Lineages report for EDIN

This report gives summaries of UK specific lineages sequenced by EDIN for week 2020-06-19. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-06-08. The analysis (eg time since last sample) is therefore undertaken from this date. 1283 sequences in the UK from the sequencing centre EDIN have been included in this analysis.

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 68 and the maximum is 485

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

92 are lineages which only contained five sequences or fewer, 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 15 that remain: 10 are pending extinction, ie last seen three weeks ago. 1 has reactivated. 4 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 expect 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	Scotland	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK109	227 (100.0%)	Mar-21, Jun-08	227	B.1.5, B.1.5.5	0	active today
UK36	159 (100.0%)	Mar-21, May-22	159	B.1	17	0.0231
UK42	127 (100.0%)	Mar-04, May-04	127	B.1.71, B.1	35	0.0138
UK5	107 (100.0%)	Mar-11, Jun-07	107	B.1.1, B.1.1.p12, B.1.1.1	1	0.8302
UK199	105 (100.0%)	Mar-20, Jun-06	105	B.1.5	2	0.375
UK2464	54 (100.0%)	Mar-20, Jun-06	54	B.1.p11	2	0.7358
UK3929	31 (100.0%)	Mar-06, May-21	31	B.1.1, B.1.1.4	18	0.1407
UK72	30 (100.0%)	Mar-11, Apr-25	30	B	44	0.0353
UK304	30 (100.0%)	Apr-16, Jun-02	30	B.1.1.14	6	0.2701
UK2913	28 (100.0%)	Mar-28, May-13	28	B.1.p11	26	0.0655

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.

