

UK lineages summary report

This report gives summaries of UK specific lineages sequenced by GLAS for week 2020-05-29. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-05-13. The analysis (eg time since last sample) is therefore undertaken from this date. 1197 sequences in the UK from the sequencing centre GLAS 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 6130 and the maximum is 9084

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

280 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 22 that remain: 7 are pending extinction, ie last seen three weeks ago. 5 lineages have gone quiet, ie haven't been seen this week. 5 lineages have reactivated. 5 lineages have been continuously circulating.

The following table contains information about lineages and the number of sequences the dataset, 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

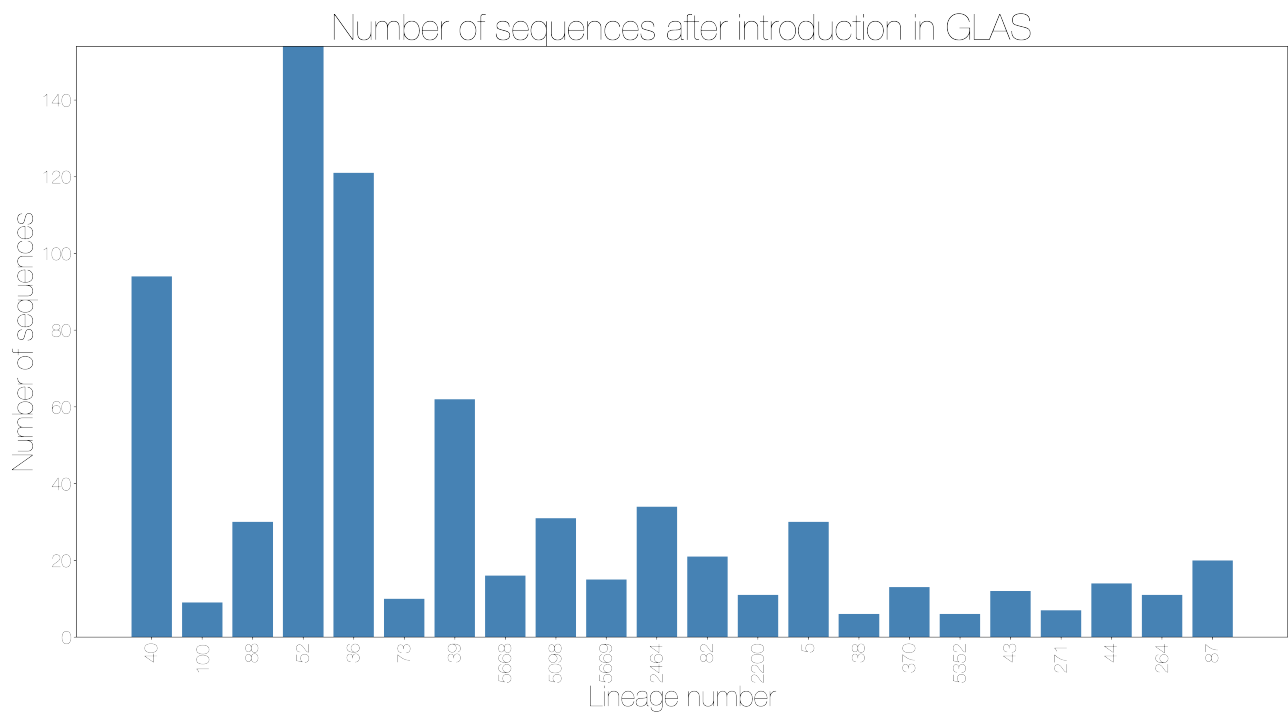
It is written to "summary_files" as "lineage_summary.tsv" for further use, and the full list of lineages is available in the same directory as "all_lineages.csv"

