UK lineages summary report

This report gives summaries of UK specific lineages sequenced by EDIN for week 2020-05-22. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-05-17. The analysis (eg time since last sample) is therefore undertaken from this date. 928 sequences in the UK from the sequencing centre EDIN have been included in this analysis. 346 lineages have been recorded, 269 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 5185 and the maximum is 8310

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

316 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 27 that remain: 11 are pending extinction, ie last seen three weeks ago. 7 lineages have gone quiet, ie haven't been seen this week. 4 lineages have reactivated. 5 lineages have been continuously circulating.

The following table contains information about lineages and the number of sequences in each country in the UK for each lineage, in reverse size order.

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.

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"

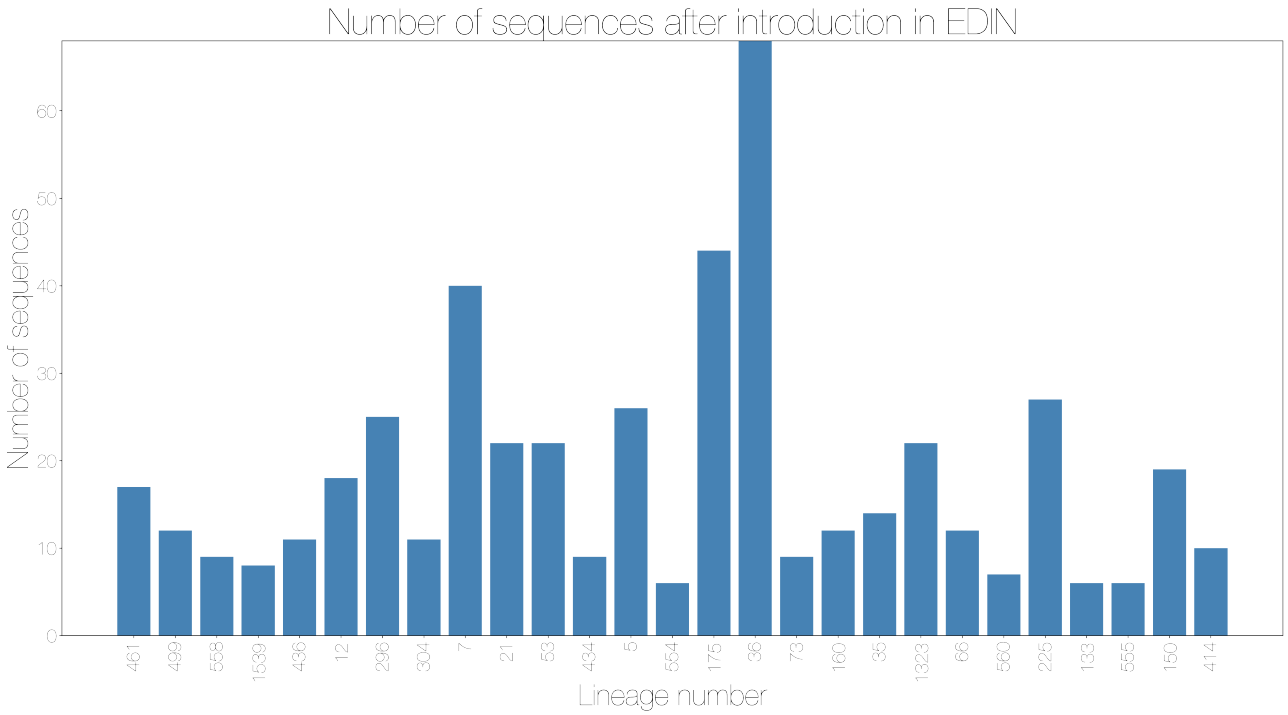
Lineage name	Scotland	Date range	Total sequences	Global lineage	Time since last sample (days)
UK36	68 (100.0%)	Mar-21, May-04	68	B.1	13
UK175	44 (100.0%)	Mar-22, May-04	44	B.1	13
UK7	40 (100.0%)	Mar-20, May-12	40	B.1.p11	5
UK225	27 (100.0%)	Mar-17, Apr-27	27	B.2, B.2.6	20
UK5	26 (100.0%)	Mar-18, May-06	26	B.1.1.1	11
UK296	25 (100.0%)	Apr-08, May-13	25	B.1.5	4
UK21	22 (100.0%)	Mar-18, May-08	22	B.1.40	9
UK1323	22 (100.0%)	Mar-25, May-01	22	B	16