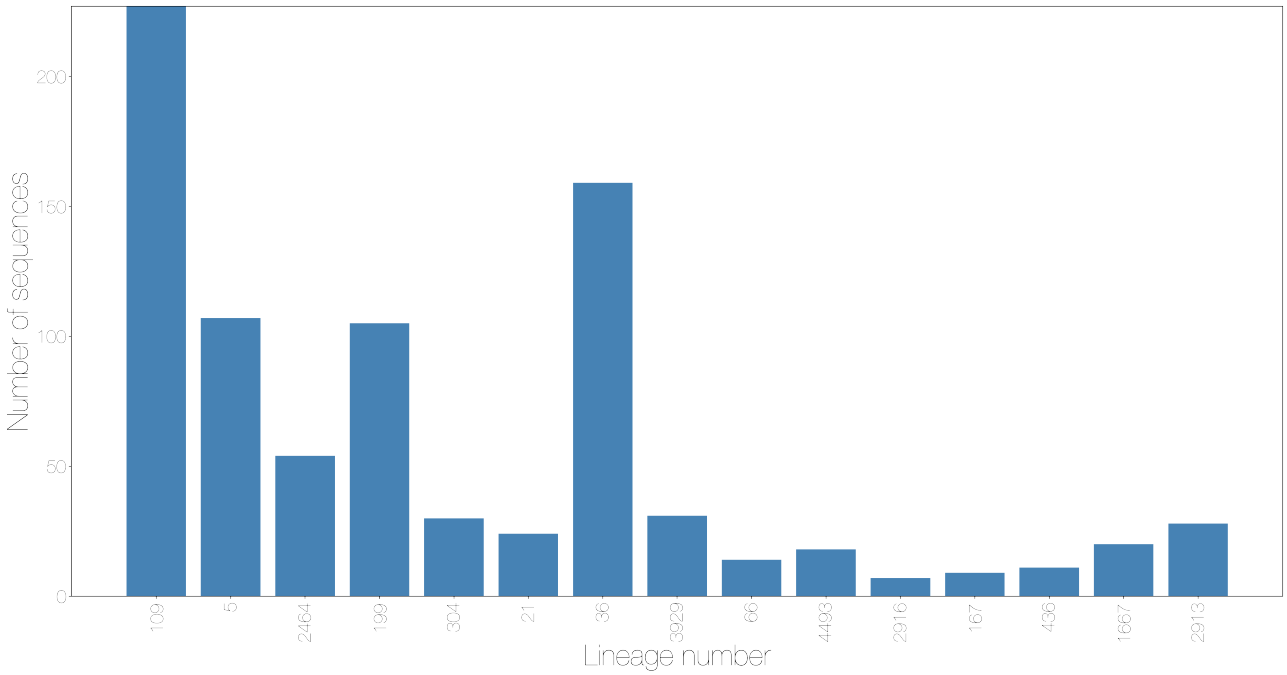

Figure 1: Number of sequences sampled in a lineage by country

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 lag for this sequencing centre is 11 days`

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.