Lineage name	Scotland	Date range	Total sequences	Global lineage	Time since last sample (days)
UK52	154 (100.0%)	Mar-16, May-12	154	B.1.p73	1
UK36	121 (100.0%)	Mar-20, May-11	121	B.1	2
UK40	94 (100.0%)	Mar-13, May-12	94	B.16, B	1
UK39	62 (100.0%)	Mar-12, May-10	62	A.2	3
UK2464	34 (100.0%)	Mar-19, May-06	34	B.1.p11	7
UK5098	31 (100.0%)	Mar-21, May-09	31	B.1.p73	4
UK5	30 (100.0%)	Mar-13, May-03	30	B.1.1.1	10
UK88	30 (100.0%)	Mar-22, May-12	30	B.1	1
UK225	23 (100.0%)	Mar-14, Mar-31	23	B.2	43

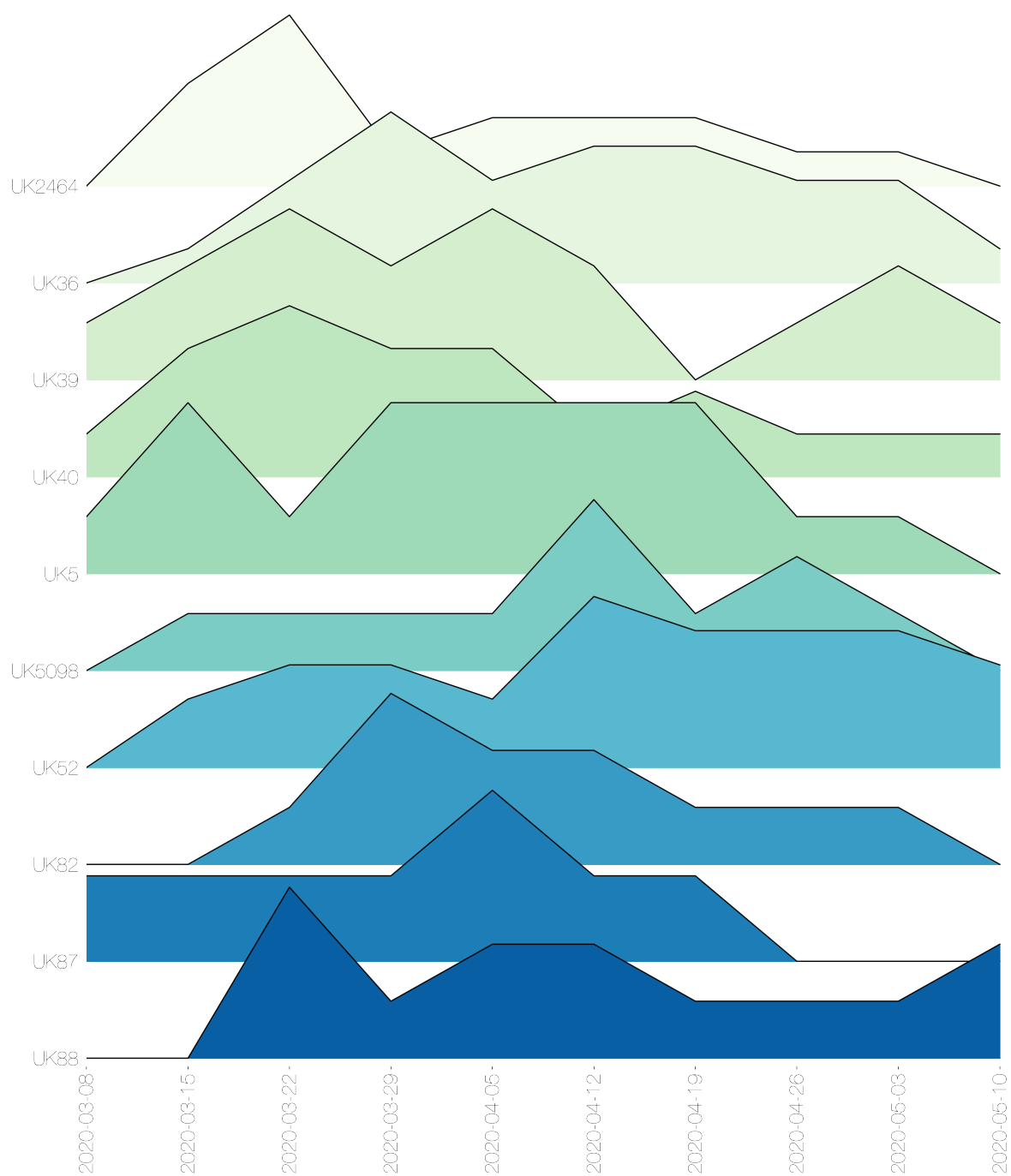
Lineage name	Scotland	Date range	Total sequences	Global lineage	Time since last sample (days)
UK82	21 (100.0%)	Mar-25, May-06	21	B.1.1, B.1.1.p11	7
UK87	20 (100.0%)	Mar-13, Apr-20	20	B.1.70	23
UK14	16 (100.0%)	Mar-14, Apr-08	16	B	35
UK5668	16 (100.0%)	Mar-13, May-09	16	B.2	4
UK502	16 (100.0%)	Mar-06, Mar-20	16	B.1.69	54
UK5669	15 (100.0%)	Mar-24, May-07	15	B.2	6
UK44	14 (100.0%)	Mar-17, Apr-23	14	B	20
UK370	13 (100.0%)	Apr-08, Apr-27	13	B.1.1.10	16
UK43	12 (100.0%)	Mar-23, Apr-26	12	A.5	17
UK261	12 (100.0%)	Mar-15, Apr-08	12	A.3	35
UK264	11 (100.0%)	Mar-29, Apr-22	11	B.1.p11	21
UK2200	11 (100.0%)	Mar-28, May-04	11	B.1.5	9
UK73	10 (100.0%)	Apr-01, May-10	10	B.1.p11	3
UK100	9 (100.0%)	Apr-11, May-12	9	B.1.5	1
UK137	7 (100.0%)	Mar-10, Mar-29	7	B.1.1	45
UK271	7 (100.0%)	Apr-15, Apr-26	7	B.1	17
UK5352	6 (100.0%)	Apr-22, Apr-27	6	B.1.1.14	16
UK198	6 (100.0%)	Mar-18, Apr-15	6	A, B.1.5	28
UK72	6 (100.0%)	Mar-14, Apr-01	6	B.10	42
UK38	6 (100.0%)	Mar-16, May-01	6	B.2.1	12