Lineage name	Scotland	Date range	Total sequences	Global lineage	Time since last sample (days)
UK53	22 (100.0%)	Apr-16, May-08	22	B.1.1.4	9
UK150	19 (100.0%)	Mar-21, Apr-22	19	B.1.1.p12	25
UK12	18 (100.0%)	Apr-12, May-13	18	B.1.p11	4
UK461	17 (100.0%)	Apr-18, May-17	17	B.1.5	0
UK43	14 (100.0%)	Mar-26, Apr-18	14	A.5	29
UK156	14 (100.0%)	Mar-18, Apr-18	14	B.1.71	29
UK35	14 (100.0%)	Mar-25, May-01	14	B.1.5, B.1.5.6	16
UK66	12 (100.0%)	Mar-28, Apr-28	12	B.1.1.8	19
UK160	12 (100.0%)	Apr-01, May-02	12	B.1.1	15
UK499	12 (100.0%)	Apr-24, May-15	12	B.1.5	2
UK304	11 (100.0%)	Apr-16, May-12	11	B.1.1.14	5
UK436	11 (100.0%)	Apr-13, May-14	11	B.1.5	3
UK414	10 (100.0%)	Apr-05, Apr-22	10	B.1.5	25
UK434	9 (100.0%)	Apr-20, May-06	9	B.1.5	11
UK73	9 (100.0%)	Mar-28, May-02	9	B.1.p11	15
UK558	9 (100.0%)	Apr-24, May-15	9	B.1.5	2
UK1539	8 (100.0%)	May-09, May-15	8	B.1.5	2
UK560	7 (100.0%)	Apr-15, Apr-27	7	B.1.5	20
UK554	6 (100.0%)	Apr-23, May-05	6	B.1.5	12
UK133	6 (100.0%)	Mar-22, Apr-25	6	B.1	22
UK555	6 (100.0%)	Apr-13, Apr-25	6	B.1.5	22
UK931	6 (100.0%)	Mar-30, Apr-04	6	B.1.1	43

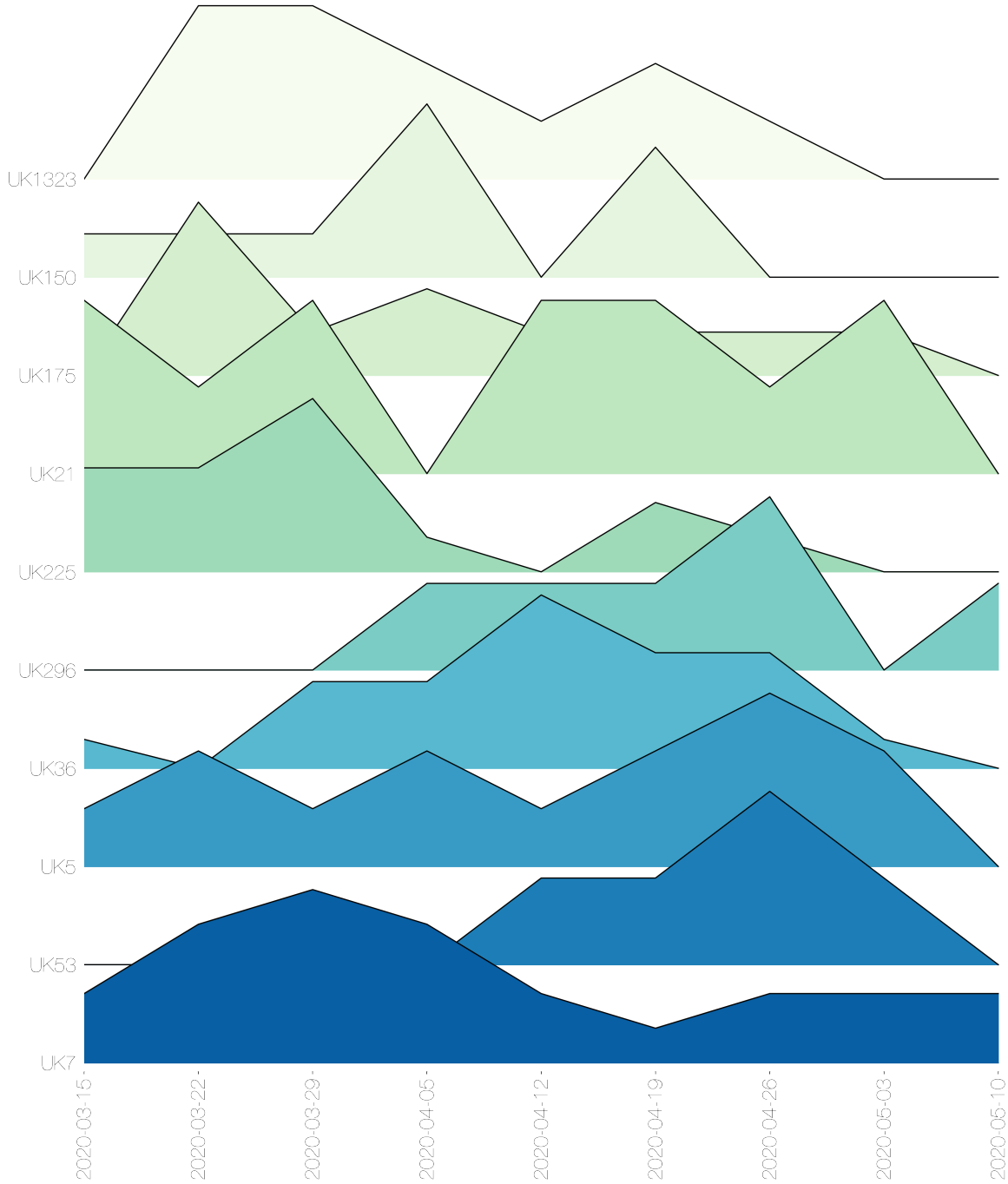
These data is represented in the stacked bar chart below. Note that the number of sequences is likely to be due more to differing sampling efforts in different regions, rather than genuine differences in numbers of

