UK lineages summary report

This report gives summaries of UK specific lineages sequenced by LIVE for week 2020-05-29. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-05-15. The analysis (eg time since last sample) is therefore undertaken from this date. 545 sequences in the UK from the sequencing centre LIVE 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.

164 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 17 that remain: 11 are pending extinction, ie last seen three weeks ago. 4 lineages have gone quiet, ie haven't been seen this week. 2 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 lienages 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	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK5	71 (100.0%)	Mar-21, May-07	71	B.1.1.1	8	0.0839
UK36	30 (100.0%)	Mar-29, May-12	30	B.1	3	0.5057
UK293	20 (100.0%)	Mar-24, Apr-28	20	B.1	17	0.1084
UK307	19 (100.0%)	Mar-28, May-04	19	B.1.1	11	0.1869
UK632	17 (100.0%)	Mar-23, Apr-25	17	B.1.1	20	0.1031
UK193	17 (100.0%)	Apr-07, May-01	17	B.1.1	14	0.1071

Lineage name	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK236	14 (100.0%)	Mar-27, Apr-22	14	B.1.1	23	0.087
UK249	13 (100.0%)	Apr-01, Apr-25	13	B.1.1	20	0.1
UK92	13 (100.0%)	Mar-23, Apr-22	13	B.1.1	23	0.1087
UK274	13 (100.0%)	Mar-23, Apr-15	13	B.3	30	0.0639

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.

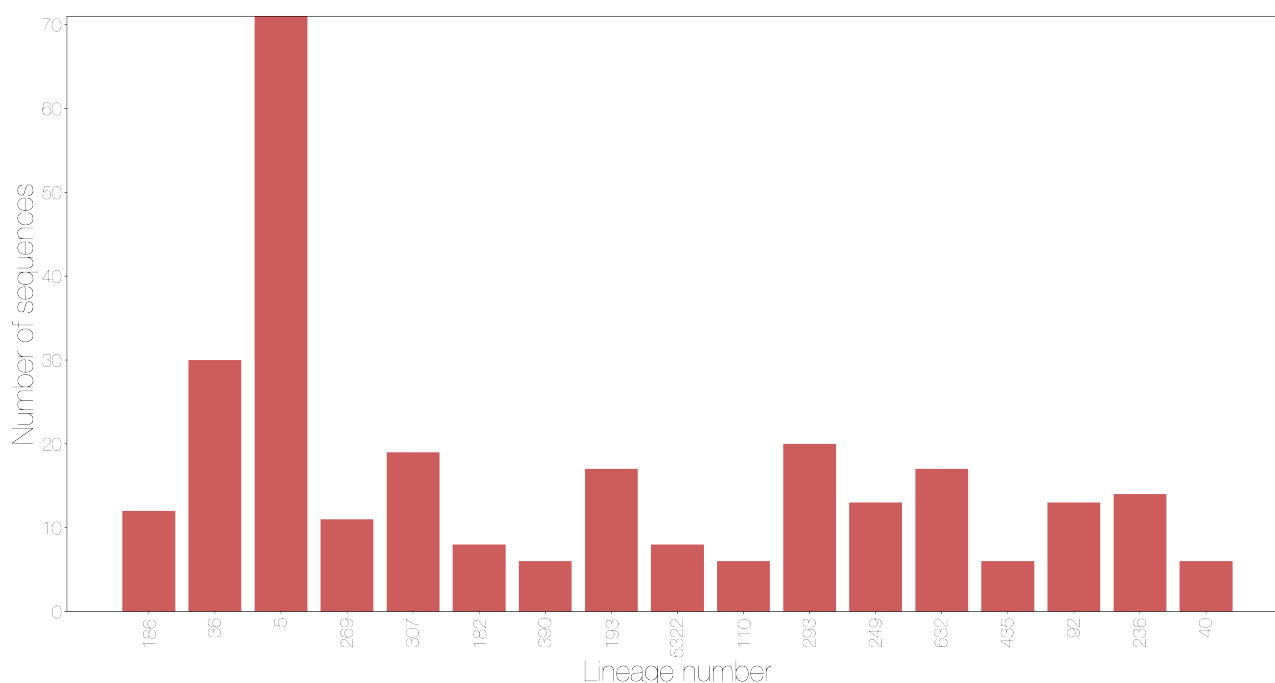The raw data for this bar chart are in the table above.

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.

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.