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.

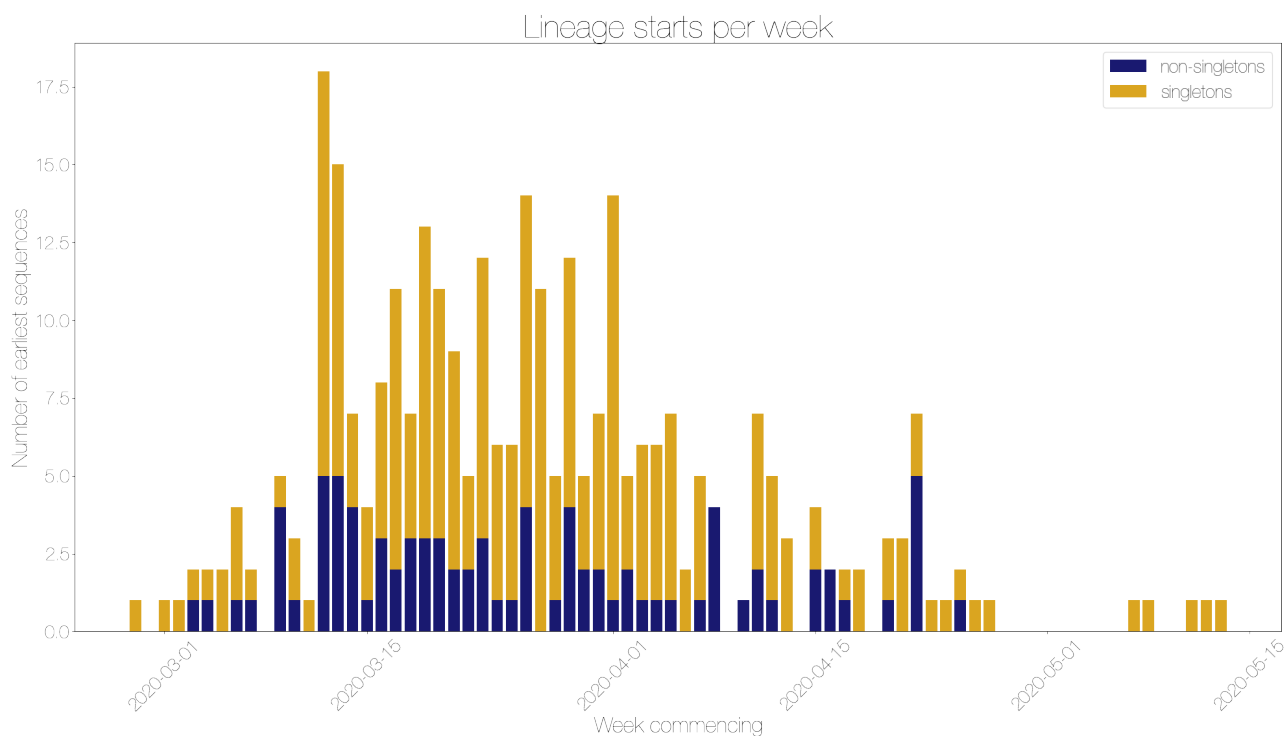
The raw data for this bar chart are in the table above.



