

## UK lineages summary report

This report gives summaries of UK specific lineages sequenced by SHEF for week 2020-05-22. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-05-03. The analysis (eg time since last sample) is therefore undertaken from this date. 1044 sequences in the UK from the sequencing centre SHEF have been included in this analysis. 306 lineages have been recorded, 231 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.

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 23 that remain: 9 are pending extinction, ie last seen three weeks ago. 13 lineages have gone quiet, ie haven't been seen this week. 1 lineage has 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

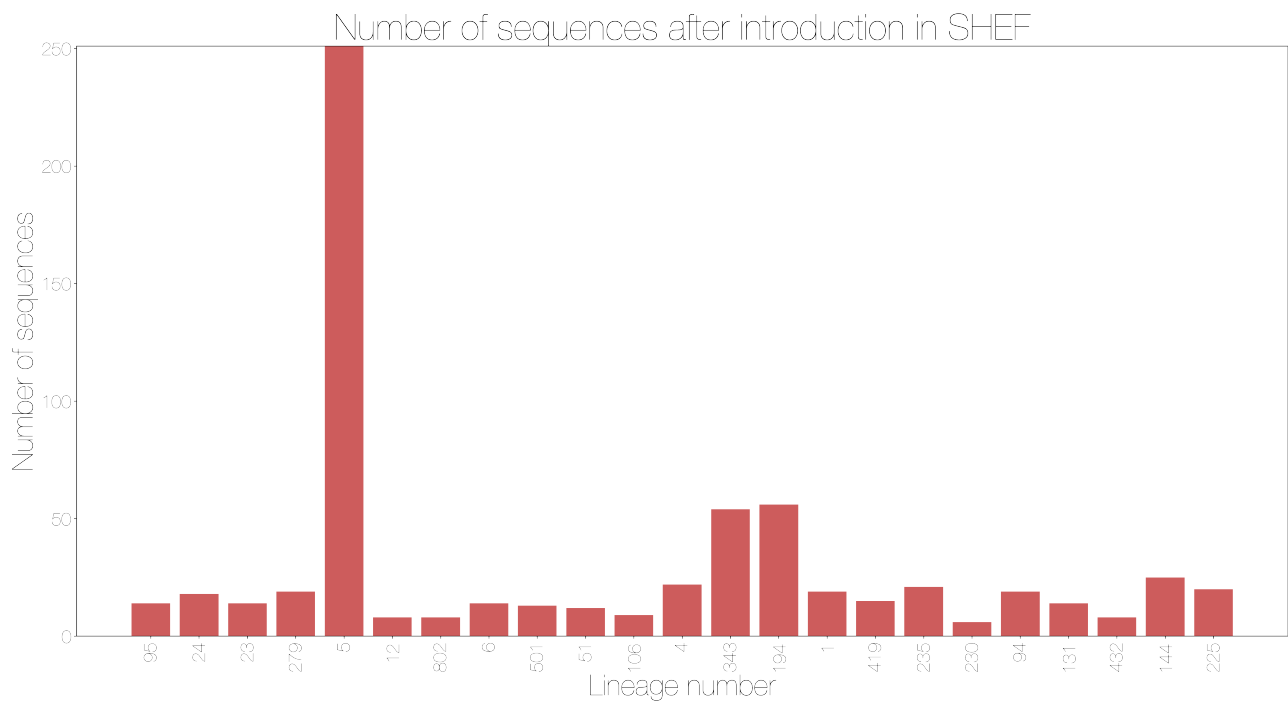
It is written to "summary\_files" as "lineage\_summary.tsv" for further use, and the full list of lineages is available in the same directory as "all\_lineages.csv"