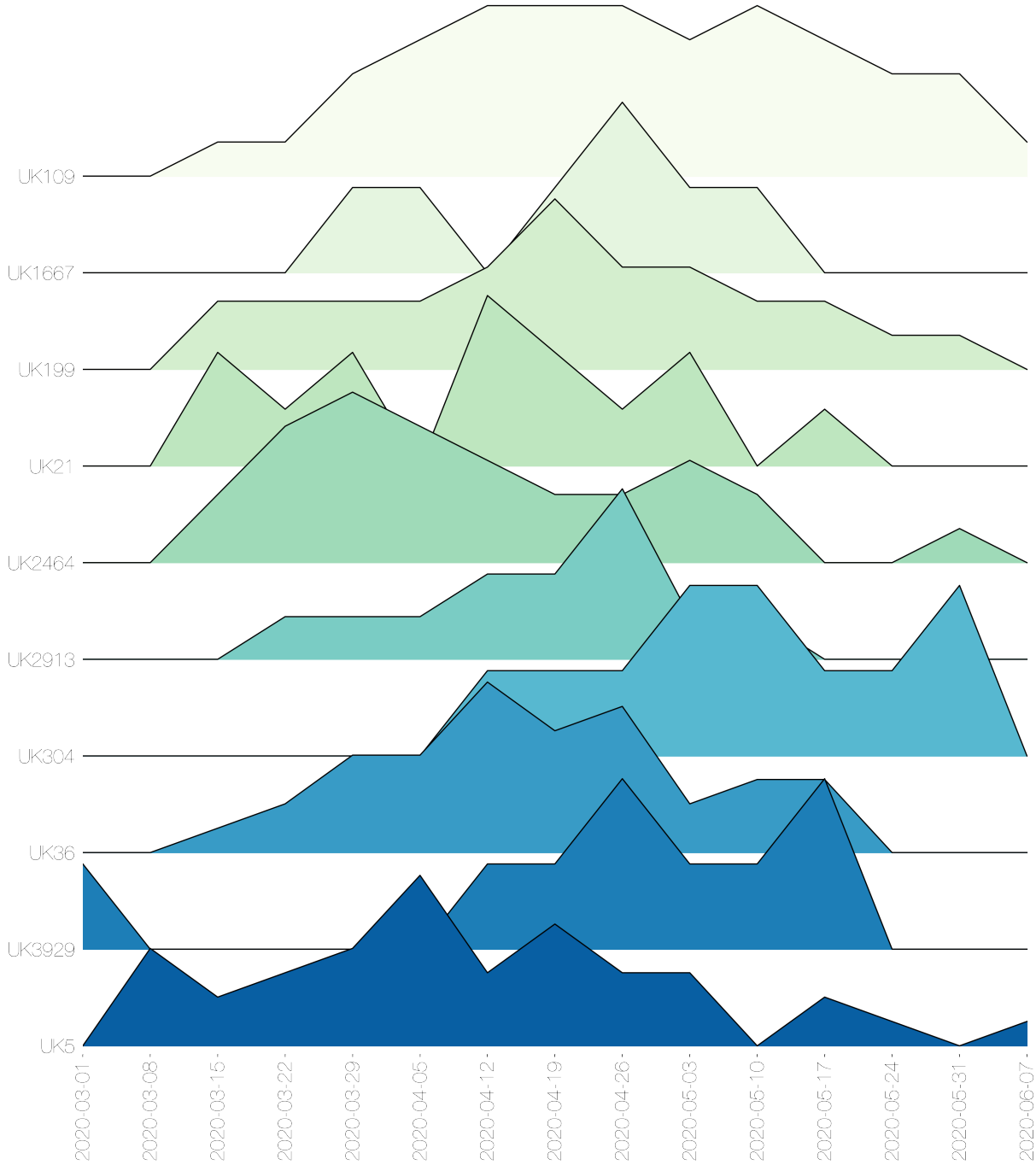

Figure 2: Lineages by number of adm2 regions present by epiweek

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.

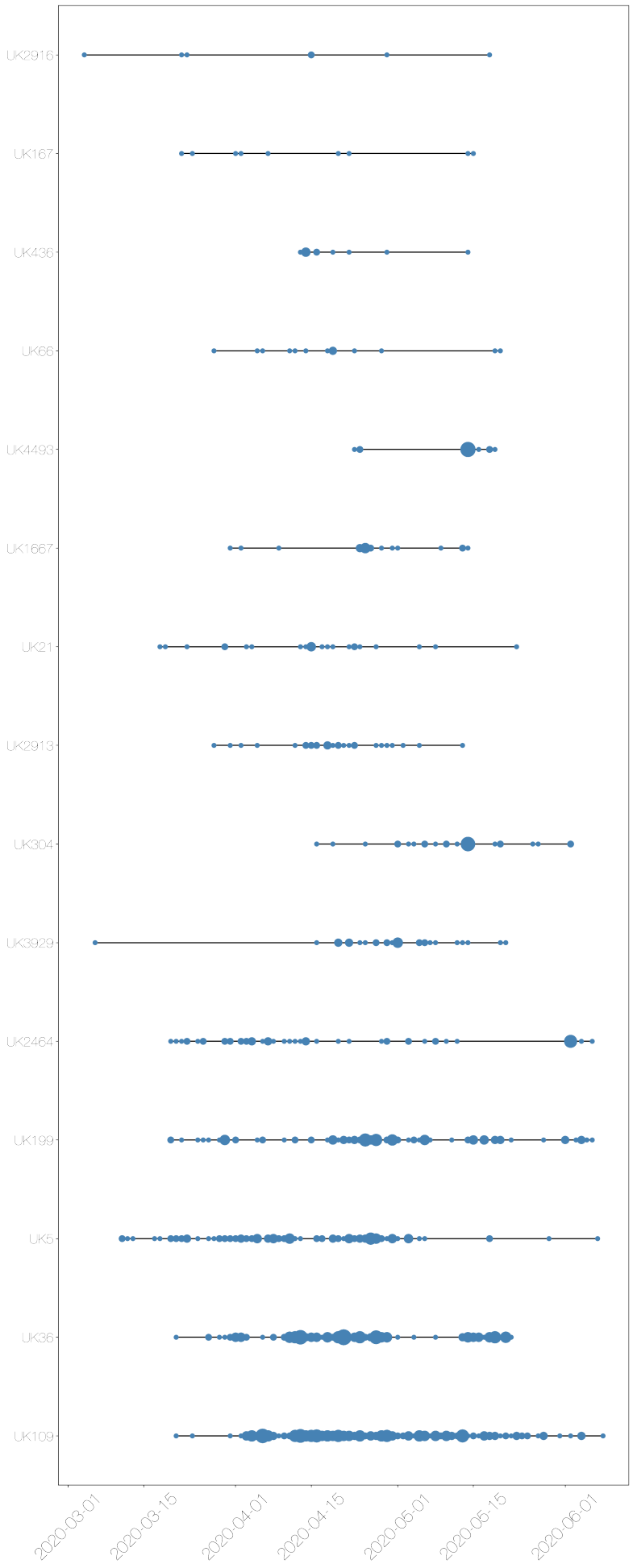

Figure 3: Timeline of lineages, sized by number of sequences from each country.

