Lineages report for Northern Ireland

This report gives summaries of lineages sampled in Northern Ireland for week 2020-09-13. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-08-26. The analysis (eg time since last sample) is therefore undertaken from this date. 1528 sequences from Northern_Ireland have been included in this analysis. 157 lineages have been recorded, 96 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 6 and the maximum is 419

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

136 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 21 that remain: 1 is pending extinction ie last seen three weeks ago. 2 lineages have gone quiet, ie haven't been seen this week. 2 lineages have reactivated. 16 lineages have been continuously circulating.

The following table contains information about the ten largest lineages lineages and the number of sequences 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 2020-07-20

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	Total	Global lineage	Time since last sample	Activity score
UK5	Mar-10, Aug-26	607 taxa	B.1.1, B.1.1.10	0 days	active today
UK1535	Mar-25, Aug-26	191 taxa	B.1.1	0 days	active today
UK1855	Jul-12, Aug-26	147 taxa	B.1.1	0 days	active today
UK601	Mar-11, May-15	71 taxa	B, B.10	103 days	0.0055
UK167	Mar-14, Jul-15	31 taxa	B.1	42 days	0.007
UK2913	Apr-06, Jul-15	19 taxa	B.1.11	42 days	0.0043
UK175	Mar-22, Aug-21	18 taxa	B.1, B.1.5	5 days	0.0137
UK107	Mar-17, Jul-20	18 taxa	B.2.1	37 days	0.0024
UK1266	Mar-25, Jun-17	17 taxa	B.1.1	70 days	0.0304
UK461	Aug-15, Aug-26	15 taxa	B.1	0 days	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.

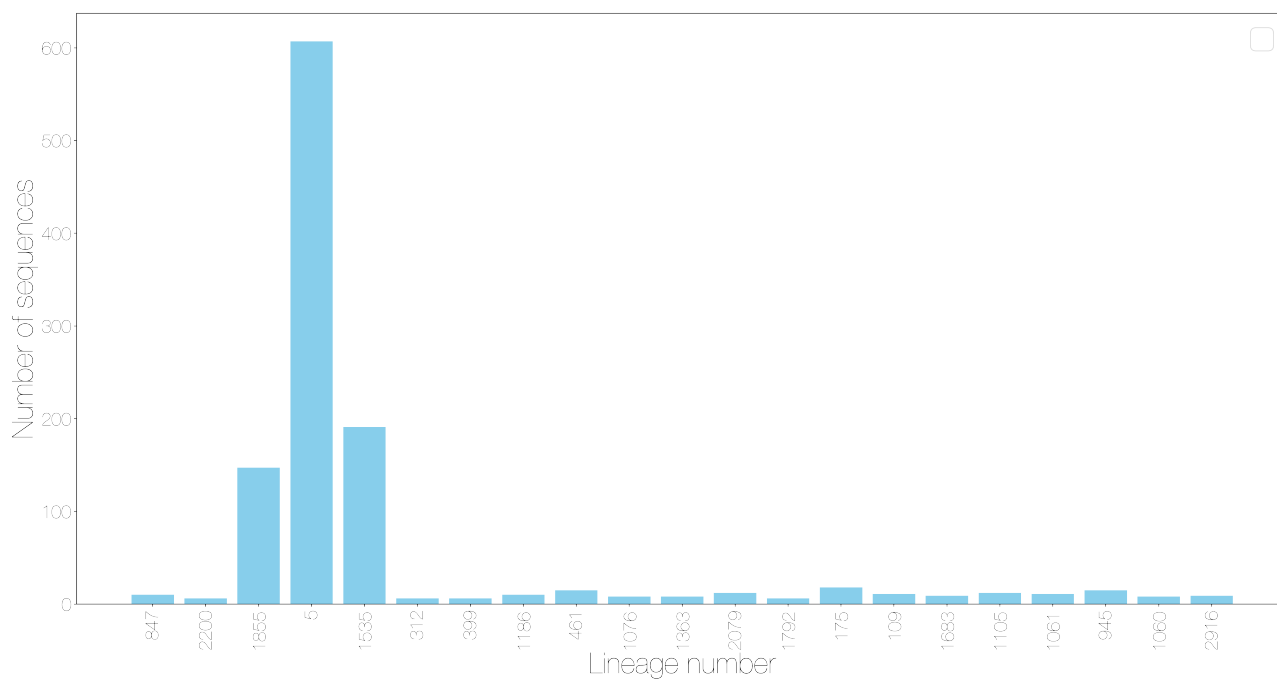The raw data for this bar chart are in the table above.