Lineage name	England	Date range	Total sequences	Global lineage	Time since last sample (days)
UK5	251 (100.0%)	Mar-13, Apr-23	251	B.1.1.1	10
UK194	56 (100.0%)	Mar-19, Apr-20	56	B.1.1	13
UK343	54 (100.0%)	Mar-28, Apr-21	54	B.1	12
UK144	25 (100.0%)	Mar-05, Apr-07	25	B.2.1	26
UK4	22 (100.0%)	Mar-07, Apr-21	22	B	12
UK235	21 (100.0%)	Mar-21, Apr-16	21	B.1.1	17
UK225	20 (100.0%)	Mar-06, Apr-06	20	B.2, B.2.2	27
UK94	19 (100.0%)	Mar-21, Apr-10	19	B.2.1	23

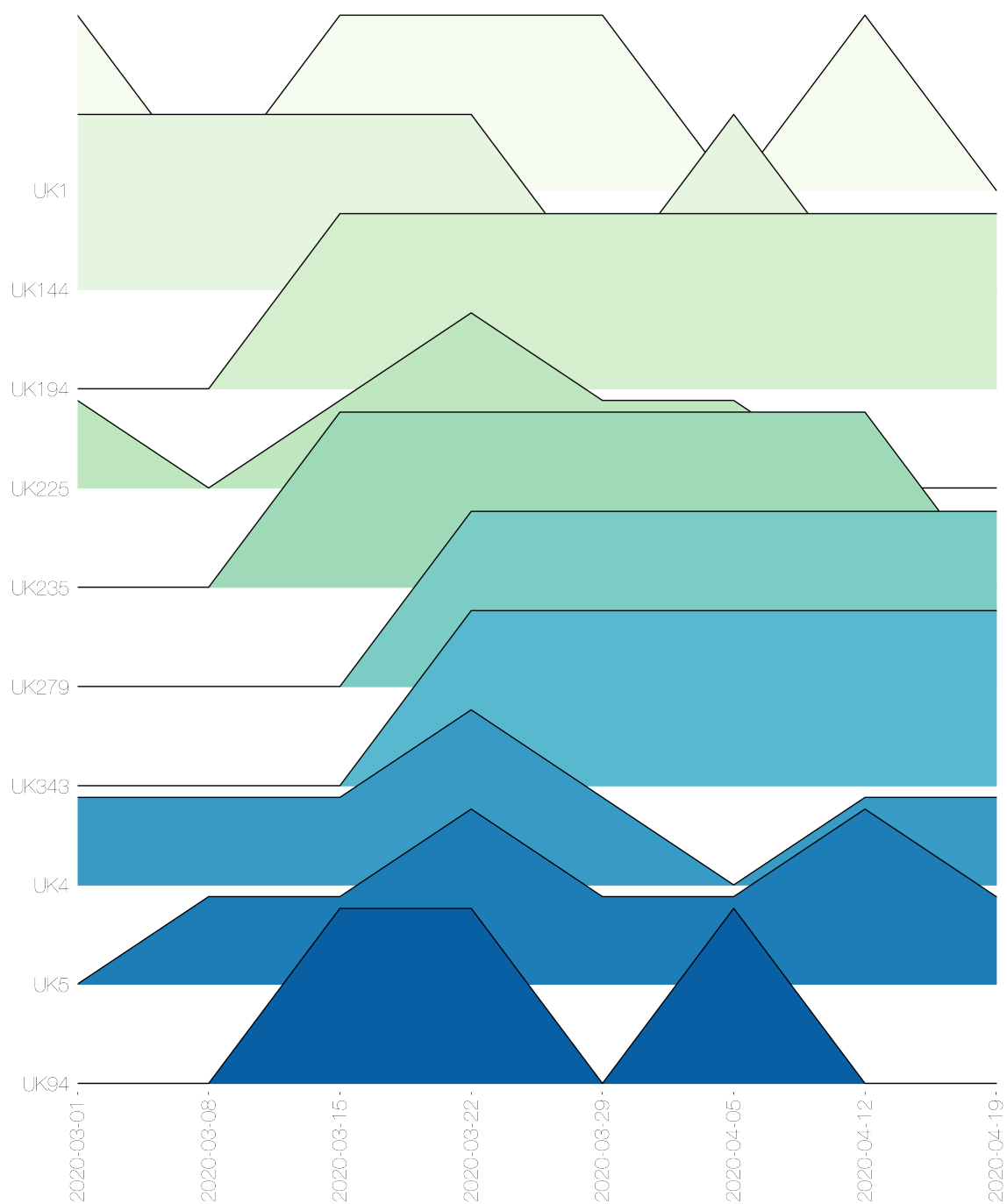
Lineage name	England	Date range	Total sequences	Global lineage	Time since last sample (days)
UK279	19 (100.0%)	Mar-26, Apr-23	19	B.1.1	10
UK1	19 (100.0%)	Mar-03, Apr-16	19	B.1	17
UK24	18 (100.0%)	Mar-19, Apr-23	18	B.1.1	10
UK384	16 (100.0%)	Mar-15, Apr-02	16	B.2.1	31
UK419	15 (100.0%)	Mar-30, Apr-16	15	B.1.1	17
UK131	14 (100.0%)	Mar-20, Apr-10	14	B.15	23
UK95	14 (100.0%)	Mar-21, May-03	14	B.2.1	0
UK23	14 (100.0%)	Mar-23, Apr-23	14	B.9	10
UK6	14 (100.0%)	Mar-22, Apr-22	14	B.1	11
UK501	13 (100.0%)	Apr-03, Apr-22	13	B, B.1	11
UK51	12 (100.0%)	Mar-25, Apr-21	12	B.1.36	12
UK106	9 (100.0%)	Mar-24, Apr-21	9	B.1	12
UK802	8 (100.0%)	Mar-24, Apr-22	8	B.1	11
UK432	8 (100.0%)	Mar-24, Apr-09	8	B.3	24
UK12	8 (100.0%)	Mar-26, Apr-22	8	B.1.p11	11
UK230	6 (100.0%)	Mar-29, Apr-16	6	B.1	17
UK173	6 (100.0%)	Mar-22, Mar-30	6	B	34
UK36	6 (100.0%)	Mar-24, Apr-02	6	B.1	31