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

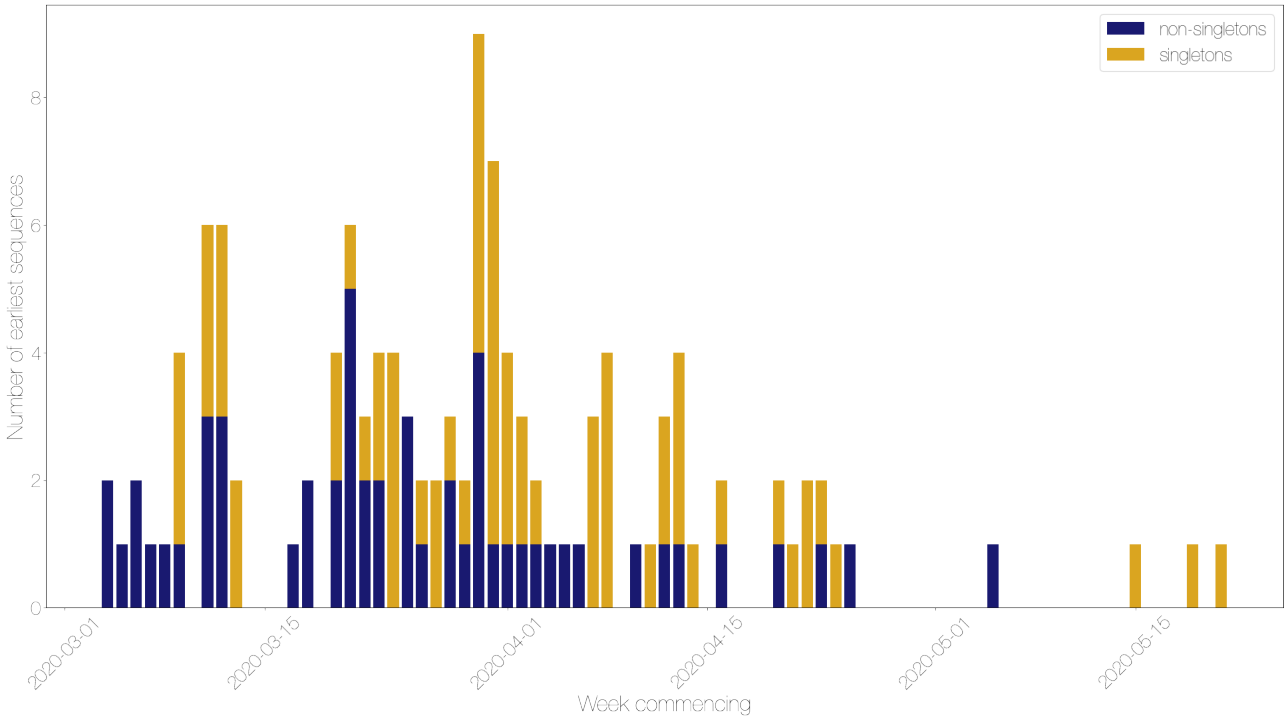

Figure 4: Lineage starts per week, split by singletons and non-singletons

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.