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

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.
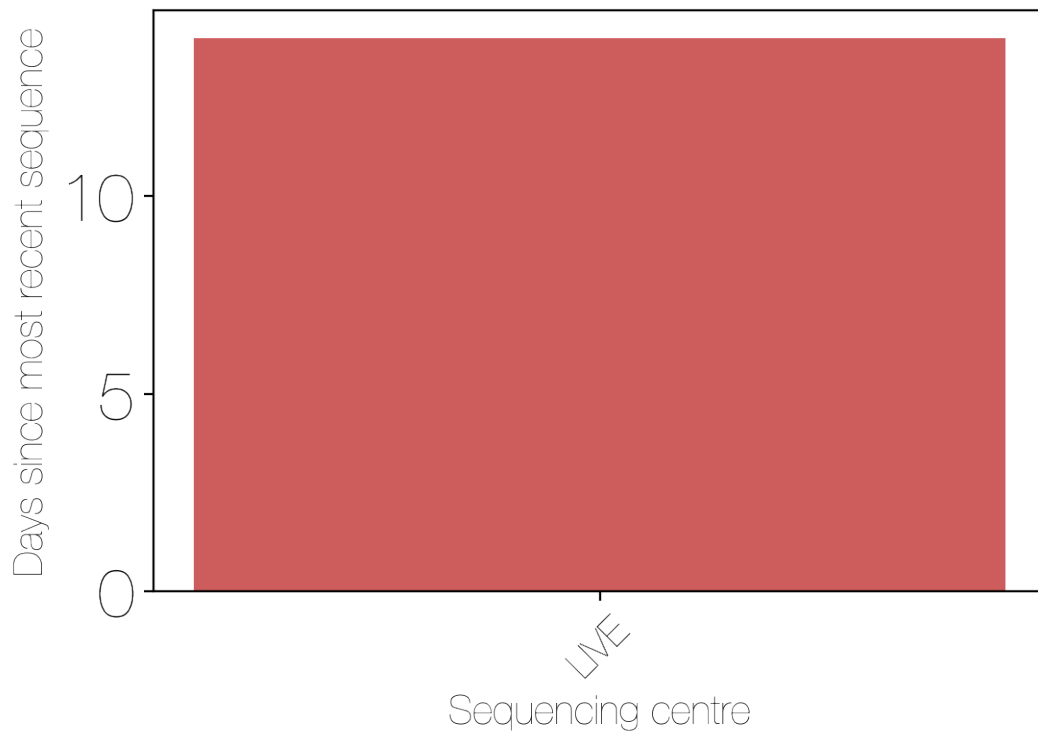
Figure 2: Lag since the most recent sequence from each sequencing centre to most current date

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 no sequences with geographical information for this sequencing centre