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.

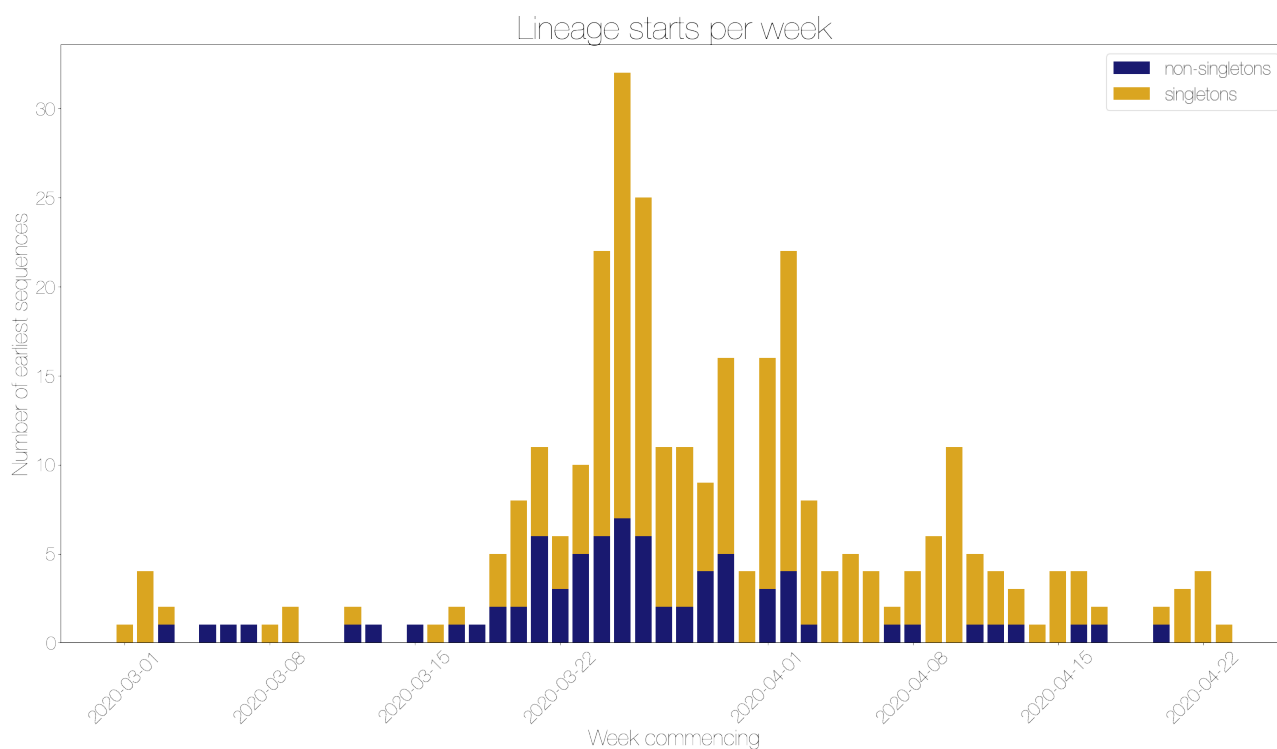


The relative growth and decline of the ten most sampled lineages in terms of number of counties they are present in is shown below. The raw data for the plot is shown below it, with each column representing a lineage, and the number of admin2 regions it is present in in each week.



Week commencing	UK5	UK194	UK343	UK144	UK4	UK235	UK225	UK94	UK1	UK279