cases.

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

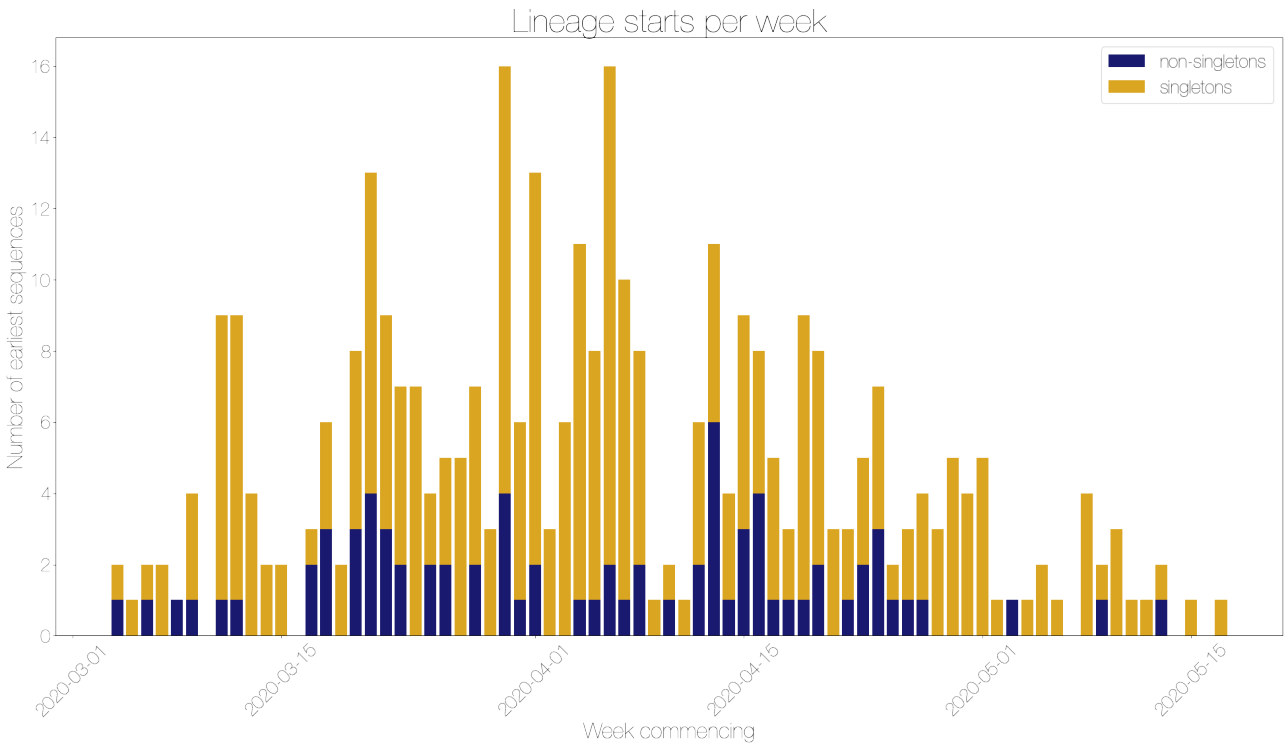


The relative growth and decline of the ten most sampled lineages in terms of number of counties they are present in is shown below. The raw data for the plot is shown below it, with each column representing a lineage, and the number of admin2 regions it is present in in each week.



Week commencing	UK36	UK175	UK7	UK225	UK5	UK296	UK1323	UK53	UK21	UK150