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.



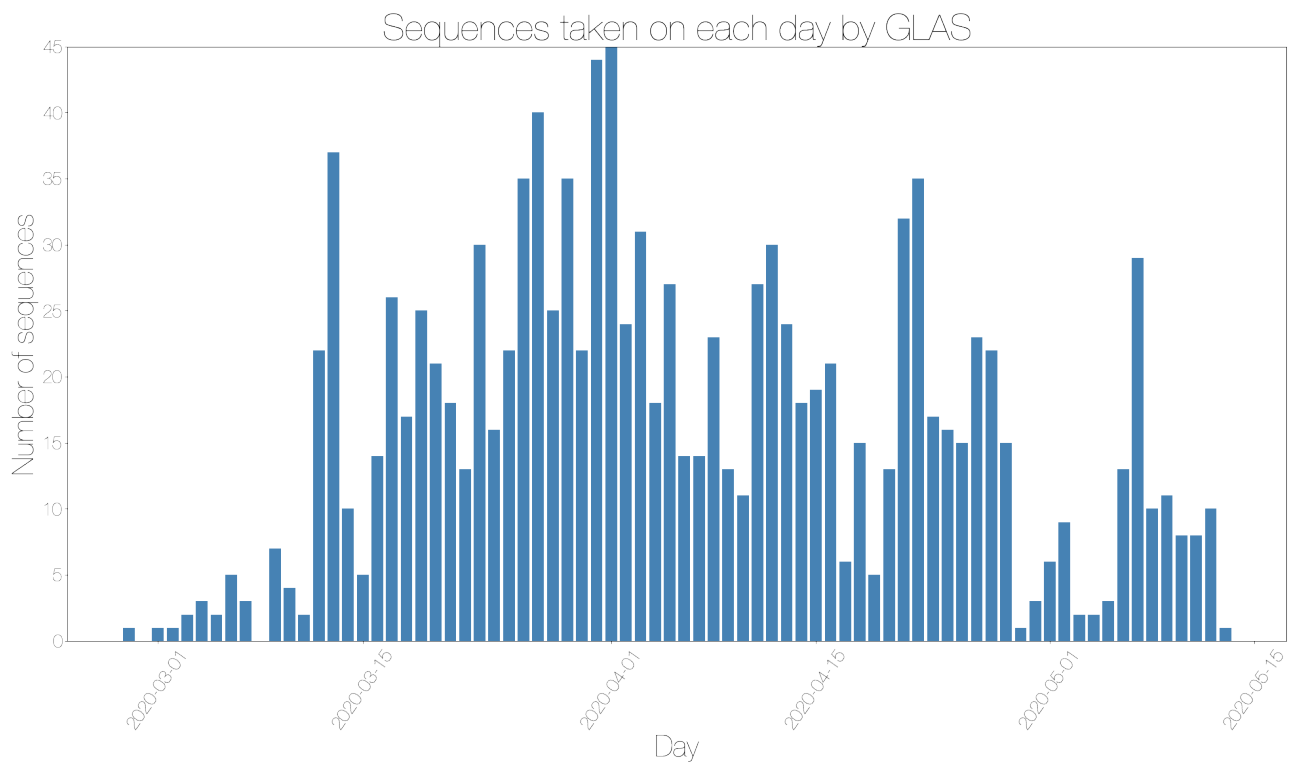
The date of first sequence in the cluster is shown below for every cluster with date information.