2020-03-01	0	0	0	1	1	0	1	0	1	0
2020-03-08	1	0	0	1	1	0	0	0	0	0
2020-03-15	1	1	0	1	1	1	1	1	1	0
2020-03-22	2	1	1	1	2	1	2	1	1	1
2020-03-29	1	1	1	0	1	1	1	0	1	1
2020-04-05	1	1	1	1	0	1	1	1	0	1
2020-04-12	2	1	1	0	1	1	0	0	1	1
2020-04-19	1	1	1	0	1	0	0	0	0	1

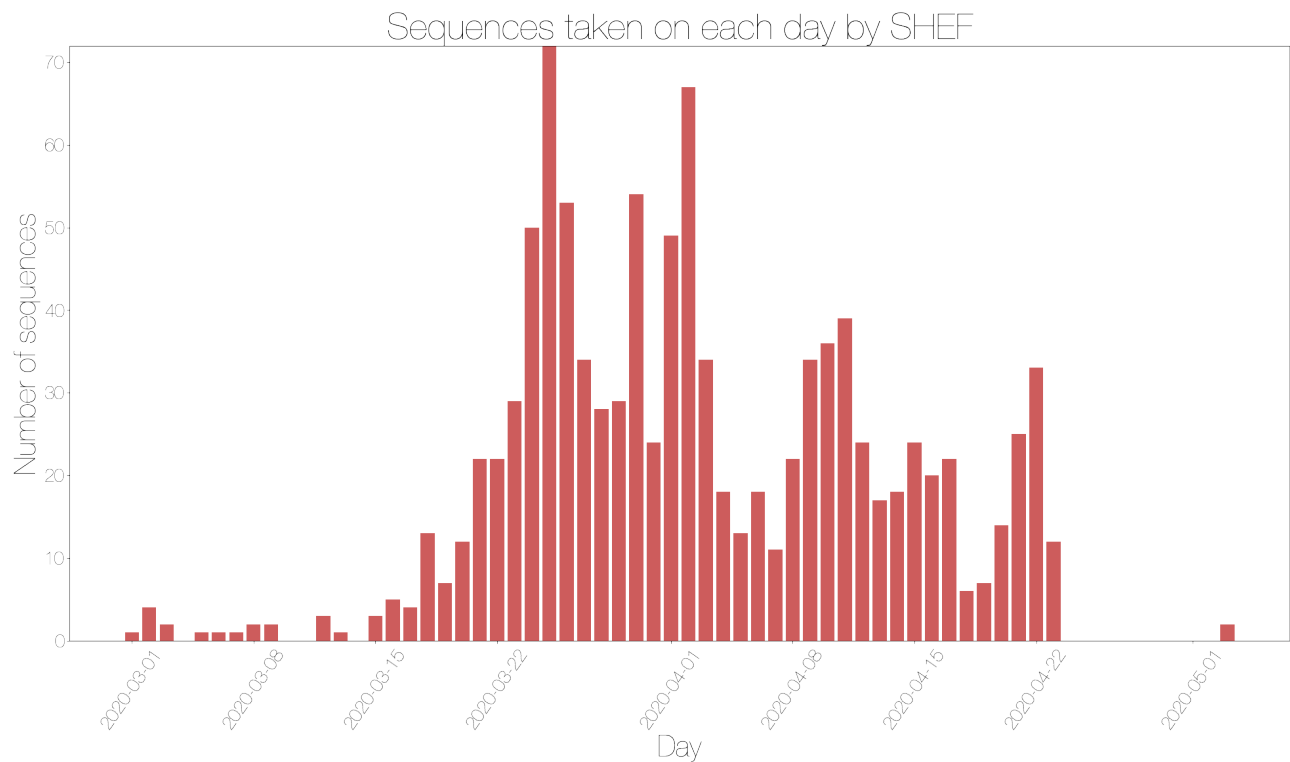
The date of first sequence in the cluster is shown below for every cluster with date information.