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.

Figure 2: Lineages by number of adm2 regions present by epiweek
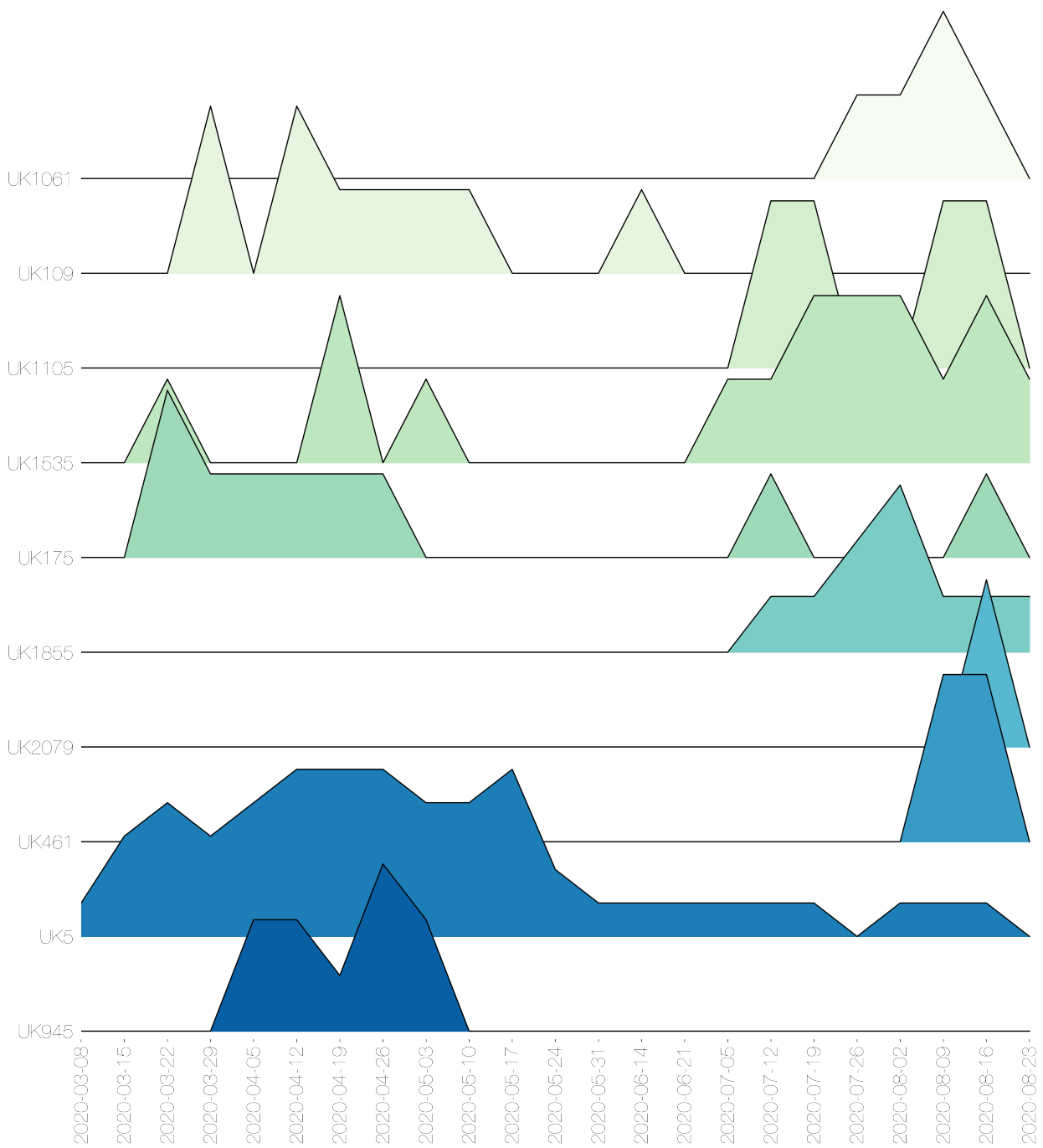
These lineages are shown on the timeline. Each line represents the length of the cluster, from oldest to most recent sampling