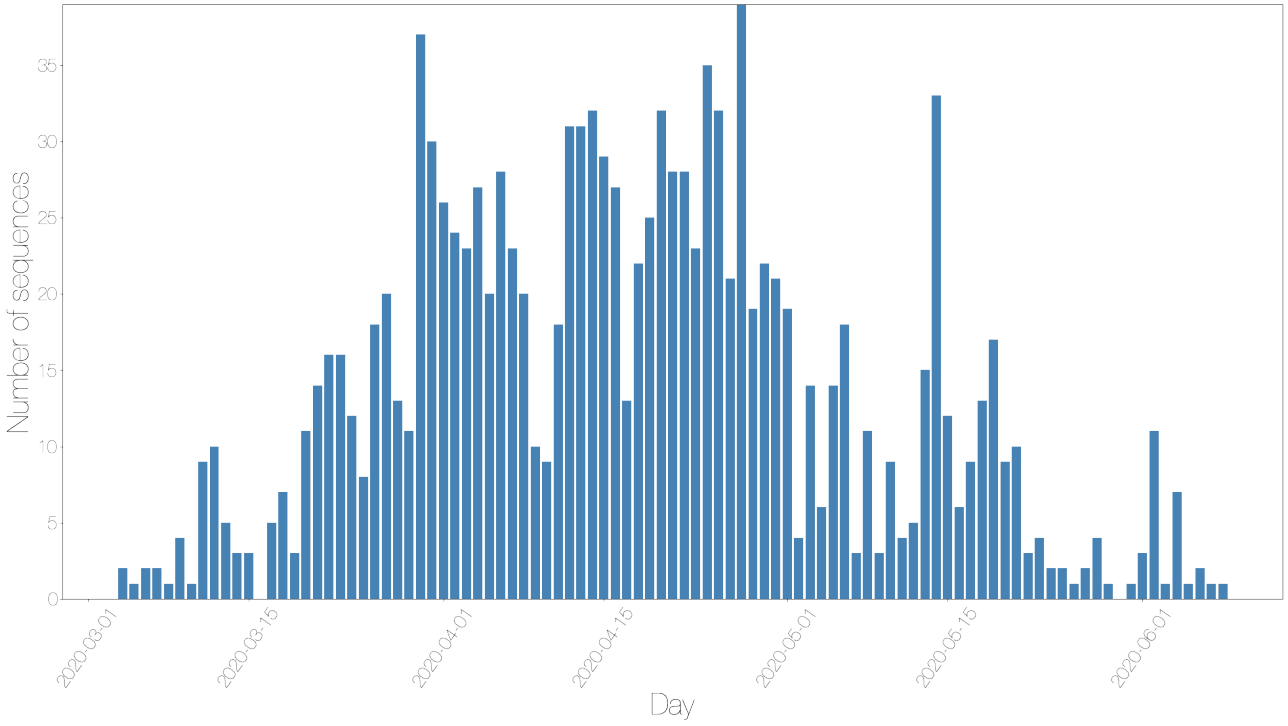

Figure 5: Sequences taken on each day by country

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.

```
There are 127 sequences without enough geographical information to map
```

from this centre.