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

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-03-27	11	0	11
2020-03-28	4	1	5
2020-03-29	8	4	12
2020-03-30	3	2	5
2020-03-31	5	2	7
2020-04-01	13	1	14
2020-04-02	3	2	5
2020-04-03	5	1	6
2020-04-04	5	1	6
2020-04-05	6	1	7
2020-04-06	2	0	2
2020-04-07	4	1	5
2020-04-08	0	4	4
2020-04-10	0	1	1
2020-04-11	5	2	7
2020-04-12	4	1	5
2020-04-13	3	0	3
2020-04-15	2	2	4
2020-04-16	0	2	2
2020-04-17	1	1	2
2020-04-18	2	0	2
2020-04-20	2	1	3
2020-04-21	3	0	3
2020-04-22	2	5	7
2020-04-23	1	0	1
2020-04-24	1	0	1
2020-04-25	1	1	2
2020-04-26	1	0	1
2020-04-27	1	0	1
2020-05-07	1	0	1
2020-05-08	1	0	1
2020-05-11	1	0	1
2020-05-12	1	0	1
2020-05-13	1	0	1

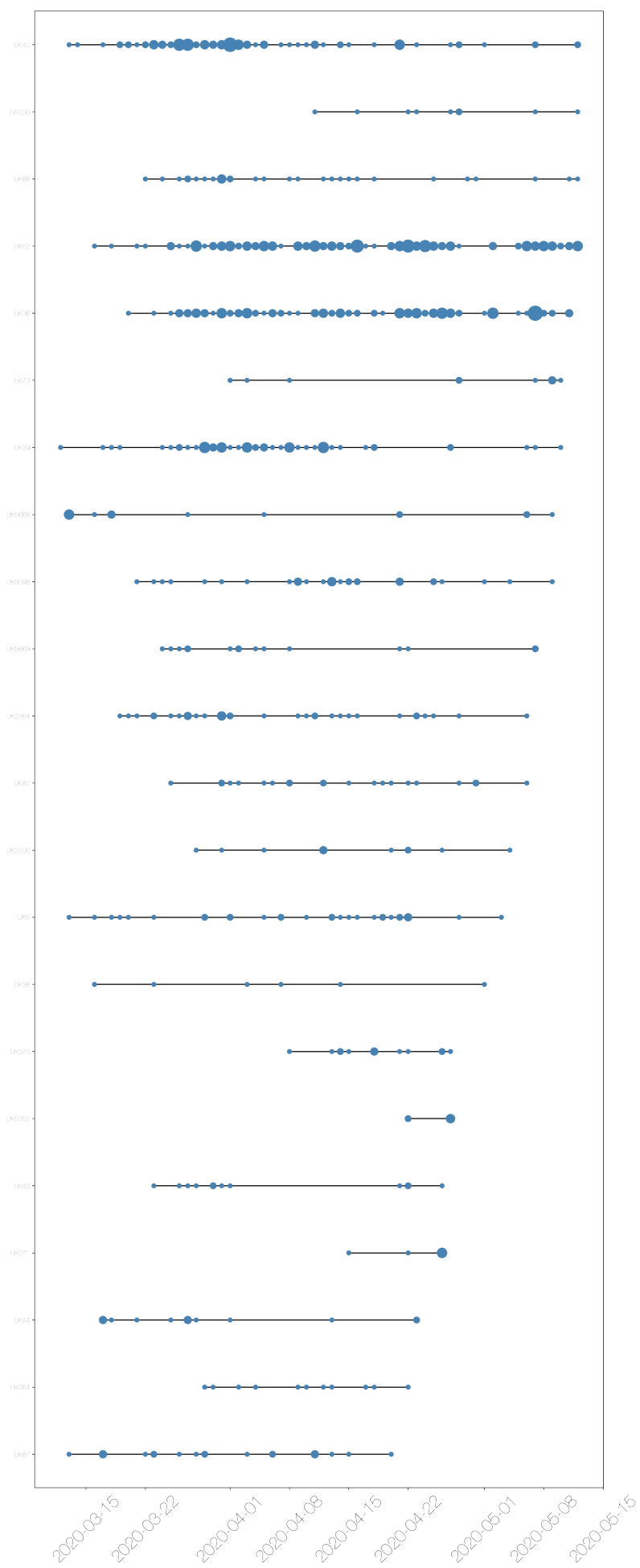
For comparison, here is a plot of the day that every sequence was taken, coloured by country. Note that sequences without dates were not included.