date. The dots are sized by the number of sequences taken on that date, and again are colour coded by country. The raw data has been written to a summary file.
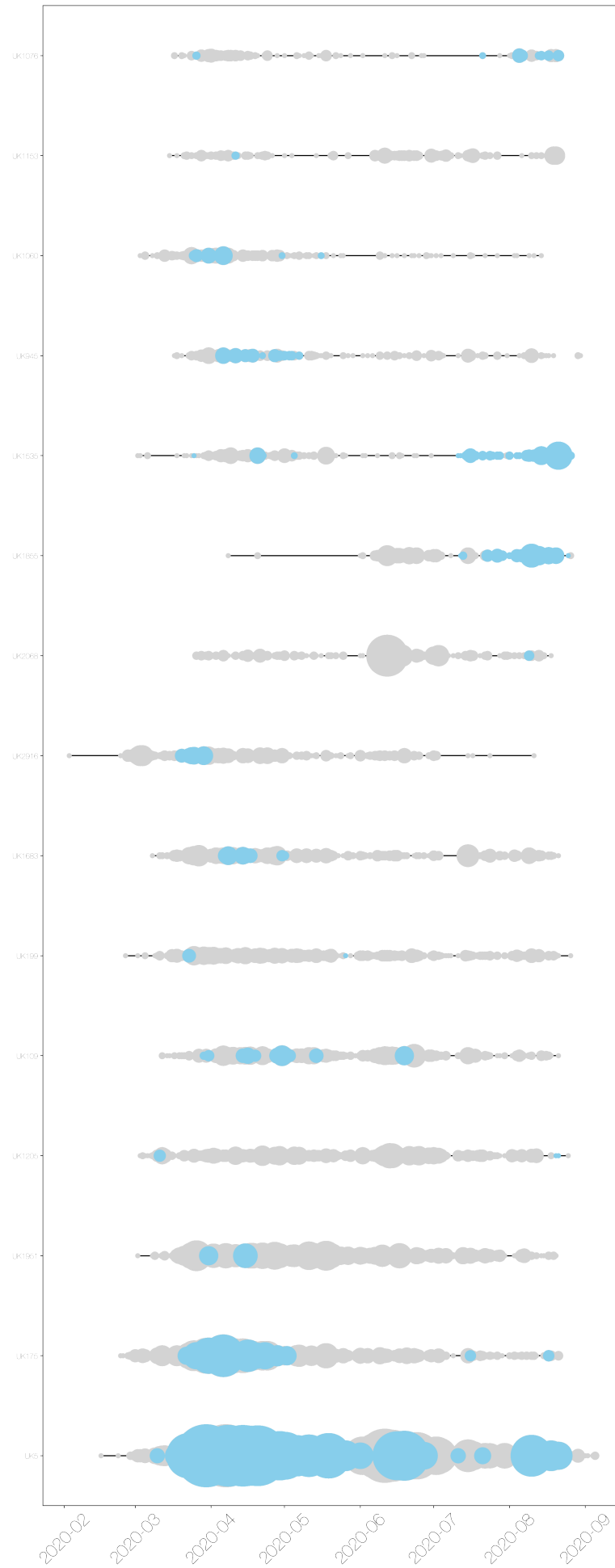
Figure 3: Timeline of lineages, sized by number of sequences from each country.