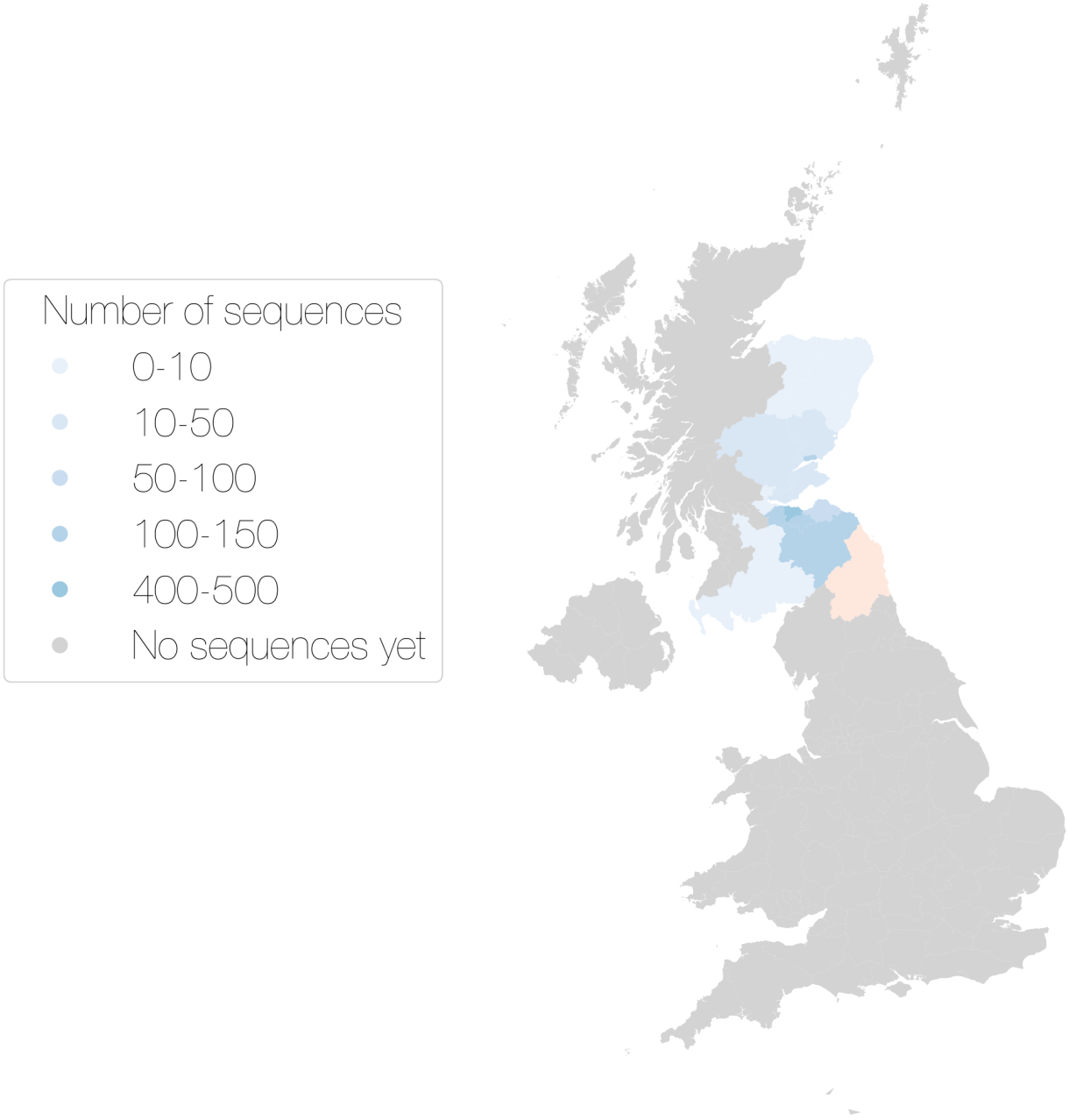

Figure 6: Map showing the number of sequences sampled by adm2 region

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	Scotland	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK109	227 (100.0%)	Mar-21, Jun-08	227	B.1.5, B.1.5.5	0	active today
UK36	159 (100.0%)	Mar-21, May-22	159	B.1	17	0.0231
UK42	127 (100.0%)	Mar-04, May-04	127	B.1.71, B.1	35	0.0138
UK5	107 (100.0%)	Mar-11, Jun-07	107	B.1.1, B.1.1.p12, B.1.1.1	1	0.8302
UK199	105 (100.0%)	Mar-20, Jun-06	105	B.1.5	2	0.375
UK2464	54 (100.0%)	Mar-20, Jun-06	54	B.1.p11	2	0.7358
UK3929	31 (100.0%)	Mar-06, May-21	31	B.1.1, B.1.1.4	18	0.1407
UK72	30 (100.0%)	Mar-11, Apr-25	30	B	44	0.0353
UK304	30 (100.0%)	Apr-16, Jun-02	30	B.1.1.14	6	0.2701
UK2913	28 (100.0%)	Mar-28, May-13	28	B.1.p11	26	0.0655
UK5676	24 (100.0%)	Mar-17, Apr-27	24	B.2	42	0.0424
UK21	24 (100.0%)	Mar-18, May-23	24	B.1.40	16	0.1793
UK1667	20 (100.0%)	Mar-31, May-14	20	B.1.p9	25	0.0926
UK2735	19 (100.0%)	Mar-18, May-02	19	B.1.1	37	0.0676
UK4493	18 (100.0%)	Apr-23, May-19	18	B.1	20	0.0765
UK107	16 (100.0%)	Mar-21, Apr-24	16	B.2.1	45	0.0504
UK2200	15 (100.0%)	Mar-25, May-05	15	B.1.5, B.1.5.6	34	0.0861
UK66	14 (100.0%)	Mar-28, May-20	14	B.1.1.8	19	0.2146
UK44	13 (100.0%)	Mar-25, Apr-19	13	B	50	0.0417
UK43	13 (100.0%)	Mar-26, Apr-18	13	A.5	51	0.0376
UK436	11 (100.0%)	Apr-13, May-14	11	B.1.5	25	0.124
UK198	10 (100.0%)	Apr-05, Apr-22	10	B.1.5	47	0.0402
UK14	9 (100.0%)	Mar-21, Apr-27	9	B	42	0.1101
UK5498	9 (100.0%)	Mar-12, Apr-27	9	B.2, B	42	0.1369
UK167	9 (100.0%)	Mar-22, May-15	9	B.1	24	0.2812
UK594	8 (100.0%)	Apr-20, May-01	8	B	38	0.0414

Lineage name	Scotland	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK2916	7 (100.0%)	Mar-04, May-18	7	B.1	21	0.5952