Day	Scotland
2020-02-28	1
2020-03-01	1
2020-03-02	1
2020-03-03	2
2020-03-04	3
2020-03-05	2
2020-03-06	5
2020-03-07	3
2020-03-09	7
2020-03-10	4
2020-03-11	2
2020-03-12	22
2020-03-13	37
2020-03-14	10
2020-03-15	5
2020-03-16	14
2020-03-17	26
2020-03-18	17
2020-03-19	25
2020-03-20	21
2020-03-21	18
2020-03-22	13
2020-03-23	30
2020-03-24	16
2020-03-25	22
2020-03-26	35
2020-03-27	40

Day	Scotland
2020-03-28	25
2020-03-29	35
2020-03-30	22
2020-03-31	44
2020-04-01	45
2020-04-02	24
2020-04-03	31
2020-04-04	18
2020-04-05	27
2020-04-06	14
2020-04-07	14
2020-04-08	23
2020-04-09	13
2020-04-10	11
2020-04-11	27
2020-04-12	30
2020-04-13	24
2020-04-14	18
2020-04-15	19
2020-04-16	21
2020-04-17	6
2020-04-18	15
2020-04-19	5
2020-04-20	13
2020-04-21	32
2020-04-22	35
2020-04-23	17
2020-04-24	16
2020-04-25	15
2020-04-26	23
2020-04-27	22
2020-04-28	15
2020-04-29	1
2020-04-30	3
2020-05-01	6
2020-05-02	9
2020-05-03	2
2020-05-04	2
2020-05-05	3
2020-05-06	13
2020-05-07	29
2020-05-08	10
2020-05-09	11
2020-05-10	8
2020-05-11	8
2020-05-12	10
2020-05-13	1