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
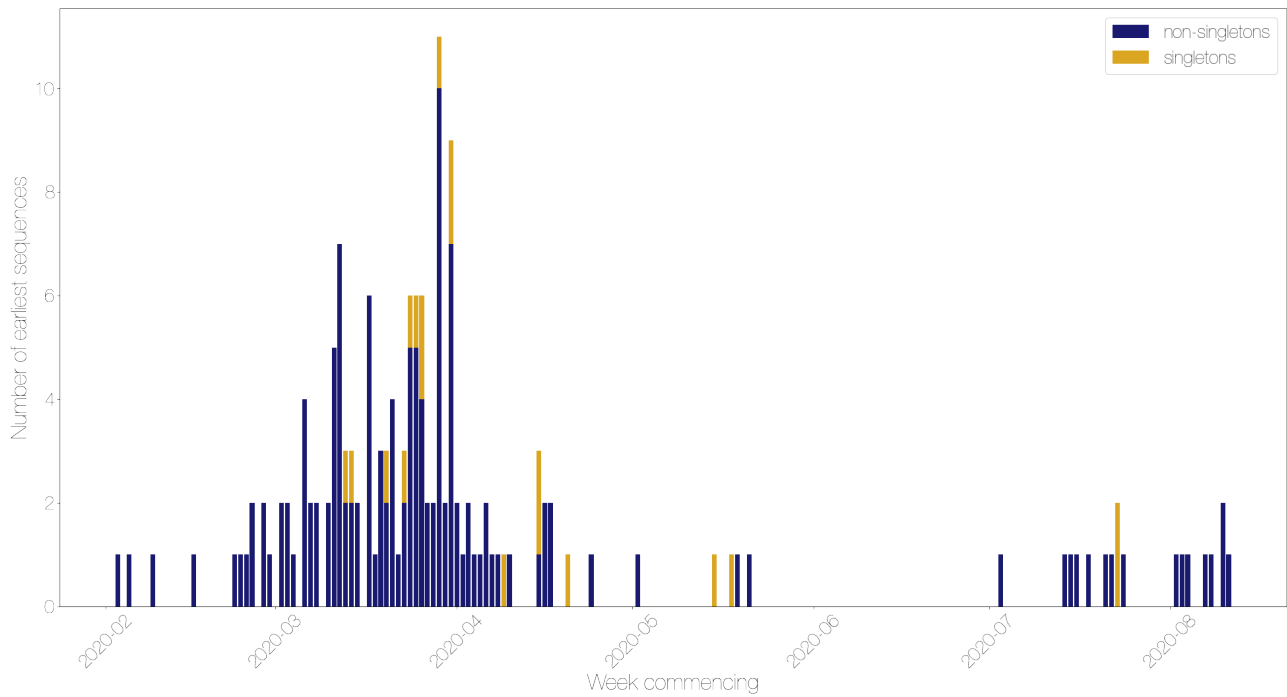
Figure 4: Lineage starts per week, split by singletons and non-singletons

