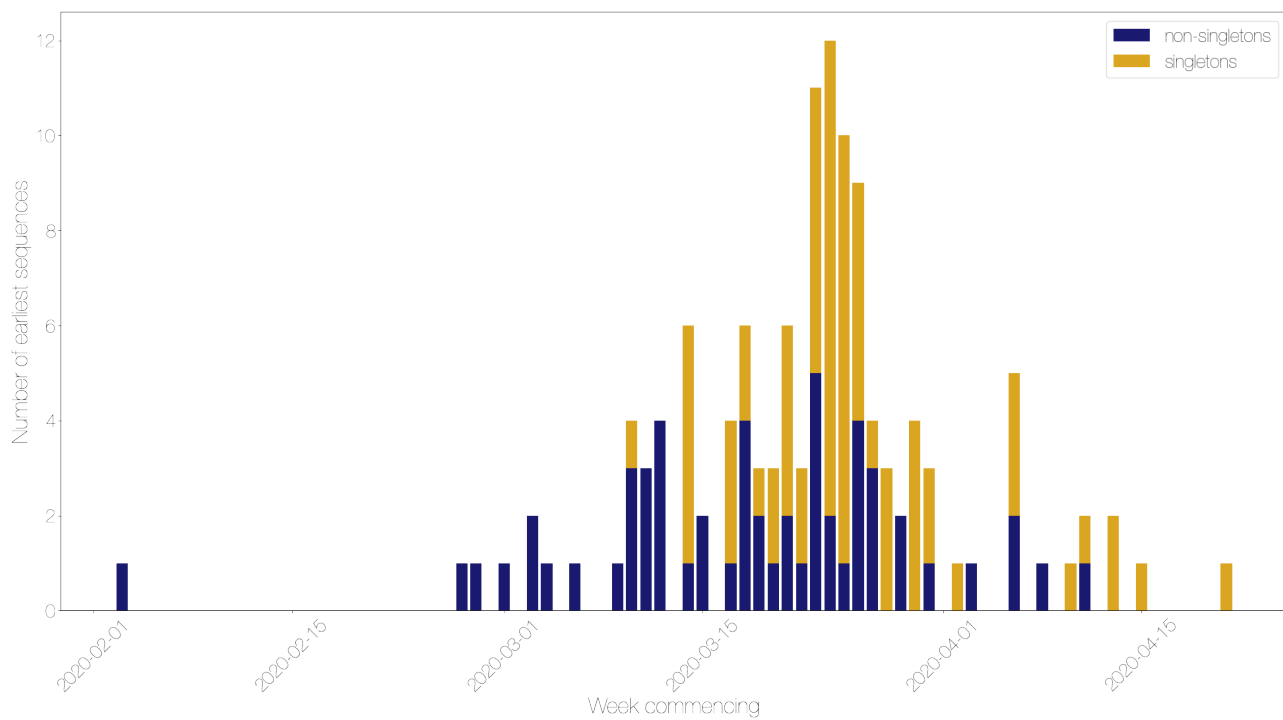


Figure 4: Timeline of lineages, sized by number of sequences from each country.