Other results modules for UK lineage analysis can be added in here if required.
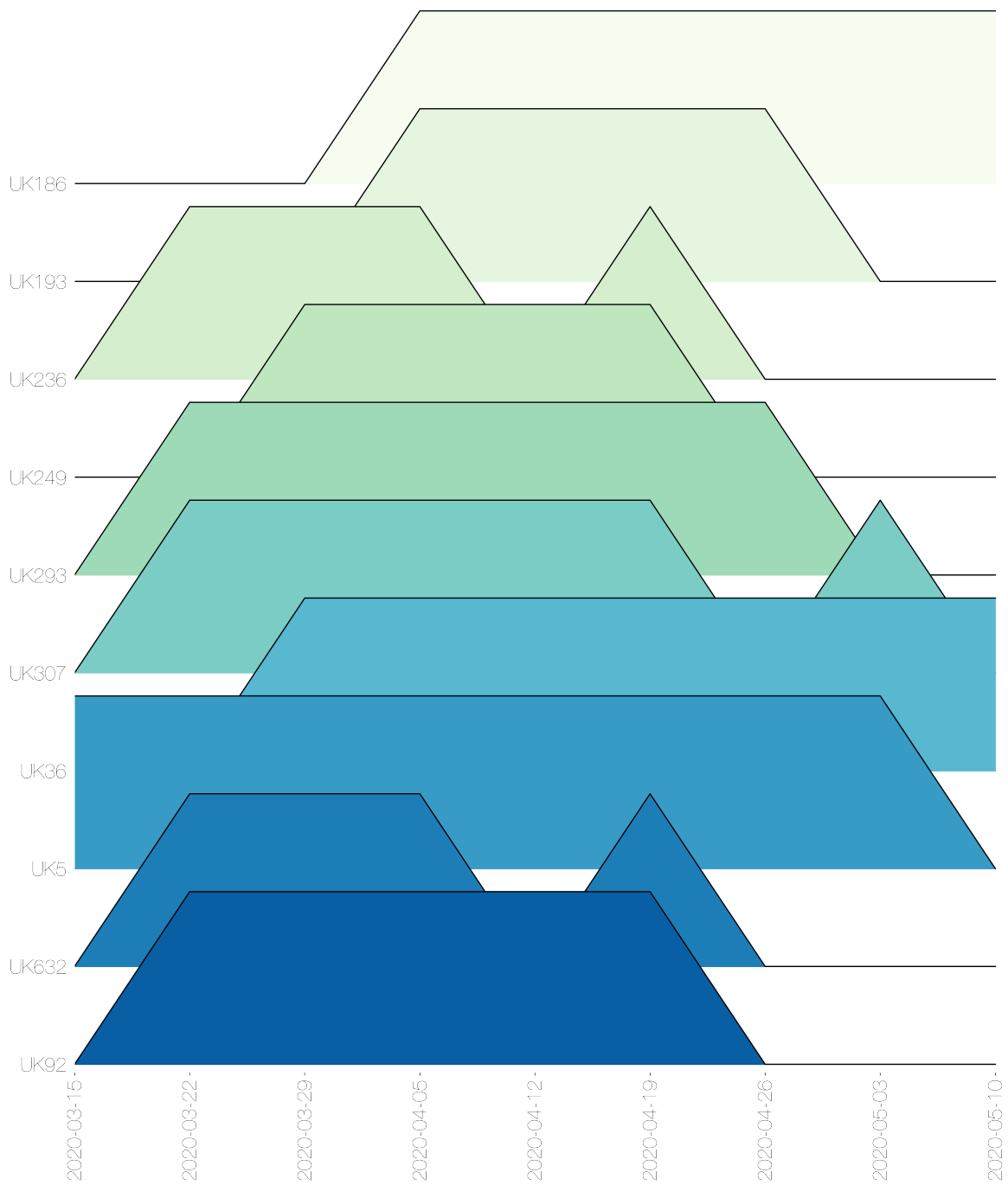