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.
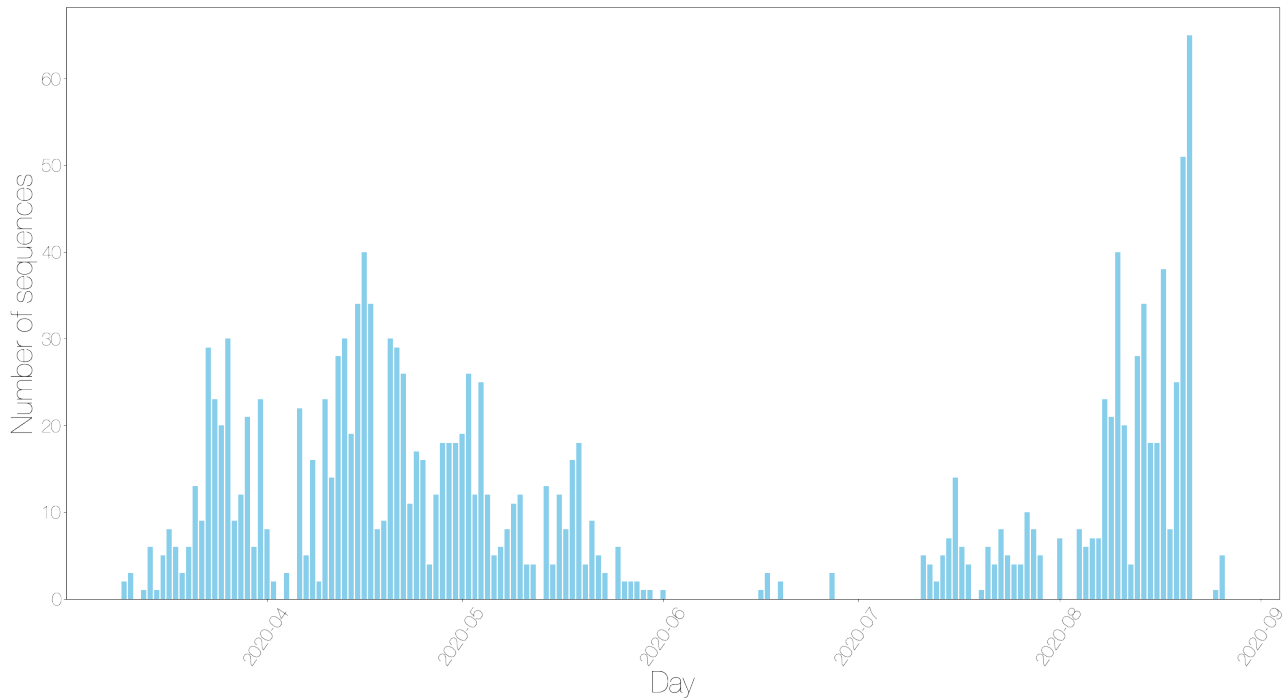
Figure 5: Sequences taken on each day by country

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.
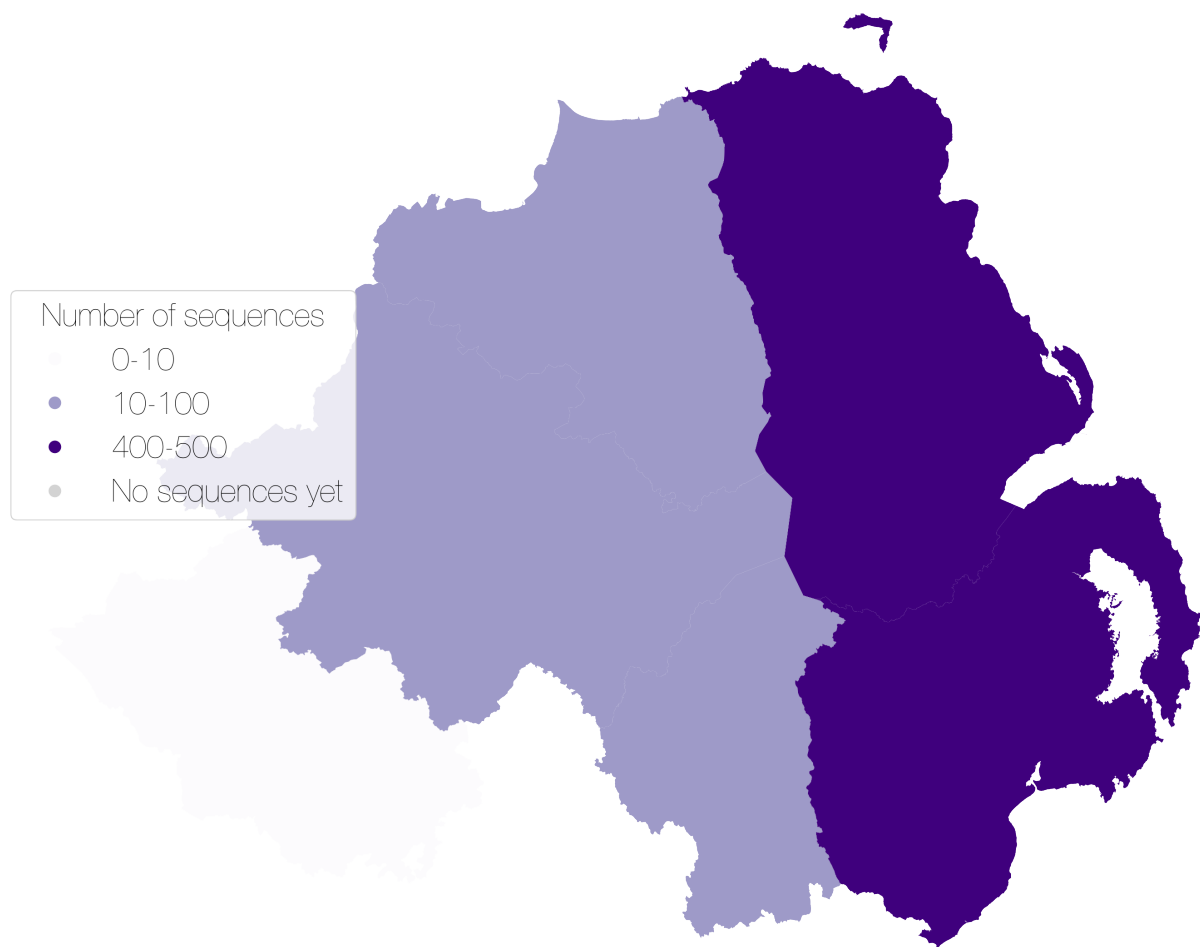
Figure 6: Map showing the number of sequences sampled by adm2 region