Day	Number of singleton starts	Number of non-singleton starts	Total
2020-03-01	1	0	1
2020-03-02	4	0	4
2020-03-03	1	1	2
2020-03-05	0	1	1
2020-03-06	0	1	1
2020-03-07	0	1	1
2020-03-08	1	0	1
2020-03-09	2	0	2
2020-03-12	1	1	2
2020-03-13	0	1	1
2020-03-15	0	1	1
2020-03-16	1	0	1
2020-03-17	1	1	2
2020-03-18	0	1	1
2020-03-19	3	2	5
2020-03-20	6	2	8
2020-03-21	5	6	11
2020-03-22	3	3	6
2020-03-23	5	5	10
2020-03-24	16	6	22
2020-03-25	25	7	32
2020-03-26	19	6	25
2020-03-27	9	2	11
2020-03-28	9	2	11
2020-03-29	5	4	9
2020-03-30	11	5	16
2020-03-31	4	0	4
2020-04-01	13	3	16

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-04-02	18	4	22
2020-04-03	7	1	8
2020-04-04	4	0	4
2020-04-05	5	0	5
2020-04-06	4	0	4
2020-04-07	1	1	2
2020-04-08	3	1	4
2020-04-09	6	0	6
2020-04-10	11	0	11
2020-04-11	4	1	5
2020-04-12	3	1	4
2020-04-13	2	1	3
2020-04-14	1	0	1
2020-04-15	4	0	4
2020-04-16	3	1	4
2020-04-17	1	1	2
2020-04-20	1	1	2
2020-04-21	3	0	3
2020-04-22	4	0	4
2020-04-23	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.