COVID-19 sequences from each Admn2 region in Northern_Ireland

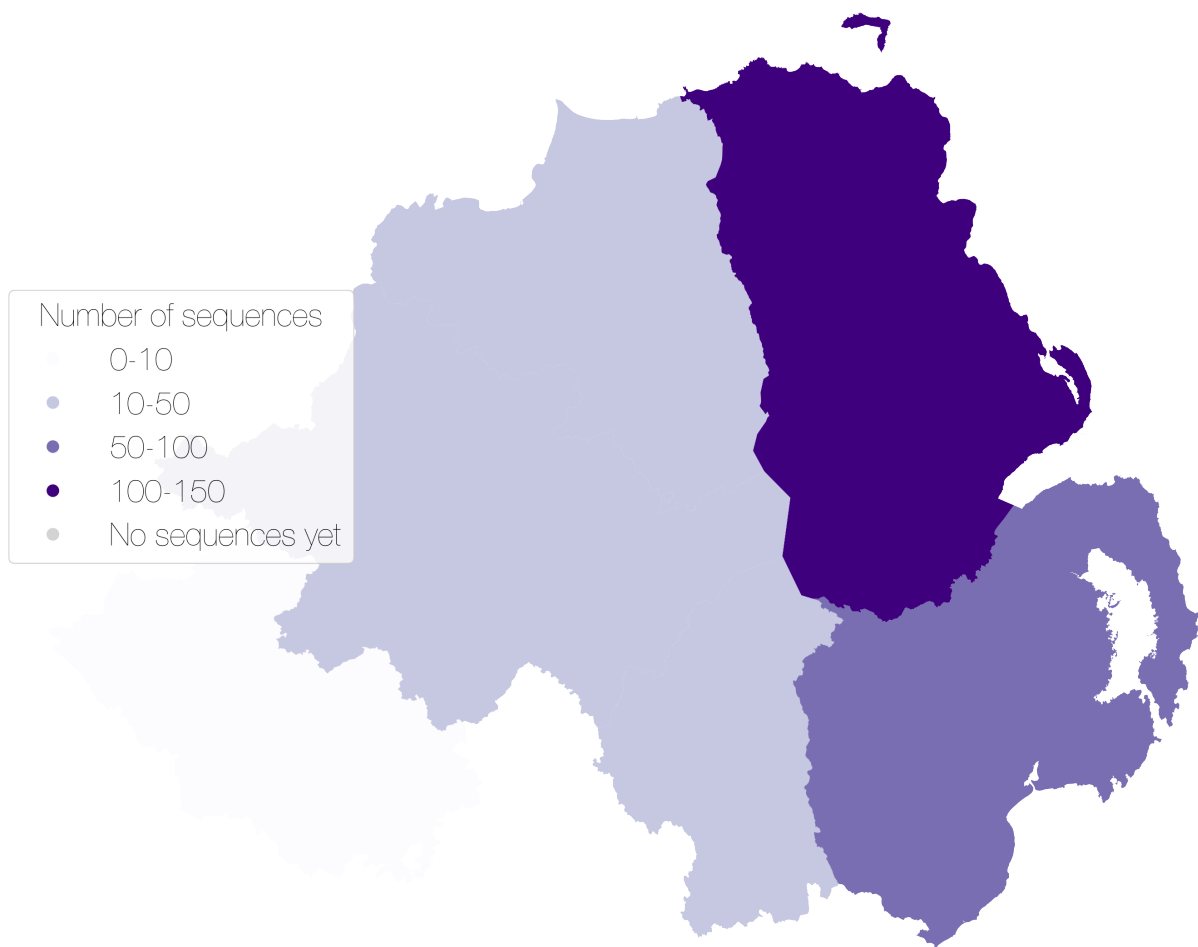


Figure 7: Map showing the number of sequences sampled by adm2 region

Table S2 Raw data for figure three showing the number of admin2 regions a lineage is present in over time

Week commencing	UK72	UK760	UK295	UK2904	UK701	UK187	UK225	UK706	UK2200	UK5
2020-03-08	1	0	1	0	0	0	0	0	0	1
2020-03-15	3	1	2	0	1	0	0	0	0	0
2020-03-22	5	3	3	0	4	1	1	1	1	1
2020-03-29	2	0	1	0	1	1	0	0	0	1
2020-04-05	1	2	0	2	0	0	0	1	0	0
2020-04-12	0	2	0	1	0	0	0	0	0	0
2020-04-19	1	2	1	1	0	1	0	0	0	0

Table S3 is not appropriate for this report and so has been omitted.

Table S4 Raw data for figure six showing when lineages started per day, divided by singletons and non-singletons

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-02-03	0	1	1
2020-02-27	0	1	1
2020-02-28	0	1	1
2020-03-01	0	1	1
2020-03-03	0	2	2
2020-03-04	0	1	1
2020-03-06	0	1	1
2020-03-09	0	1	1
2020-03-10	1	3	4
2020-03-11	0	3	3
2020-03-12	0	4	4
2020-03-14	5	1	6
2020-03-15	0	2	2
2020-03-17	3	1	4
2020-03-18	2	4	6
2020-03-19	1	2	3
2020-03-20	2	1	3
2020-03-21	4	2	6
2020-03-22	2	1	3
2020-03-23	6	5	11
2020-03-24	10	2	12
2020-03-25	9	1	10
2020-03-26	5	4	9
2020-03-27	1	3	4
2020-03-28	3	0	3
2020-03-29	0	2	2
2020-03-30	4	0	4
2020-03-31	2	1	3
2020-04-02	1	0	1
2020-04-03	0	1	1
2020-04-06	3	2	5
2020-04-08	0	1	1
2020-04-10	1	0	1
2020-04-11	1	1	2
2020-04-13	2	0	2
2020-04-15	1	0	1
2020-04-21	1	0	1

Table S5 Raw data for figure seven showing the number of sequences taken over time.

Day	Northern Ireland
2020-03-10	2
2020-03-11	3
2020-03-13	1
2020-03-14	6
2020-03-16	5
2020-03-17	7
2020-03-18	6
2020-03-19	3
2020-03-20	6
2020-03-21	13
2020-03-22	8
2020-03-23	29
2020-03-24	23
2020-03-25	16
2020-03-26	27
2020-03-27	7
2020-03-28	11
2020-03-29	5
2020-03-30	6
2020-03-31	7
2020-04-02	1
2020-04-04	1
2020-04-06	13
2020-04-10	11
2020-04-11	8
2020-04-12	7
2020-04-13	5
2020-04-15	8
2020-04-19	1
2020-04-20	2
2020-04-21	12
2020-04-22	14

Table S6 Raw data for the map with the number of sequences assigned to each admin2 region.

Admin2	Country	Number of sequences	Sequence group
ANTRIM	Northern Ireland	140	100-150
ARMAGH	Northern Ireland	12	10-50
DOWN	Northern Ireland	90	50-100
FERMANAGH	Northern Ireland	3	1-10
LONDONDERRY	Northern Ireland	15	10-50
TYRONE	Northern Ireland	15	10-50