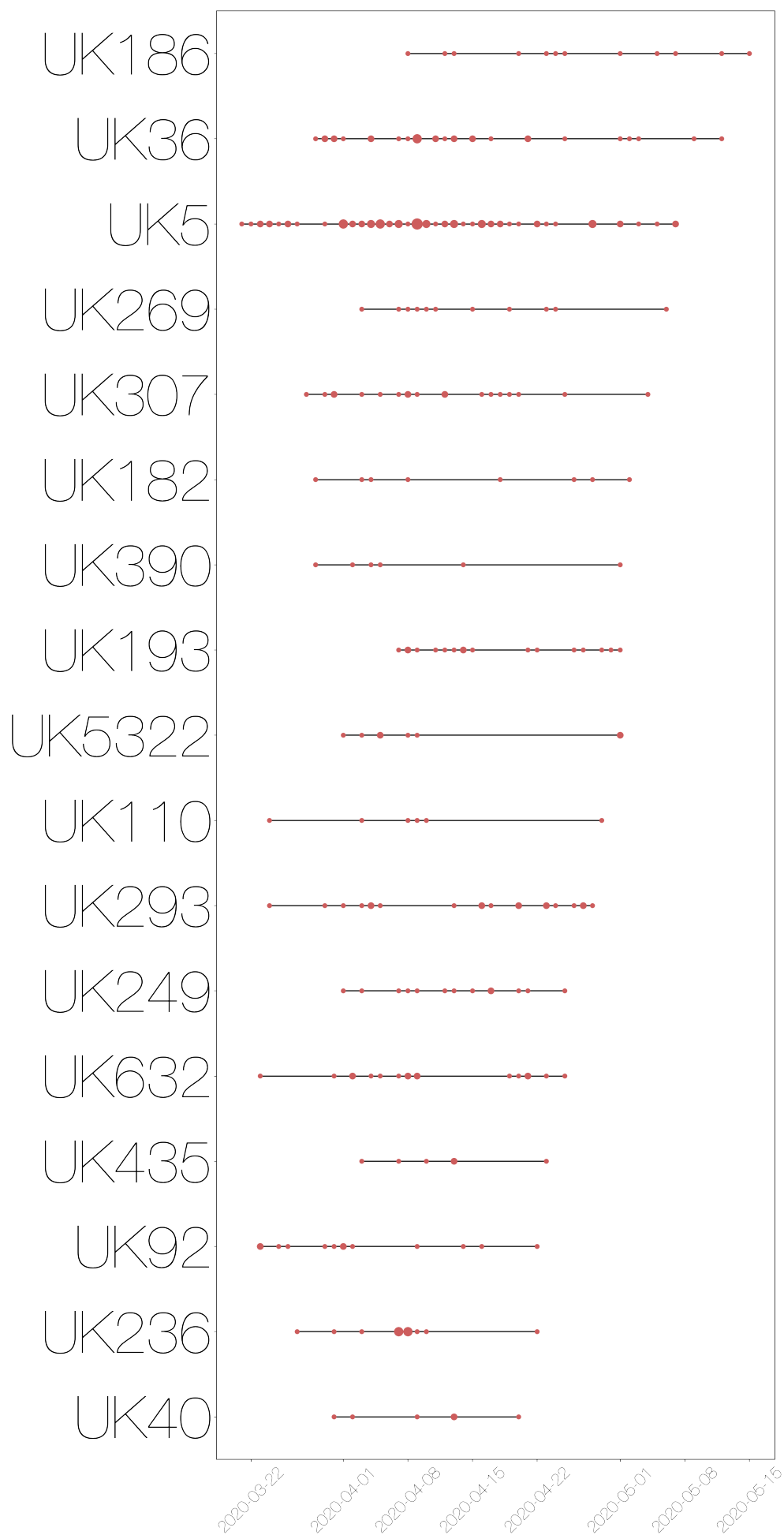Figure 3: Lineages by number of adm2 regions present by epiweek