2020-03-15	1	0	2	3	1	0	0	0	2	1
2020-03-22	0	4	4	3	2	0	3	0	1	1
2020-03-29	3	1	5	5	1	0	3	0	2	1
2020-04-05	3	2	4	1	2	1	2	0	0	4
2020-04-12	6	1	2	0	1	1	1	1	2	0
2020-04-19	4	1	1	2	2	1	2	1	2	3
2020-04-26	4	1	2	1	3	2	1	2	1	0
2020-05-03	1	1	2	0	2	0	0	1	2	0
2020-05-10	0	0	2	0	0	1	0	0	0	0

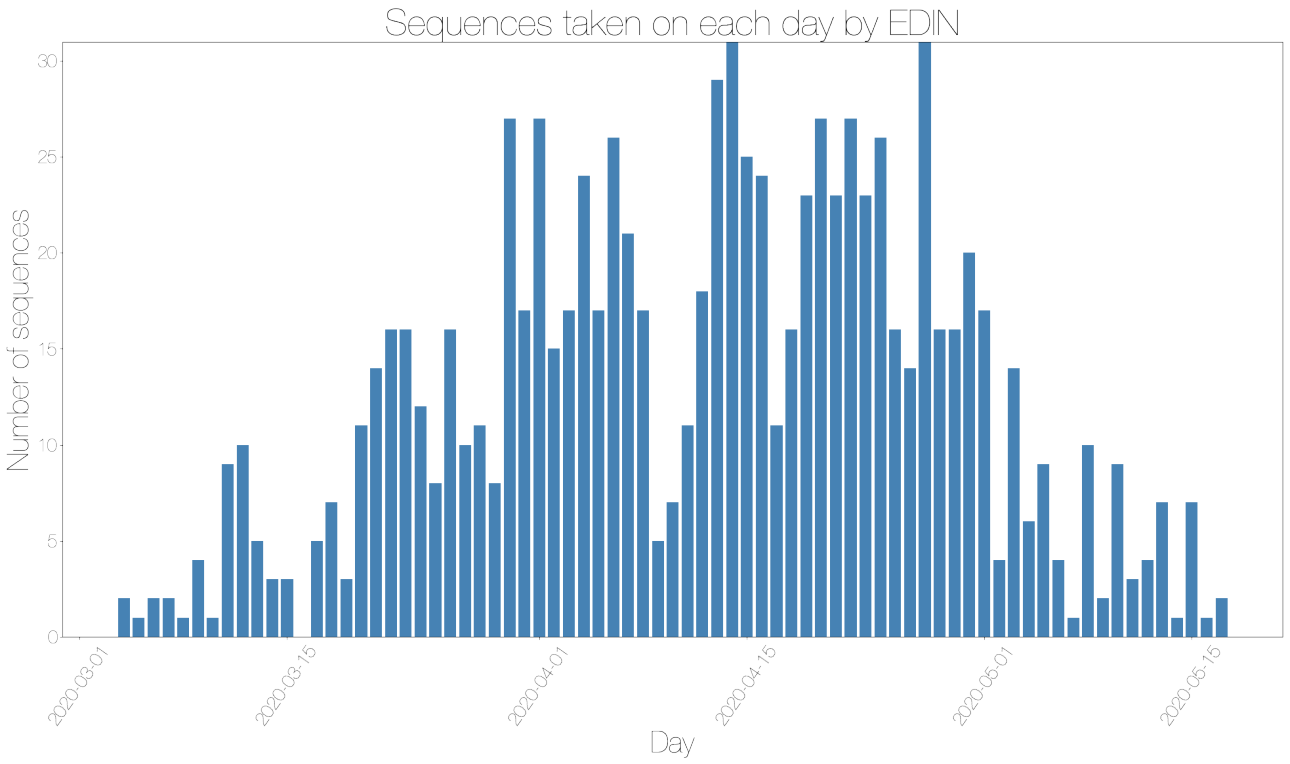
The date of first sequence in the cluster is shown below for every cluster with date information.