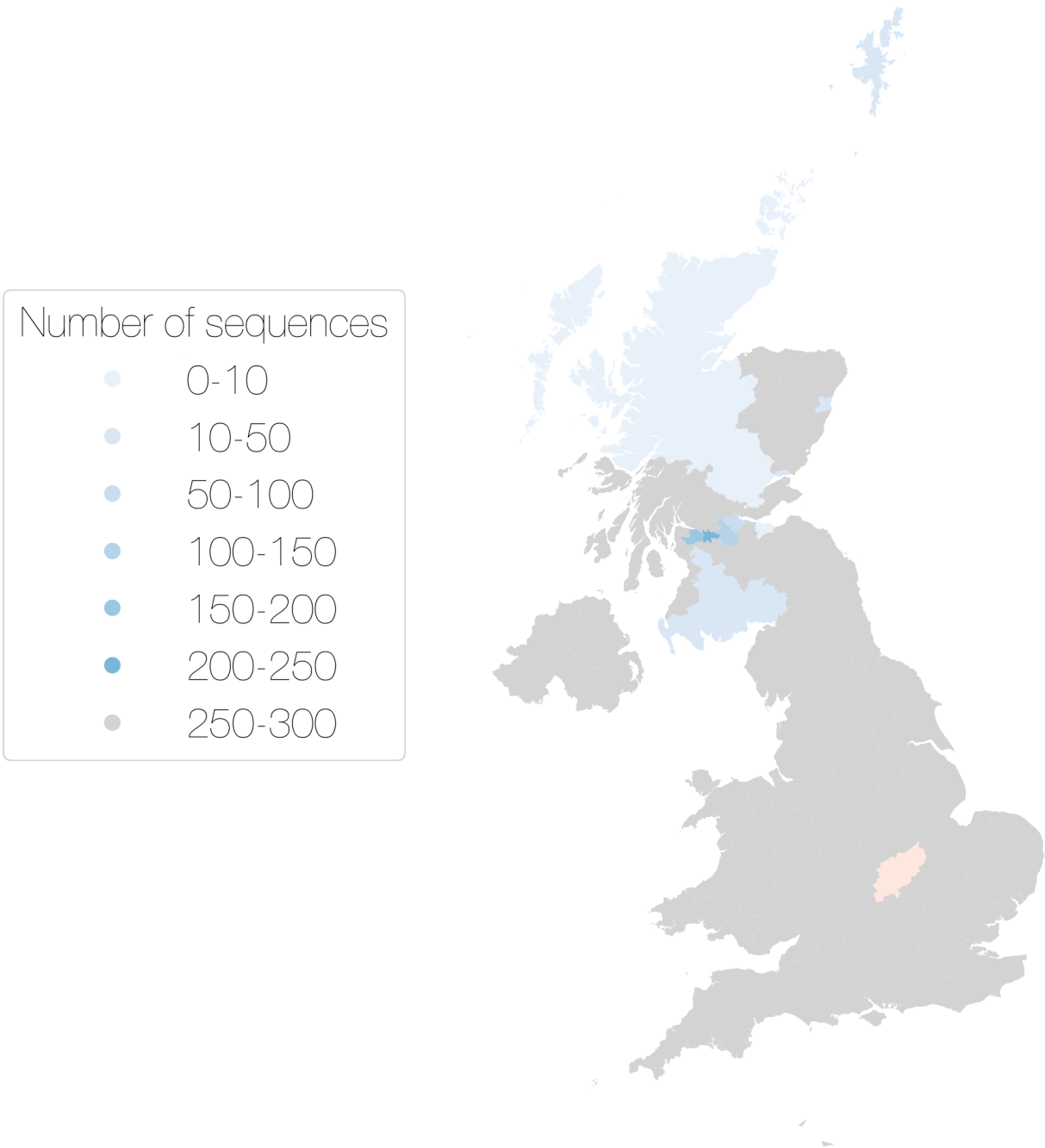
These lineages are shown on the timeline below. Each line represents the length of the cluster, from oldest to most recent sampling date. The dots are sized by the number of sequences taken on that date, and again are colour coded by country. The raw data has been written to a summary file.



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.

All sequences have been assigned clean adm2 data this week.

COVID-19 sequences from each Admn2 region UK



Admin2	Country	Number of sequences	Sequence group
ABERDEEN	Scotland	19	10-50
DUMFRIES AND GALLOWAY	Scotland	37	10-50
DUNDEE	Scotland	11	10-50
EAST AYRSHIRE	Scotland	44	10-50
EDINBURGH	Scotland	9	1-10
EILEAN SIAR	Scotland	2	1-10
FALKIRK	Scotland	66	50-100
GLASGOW	Scotland	669	>500
HIGHLAND	Scotland	9	1-10
NORTH LANARKSHIRE	Scotland	126	100-150
NORTHAMPTONSHIRE	England	1	1-10
ORKNEY ISLANDS	Scotland	1	1-10
PERTHSHIRE AND KINROSS	Scotland	2	1-10
RENFREWSHIRE	Scotland	187	150-200
SHETLAND ISLANDS	Scotland	14	10-50

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