Figure 4: Timeline of lineages, sized by number of sequences from each country.
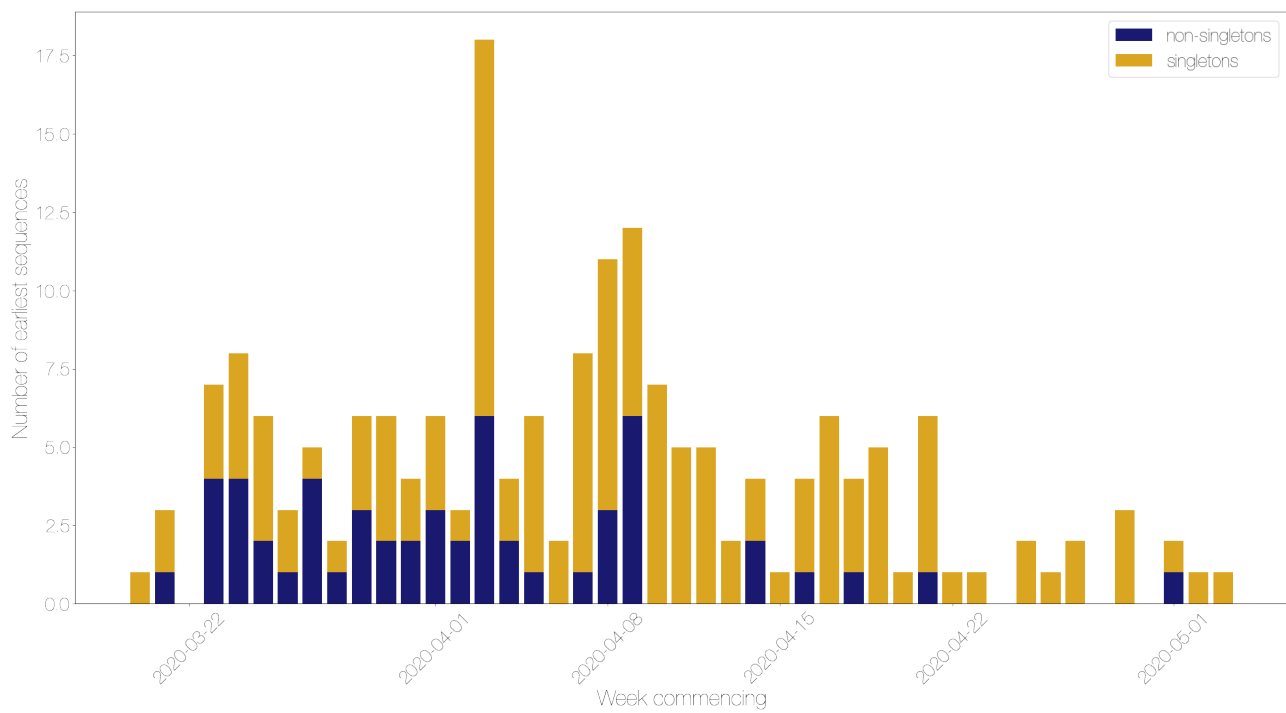
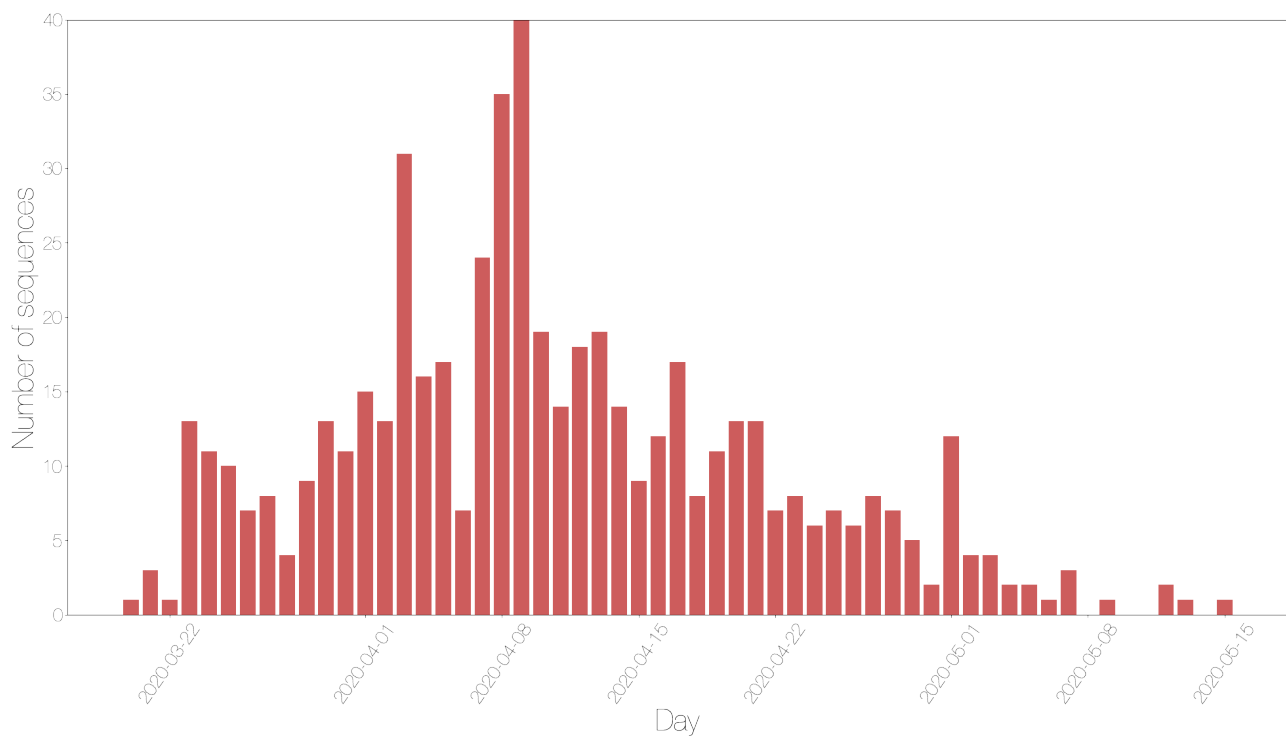Figure 5: Lineage starts per week, split by singletons and non-singletons