Day	Number of singleton starts	Number of non-singleton starts	Total
2020-03-04	1	1	2
2020-03-05	1	0	1
2020-03-06	1	1	2
2020-03-07	2	0	2
2020-03-08	0	1	1
2020-03-09	3	1	4
2020-03-11	8	1	9
2020-03-12	8	1	9
2020-03-13	4	0	4
2020-03-14	2	0	2
2020-03-15	2	0	2
2020-03-17	1	2	3
2020-03-18	3	3	6
2020-03-19	2	0	2
2020-03-20	5	3	8
2020-03-21	9	4	13
2020-03-22	6	3	9
2020-03-23	5	2	7
2020-03-24	7	0	7
2020-03-25	2	2	4
2020-03-26	3	2	5
2020-03-27	5	0	5
2020-03-28	5	2	7
2020-03-29	3	0	3
2020-03-30	12	4	16
2020-03-31	5	1	6
2020-04-01	11	2	13
2020-04-02	3	0	3

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-04-03	6	0	6
2020-04-04	10	1	11
2020-04-05	7	1	8
2020-04-06	14	2	16
2020-04-07	9	1	10
2020-04-08	6	2	8
2020-04-09	1	0	1
2020-04-10	1	1	2
2020-04-11	1	0	1
2020-04-12	4	2	6
2020-04-13	5	6	11
2020-04-14	3	1	4
2020-04-15	6	3	9
2020-04-16	4	4	8
2020-04-17	4	1	5
2020-04-18	2	1	3
2020-04-19	8	1	9
2020-04-20	6	2	8
2020-04-21	3	0	3
2020-04-22	2	1	3
2020-04-23	3	2	5
2020-04-24	4	3	7
2020-04-25	1	1	2
2020-04-26	2	1	3
2020-04-27	3	1	4
2020-04-28	3	0	3
2020-04-29	5	0	5
2020-04-30	4	0	4
2020-05-01	5	0	5
2020-05-02	1	0	1
2020-05-03	0	1	1
2020-05-04	1	0	1
2020-05-05	2	0	2
2020-05-06	1	0	1
2020-05-08	4	0	4
2020-05-09	1	1	2
2020-05-10	3	0	3
2020-05-11	1	0	1
2020-05-12	1	0	1
2020-05-13	1	1	2
2020-05-15	1	0	1
2020-05-17	1	0	1