UK133	6 (100.0%)	Mar-22, Apr-25	6	B.1	44	0.1545
UK3692	6 (100.0%)	Mar-12, Apr-15	6	B.1.1	54	0.1259

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

	Centre	Lag in days
0	EDIN	11

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

Week commencing	UK109	UK36	UK5	UK199	UK2464	UK3929	UK304	UK2913	UK21	UK1667
2020-03-01	0	0	0	0	0	1	0	0	0	0
2020-03-08	0	0	4	0	0	0	0	0	0	0
2020-03-15	1	1	2	2	2	0	0	0	2	0
2020-03-22	1	2	3	2	4	0	0	1	1	0
2020-03-29	3	4	4	2	5	0	0	1	2	1
2020-04-05	4	4	7	2	4	0	0	1	0	1
2020-04-12	5	7	3	3	3	1	1	2	3	0
2020-04-19	5	5	5	5	2	1	1	2	2	1
2020-04-26	5	6	3	3	2	2	1	4	1	2
2020-05-03	4	2	3	3	3	1	2	1	2	1
2020-05-10	5	3	0	2	2	1	2	1	0	1
2020-05-17	4	3	2	2	0	2	1	0	1	0
2020-05-24	3	0	1	1	0	0	1	0	0	0
2020-05-31	3	0	0	1	1	0	2	0	0	0
2020-06-07	1	0	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
2020-03-04	0	2	2
2020-03-05	0	1	1
2020-03-06	0	2	2
2020-03-07	0	1	1
2020-03-08	0	1	1
2020-03-09	3	1	4
2020-03-11	3	3	6
2020-03-12	3	3	6
2020-03-13	2	0	2
2020-03-17	0	1	1
2020-03-18	0	2	2
2020-03-20	2	2	4
2020-03-21	1	5	6
2020-03-22	1	2	3
2020-03-23	2	2	4
2020-03-24	4	0	4
2020-03-25	0	3	3
2020-03-26	1	1	2
2020-03-27	2	0	2
2020-03-28	1	2	3
2020-03-29	1	1	2
2020-03-30	5	4	9
2020-03-31	6	1	7
2020-04-01	3	1	4
2020-04-02	2	1	3
2020-04-03	1	1	2
2020-04-04	0	1	1
2020-04-05	0	1	1
2020-04-06	0	1	1
2020-04-07	3	0	3
2020-04-08	4	0	4
2020-04-10	0	1	1
2020-04-11	1	0	1
2020-04-12	2	1	3
2020-04-13	3	1	4
2020-04-14	1	0	1
2020-04-16	1	1	2
2020-04-20	1	1	2
2020-04-21	1	0	1
2020-04-22	2	0	2
2020-04-23	1	1	2
2020-04-24	1	0	1
2020-04-25	0	1	1
2020-05-05	0	1	1
2020-05-15	1	0	1
2020-05-19	1	0	1
2020-05-21	1	0	1