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.

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	Total	Global lineage	Time since last sample	Activity score
UK5	Mar-10, Aug-26	607 taxa	B.1.1, B.1.1.10	0 days	active today
UK1535	Mar-25, Aug-26	191 taxa	B.1.1	0 days	active today
UK1855	Jul-12, Aug-26	147 taxa	B.1.1	0 days	active today
UK601	Mar-11, May-15	71 taxa	B, B.10	103 days	0.0055
UK167	Mar-14, Jul-15	31 taxa	B.1	42 days	0.007
UK2913	Apr-06, Jul-15	19 taxa	B.1.11	42 days	0.0043
UK175	Mar-22, Aug-21	18 taxa	B.1, B.1.5	5 days	0.0137

Lineage name	Date range	Total	Global lineage	Time since last sample	Activity score
UK107	Mar-17, Jul-20	18 taxa	B.2.1	37 days	0.0024
UK1266	Mar-25, Jun-17	17 taxa	B.1.1	70 days	0.0304
UK461	Aug-15, Aug-26	15 taxa	B.1	0 days	active today
UK187	Mar-26, Jun-05	15 taxa	B.1, B.1.77	82 days	0.0112
UK945	Apr-06, Aug-19	15 taxa	B.1.1	7 days	0.0632
UK72	Mar-11, Jun-23	14 taxa	B, B.2	64 days	0.0047
UK2079	Aug-19, Aug-21	12 taxa	B.1.1, B.1.1.25	5 days	2.1385
UK1105	Jul-14, Aug-20	12 taxa	B.1	6 days	0.4211
UK109	Mar-29, Aug-21	11 taxa	B.1	5 days	0.0342
UK1061	Jul-28, Aug-20	11 taxa	B.1	6 days	0.6944
UK1179	May-03, May-16	10 taxa	B.1.1	102 days	0.0471
UK871	Mar-31, May-01	10 taxa	B.1.1	117 days	0.0265
UK847	Aug-04, Aug-26	10 taxa	B.1.36	0 days	active today
UK1186	Aug-05, Aug-26	10 taxa	B.1, B.1.79	0 days	active today
UK1683	Apr-07, Aug-21	9 taxa	B.1.1.1	5 days	0.0408
UK2916	Mar-20, Aug-11	9 taxa	B.1	15 days	0.0214
UK1060	Mar-25, Aug-14	8 taxa	B.1.1	12 days	0.0415
UK1363	Jul-16, Aug-21	8 taxa	B.1	5 days	2.7
UK1076	Mar-26, Aug-22	8 taxa	B.1.1	4 days	0.1606
UK1792	Aug-07, Aug-21	6 taxa	B.1.1	5 days	2.35
UK2204	Apr-11, Jun-10	6 taxa	B.1	77 days	0.0952
UK312	Jul-21, Aug-26	6 taxa	B.1	0 days	active today
UK399	Aug-10, Aug-26	6 taxa	B.1	0 days	active today
UK2200	Mar-23, Aug-26	6 taxa	B.1.5	0 days	active today