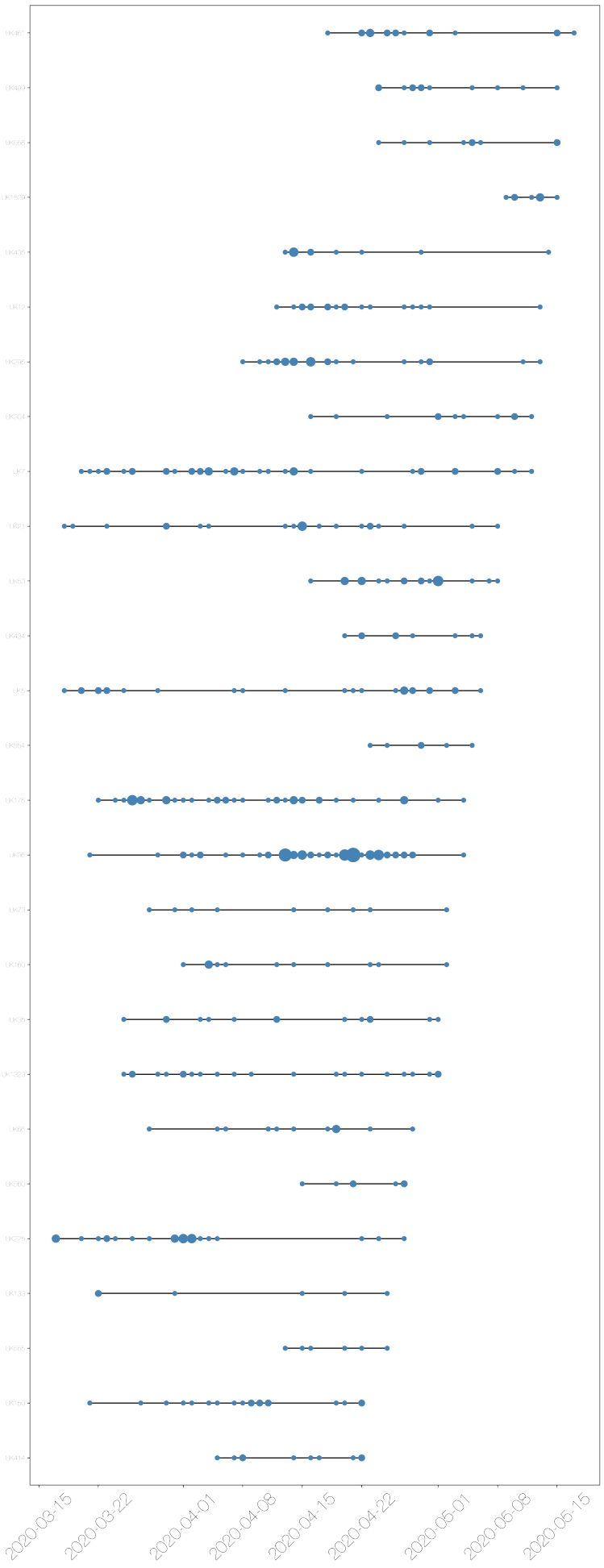
For comparison, here is a plot of the day that every sequence was taken, coloured by country. Note that sequences without dates were not included.



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	16
2020-03-27	10
2020-03-28	11
2020-03-29	8
2020-03-30	27
2020-03-31	17