Table S6 Raw data for figure six showing the number of sequences taken over time.

Day	Scotland
2020-03-04	2
2020-03-05	1
2020-03-06	2
2020-03-07	2
2020-03-08	1
2020-03-09	4
2020-03-10	1
2020-03-11	9
2020-03-12	10
2020-03-13	5
2020-03-14	3
2020-03-15	3
2020-03-17	5
2020-03-18	7
2020-03-19	3
2020-03-20	11
2020-03-21	14
2020-03-22	16
2020-03-23	16
2020-03-24	12
2020-03-25	8
2020-03-26	18
2020-03-27	20
2020-03-28	13
2020-03-29	11
2020-03-30	37
2020-03-31	30
2020-04-01	26
2020-04-02	24
2020-04-03	23
2020-04-04	27
2020-04-05	20
2020-04-06	28
2020-04-07	23
2020-04-08	20
2020-04-09	10
2020-04-10	9
2020-04-11	18
2020-04-12	31
2020-04-13	31
2020-04-14	32
2020-04-15	29
2020-04-16	27
2020-04-17	13
2020-04-18	22
2020-04-19	25
2020-04-20	32
2020-04-21	28
2020-04-22	28
2020-04-23	23
2020-04-24	35
2020-04-25	32
2020-04-26	21
2020-04-27	39
2020-04-28	19
2020-04-29	22
2020-04-30	21

Day	Scotland
2020-05-01	19
2020-05-02	4
2020-05-03	14
2020-05-04	6
2020-05-05	14
2020-05-06	18
2020-05-07	3
2020-05-08	11
2020-05-09	3
2020-05-10	9
2020-05-11	4
2020-05-12	5
2020-05-13	15
2020-05-14	33
2020-05-15	12
2020-05-16	6
2020-05-17	9
2020-05-18	13
2020-05-19	17
2020-05-20	9
2020-05-21	10
2020-05-22	3
2020-05-23	4
2020-05-24	2
2020-05-25	2
2020-05-26	1
2020-05-27	2
2020-05-28	4
2020-05-29	1
2020-05-31	1
2020-06-01	3
2020-06-02	11
2020-06-03	1
2020-06-04	7
2020-06-05	1
2020-06-06	2
2020-06-07	1
2020-06-08	1

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
ABERDEEN	Scotland	1	1-10
ABERDEENSHIRE	Scotland	5	1-10
ANGUS	Scotland	38	10-50
CLACKMANNANSHIRE	Scotland	2	1-10
DUMFRIES AND GALLOWAY	Scotland	2	1-10
DUNDEE	Scotland	129	100-150
EAST LOTHIAN	Scotland	55	50-100
EDINBURGH	Scotland	438	400-500
FALKIRK	Scotland	4	1-10
FIFE	Scotland	45	10-50
GLASGOW	Scotland	1	1-10
MIDLOTHIAN	Scotland	134	100-150
MORAY	Scotland	6	1-10
NORTHUMBERLAND	England	3	1-10
PERTHSHIRE AND KINROSS	Scotland	27	10-50
SCOTTISH BORDERS	Scotland	142	100-150
SOUTH LANARKSHIRE	Scotland	4	1-10
WEST LOTHIAN	Scotland	119	100-150