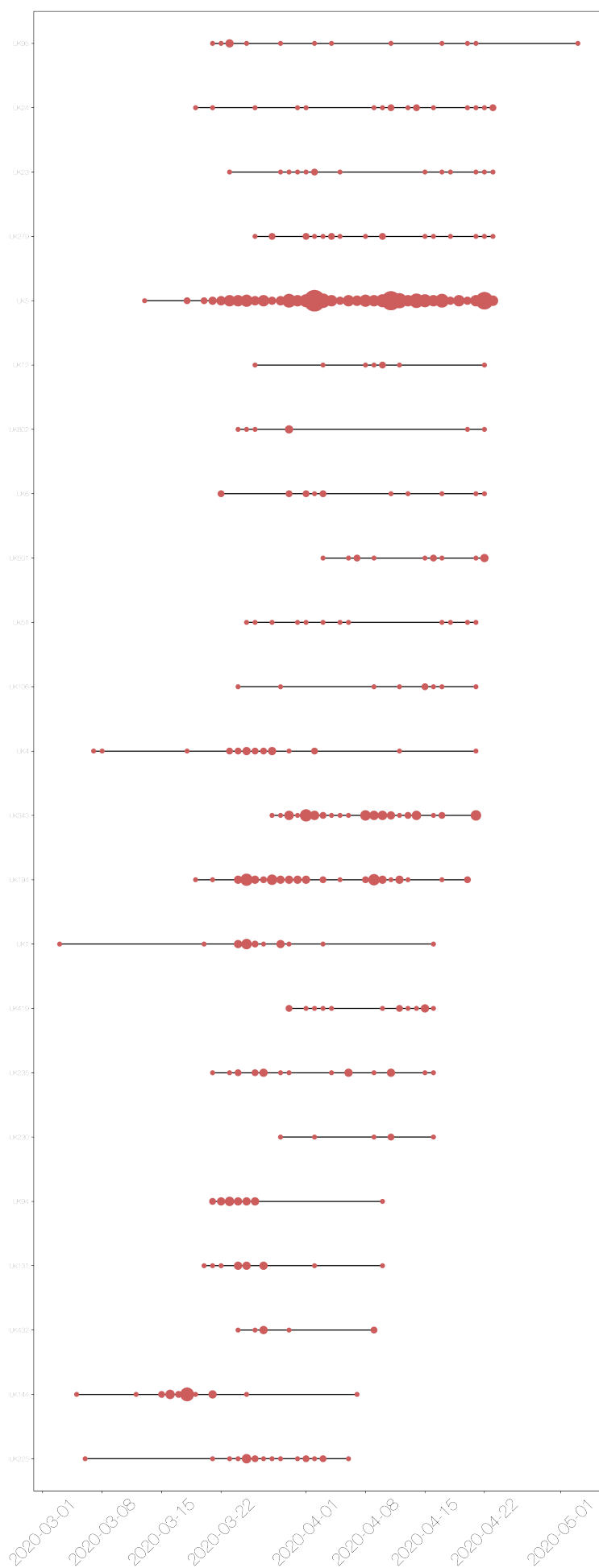
Day	England
2020-03-01	1

Day	England
2020-03-02	4
2020-03-03	2
2020-03-05	1
2020-03-06	1
2020-03-07	1
2020-03-08	2
2020-03-09	2
2020-03-12	3
2020-03-13	1
2020-03-15	3
2020-03-16	5
2020-03-17	4
2020-03-18	13
2020-03-19	7
2020-03-20	12
2020-03-21	22
2020-03-22	22
2020-03-23	29
2020-03-24	50
2020-03-25	72
2020-03-26	53
2020-03-27	34
2020-03-28	28
2020-03-29	29
2020-03-30	54
2020-03-31	24
2020-04-01	49
2020-04-02	67
2020-04-03	34
2020-04-04	18
2020-04-05	13
2020-04-06	18
2020-04-07	11
2020-04-08	22
2020-04-09	34
2020-04-10	36
2020-04-11	39
2020-04-12	24
2020-04-13	17
2020-04-14	18
2020-04-15	24
2020-04-16	20
2020-04-17	22
2020-04-18	6
2020-04-19	7
2020-04-20	14
2020-04-21	25
2020-04-22	33
2020-04-23	12

Day	England
2020-05-03	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.