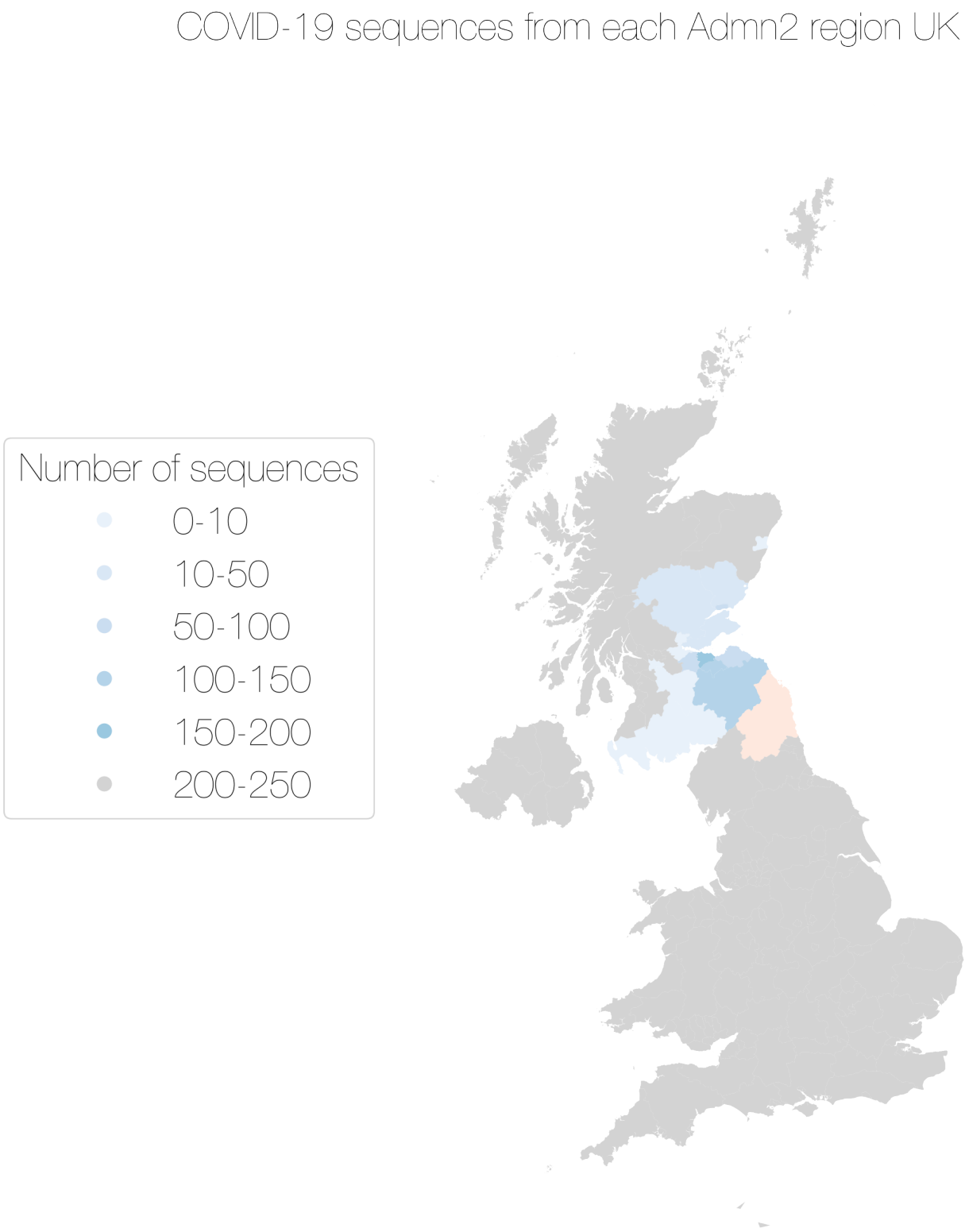
Day	Scotland
2020-04-01	27
2020-04-02	15
2020-04-03	17
2020-04-04	24
2020-04-05	17
2020-04-06	26
2020-04-07	21
2020-04-08	17
2020-04-09	5
2020-04-10	7
2020-04-11	11
2020-04-12	18
2020-04-13	29
2020-04-14	31
2020-04-15	25
2020-04-16	24
2020-04-17	11
2020-04-18	16
2020-04-19	23
2020-04-20	27
2020-04-21	23
2020-04-22	27
2020-04-23	23
2020-04-24	26
2020-04-25	16
2020-04-26	14
2020-04-27	31
2020-04-28	16
2020-04-29	16
2020-04-30	20
2020-05-01	17
2020-05-02	4
2020-05-03	14
2020-05-04	6
2020-05-05	9
2020-05-06	4
2020-05-07	1
2020-05-08	10
2020-05-09	2
2020-05-10	9
2020-05-11	3
2020-05-12	4
2020-05-13	7
2020-05-14	1
2020-05-15	7
2020-05-16	1
2020-05-17	2

These lineages are shown on the timeline below. 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.


2020-03-15
2020-03-22
2020-04-01
2020-04-08
2020-04-15
2020-04-22
2020-05-01
2020-05-08
2020-05-15

The map below 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 44 sequences without enough geographical information to map from this centre.



Admin2	Country	Number of sequences	Sequence group
ABERDEEN	Scotland	1	1-10
ANGUS	Scotland	10	10-50
CLACKMANNANSHIRE	Scotland	2	1-10
DUMFRIES AND GALLOWAY	Scotland	1	1-10
DUNDEE	Scotland	59	50-100
EAST LOTHIAN	Scotland	51	50-100
EDINBURGH	Scotland	388	300-400
FALKIRK	Scotland	4	1-10
FIFE	Scotland	41	10-50
GLASGOW	Scotland	1	1-10
MIDLOTHIAN	Scotland	119	100-150
NORTHUMBERLAND	England	1	1-10
PERTHSHIRE AND KINROSS	Scotland	12	10-50
SCOTTISH BORDERS	Scotland	102	100-150
SOUTH LANARKSHIRE	Scotland	3	1-10
WEST LOTHIAN	Scotland	88	50-100

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