Table S2 Raw data for figure two showing lags between the most recent sequence and current date for each sequencing centre

—————————————————————————————NameError Traceback (most recent call last) in 1 if not pillar2: —-> 2 lag_df = pd.DataFrame(lag_dict) 3 print(lag_df.to_markdown()) 4 else: 5 print("Table S2 is not appropriate for this report and so has been omitted.") NameError: name 'lag_dict' is not defined

Table S3 Raw data for figure three showing the number of admin2 regions a lineage is present in over time

Week commencing	UK5	UK1535	UK1855	UK175	UK945	UK461	UK1105	UK2079	UK1061	UK109
2020-03-08	1	0	0	0	0	0	0	0	0	0
2020-03-15	3	0	0	0	0	0	0	0	0	0
2020-03-22	4	1	0	2	0	0	0	0	0	0
2020-03-29	3	0	0	1	0	0	0	0	0	2
2020-04-05	4	0	0	1	2	0	0	0	0	0

Week commencing	UK5	UK1535	UK1855	UK175	UK945	UK461	UK1105	UK2079	UK1061	UK109
2020-04-12	5	0	0	1	2	0	0	0	0	2
2020-04-19	5	2	0	1	1	0	0	0	0	1
2020-04-26	5	0	0	1	3	0	0	0	0	1
2020-05-03	4	1	0	0	2	0	0	0	0	1
2020-05-10	4	0	0	0	0	0	0	0	0	1
2020-05-17	5	0	0	0	0	0	0	0	0	0
2020-05-24	2	0	0	0	0	0	0	0	0	0
2020-05-31	1	0	0	0	0	0	0	0	0	0
2020-06-14	1	0	0	0	0	0	0	0	0	1
2020-06-21	1	0	0	0	0	0	0	0	0	0
2020-07-05	1	1	0	0	0	0	0	0	0	0
2020-07-12	1	1	1	1	0	0	1	0	0	0
2020-07-19	1	2	1	0	0	0	1	0	0	0
2020-07-26	0	2	2	0	0	0	0	0	1	0
2020-08-02	1	2	3	0	0	0	0	0	1	0
2020-08-09	1	1	1	0	0	1	1	0	2	0
2020-08-16	1	2	1	1	0	1	1	1	1	0
2020-08-23	0	1	1	0	0	0	0	0	0	0

Table S4 is not appropriate for this report and so has been omitted.

Table S5 Raw data for figure five 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
2020-02-05	0	1	1
2020-02-09	0	1	1
2020-02-16	0	1	1
2020-02-23	0	1	1
2020-02-24	0	1	1
2020-02-25	0	1	1
2020-02-26	0	2	2
2020-02-28	0	2	2
2020-02-29	0	1	1
2020-03-02	0	2	2
2020-03-03	0	2	2
2020-03-04	0	1	1
2020-03-06	0	4	4
2020-03-07	0	2	2
2020-03-08	0	2	2
2020-03-10	0	2	2
2020-03-11	0	5	5
2020-03-12	0	7	7
2020-03-13	1	2	3
2020-03-14	1	2	3
2020-03-15	0	2	2
2020-03-17	0	6	6
2020-03-18	0	1	1
2020-03-19	0	3	3
2020-03-20	1	2	3
2020-03-21	0	4	4
2020-03-22	0	1	1
2020-03-23	1	2	3
2020-03-24	1	5	6
2020-03-25	1	5	6
2020-03-26	2	4	6
2020-03-27	0	2	2
2020-03-28	0	2	2