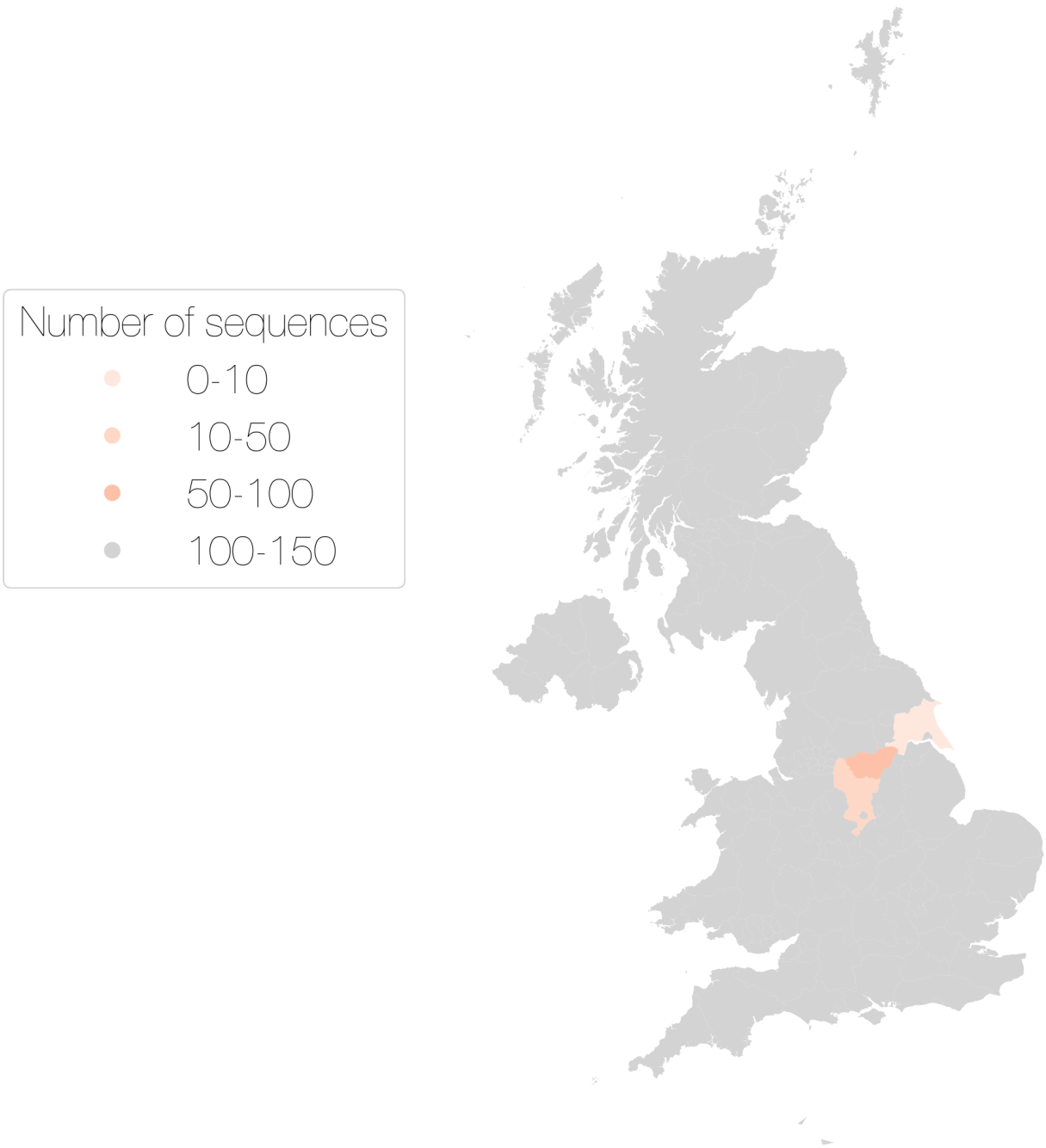




The map below shows the number of sequences sampled in each admin2 region in the UK. The colour scale is the same for all four countries, but with different underlying base colours.

There are 5 sequences without enough geographical information to map from this centre.

COVID-19 sequences from each Admn2 region UK



Admin2	Country	Number of sequences	Sequence group
DERBYSHIRE	England	14	10-50
EAST RIDING OF YORKSHIRE	England	8	1-10
SOUTH YORKSHIRE	England	1017	>500

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