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-03-29	1	10	11
2020-03-30	0	2	2
2020-03-31	2	7	9
2020-04-01	0	2	2
2020-04-02	0	1	1
2020-04-03	0	2	2
2020-04-04	0	1	1
2020-04-05	0	1	1
2020-04-06	0	2	2
2020-04-07	0	1	1
2020-04-08	0	1	1
2020-04-09	1	0	1
2020-04-10	0	1	1
2020-04-15	2	1	3
2020-04-16	0	2	2
2020-04-17	0	2	2
2020-04-20	1	0	1
2020-04-24	0	1	1
2020-05-02	0	1	1
2020-05-15	1	0	1
2020-05-18	1	0	1
2020-05-19	0	1	1
2020-05-21	0	1	1
2020-07-03	0	1	1
2020-07-14	0	1	1
2020-07-15	0	1	1
2020-07-16	0	1	1
2020-07-18	0	1	1
2020-07-21	0	1	1
2020-07-22	0	1	1
2020-07-23	2	0	2
2020-07-24	0	1	1
2020-08-02	0	1	1
2020-08-03	0	1	1
2020-08-04	0	1	1
2020-08-07	0	1	1
2020-08-08	0	1	1
2020-08-10	0	2	2
2020-08-11	0	1	1

Table S6 Raw data for figure six 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
2020-03-15	1
2020-03-16	5
2020-03-17	8
2020-03-18	6
2020-03-19	3
2020-03-20	6
2020-03-21	13
2020-03-22	9
2020-03-23	29
2020-03-24	23
2020-03-25	20

Day	Northern Ireland
2020-03-26	30
2020-03-27	9
2020-03-28	12
2020-03-29	21
2020-03-30	6
2020-03-31	23
2020-04-01	8
2020-04-02	2
2020-04-04	3
2020-04-06	22
2020-04-07	5
2020-04-08	16
2020-04-09	2
2020-04-10	23
2020-04-11	14
2020-04-12	28
2020-04-13	30
2020-04-14	19
2020-04-15	34
2020-04-16	40
2020-04-17	34
2020-04-18	8
2020-04-19	9
2020-04-20	30
2020-04-21	29
2020-04-22	26
2020-04-23	11
2020-04-24	17
2020-04-25	16
2020-04-26	4
2020-04-27	12
2020-04-28	18
2020-04-29	18
2020-04-30	18
2020-05-01	19
2020-05-02	26
2020-05-03	12
2020-05-04	25
2020-05-05	12
2020-05-06	5
2020-05-07	6
2020-05-08	8
2020-05-09	11
2020-05-10	12
2020-05-11	4
2020-05-12	4
2020-05-14	13
2020-05-15	4
2020-05-16	12
2020-05-17	8
2020-05-18	16
2020-05-19	18
2020-05-20	4
2020-05-21	9
2020-05-22	5
2020-05-23	3
2020-05-25	6
2020-05-26	2
2020-05-27	2

Day	Northern Ireland
2020-05-28	2
2020-05-29	1
2020-05-30	1
2020-06-01	1
2020-06-16	1
2020-06-17	3
2020-06-19	2
2020-06-27	3
2020-07-11	5
2020-07-12	4
2020-07-13	2
2020-07-14	5
2020-07-15	7
2020-07-16	14
2020-07-17	6
2020-07-18	4
2020-07-20	1
2020-07-21	6
2020-07-22	4
2020-07-23	8
2020-07-24	5
2020-07-25	4
2020-07-26	4
2020-07-27	10
2020-07-28	8
2020-07-29	5
2020-08-01	7
2020-08-04	8
2020-08-05	6
2020-08-06	7
2020-08-07	7
2020-08-08	23
2020-08-09	21
2020-08-10	40
2020-08-11	20
2020-08-12	4
2020-08-13	28
2020-08-14	34
2020-08-15	18
2020-08-16	18
2020-08-17	38
2020-08-18	8
2020-08-19	25
2020-08-20	51
2020-08-21	65
2020-08-25	1
2020-08-26	5

Table S7 Raw data for the figure seven with the number of sequences assigned to each admin2 region.

Admin2	Country	Number of sequences	Sequence group
ANTRIM	Northern Ireland	461	400-500
ARMAGH	Northern Ireland	44	10-100
DOWN	Northern Ireland	400	400-500
FERMANAGH	Northern Ireland	5	1-10
LONDONDERRY	Northern Ireland	50	10-100
TYRONE	Northern Ireland	35	10-100