Figure 6: Sequences taken on each day by country

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	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK5	71 (100.0%)	Mar-21, May-07	71	B.1.1.1	8	0.0839
UK36	30 (100.0%)	Mar-29, May-12	30	B.1	3	0.5057
UK293	20 (100.0%)	Mar-24, Apr-28	20	B.1	17	0.1084
UK307	19 (100.0%)	Mar-28, May-04	19	B.1.1	11	0.1869
UK632	17 (100.0%)	Mar-23, Apr-25	17	B.1.1	20	0.1031
UK193	17 (100.0%)	Apr-07, May-01	17	B.1.1	14	0.1071
UK236	14 (100.0%)	Mar-27, Apr-22	14	B.1.1	23	0.087
UK249	13 (100.0%)	Apr-01, Apr-25	13	B.1.1	20	0.1
UK92	13 (100.0%)	Mar-23, Apr-22	13	B.1.1	23	0.1087
UK274	13 (100.0%)	Mar-23, Apr-15	13	B.3	30	0.0639
UK186	12 (100.0%)	Apr-08, May-15	12	B	0	active today
UK269	11 (100.0%)	Apr-03, May-06	11	B.1.1	9	0.3667
UK5322	8 (100.0%)	Apr-01, May-01	8	B.1.1	14	0.3061
UK182	8 (100.0%)	Mar-29, May-02	8	B.1.1	13	0.3736
UK5177	7 (100.0%)	Mar-27, Apr-11	7	B.1.1.7	34	0.0735
UK131	7 (100.0%)	Mar-24, Apr-14	7	B.15	31	0.1129
UK542	6 (100.0%)	Apr-01, Apr-14	6	B.1	31	0.0839
UK110	6 (100.0%)	Mar-24, Apr-29	6	B.1	16	0.45
UK390	6 (100.0%)	Mar-29, May-01	6	B.1.5	14	0.4714
UK435	6 (100.0%)	Apr-03, Apr-23	6	B.1.5	22	0.1818

Lineage name	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK40	6 (100.0%)	Mar-31, Apr-20	6	B.16	25	0.16

Table S2 Raw data for figure three showing the number of admin2 regions a lineage is present in over time

Week commencing	UK5	UK36	UK293	UK307	UK632	UK193	UK236	UK92	UK249	UK186
2020-03-15	1	0	0	0	0	0	0	0	0	0
2020-03-22	1	0	1	1	1	0	1	1	0	0
2020-03-29	1	1	1	1	1	0	1	1	1	0
2020-04-05	1	1	1	1	1	1	1	1	1	1
2020-04-12	1	1	1	1	0	1	0	1	1	1
2020-04-19	1	1	1	1	1	1	1	1	1	1
2020-04-26	1	1	1	0	0	1	0	0	0	1
2020-05-03	1	1	0	1	0	0	0	0	0	1
2020-05-10	0	1	0	0	0	0	0	0	0	1

Table S3 is not appropriate for this report and so has been omitted.

Table S4 Raw data for figure six showing when lineages started per day, divided by singletons and non-singletons

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-03-20	1	0	1
2020-03-21	2	1	3
2020-03-23	3	4	7
2020-03-24	4	4	8
2020-03-25	4	2	6
2020-03-26	2	1	3
2020-03-27	1	4	5
2020-03-28	1	1	2
2020-03-29	3	3	6
2020-03-30	4	2	6
2020-03-31	2	2	4
2020-04-01	3	3	6
2020-04-02	1	2	3
2020-04-03	12	6	18
2020-04-04	2	2	4
2020-04-05	5	1	6
2020-04-06	2	0	2
2020-04-07	7	1	8
2020-04-08	8	3	11
2020-04-09	6	6	12
2020-04-10	7	0	7
2020-04-11	5	0	5
2020-04-12	5	0	5
2020-04-13	2	0	2
2020-04-14	2	2	4
2020-04-15	1	0	1
2020-04-16	3	1	4
2020-04-17	6	0	6
2020-04-18	3	1	4
2020-04-19	5	0	5
2020-04-20	1	0	1
2020-04-21	5	1	6
2020-04-22	1	0	1
2020-04-23	1	0	1
2020-04-25	2	0	2
2020-04-26	1	0	1
2020-04-27	2	0	2
2020-04-29	3	0	3
2020-05-01	1	1	2
2020-05-02	1	0	1
2020-05-03	1	0	1

Table S5 Raw data for figure seven showing the number of sequences taken over time.

Day	England
2020-03-20	1
2020-03-21	3
2020-03-22	1
2020-03-23	13
2020-03-24	11
2020-03-25	10
2020-03-26	7
2020-03-27	8
2020-03-28	4
2020-03-29	9
2020-03-30	13
2020-03-31	11
2020-04-01	15
2020-04-02	13
2020-04-03	31
2020-04-04	16
2020-04-05	17
2020-04-06	7
2020-04-07	24
2020-04-08	35
2020-04-09	40
2020-04-10	19
2020-04-11	14
2020-04-12	18
2020-04-13	19
2020-04-14	14
2020-04-15	9
2020-04-16	12
2020-04-17	17
2020-04-18	8
2020-04-19	11
2020-04-20	13
2020-04-21	13
2020-04-22	7
2020-04-23	8
2020-04-24	6
2020-04-25	7
2020-04-26	6
2020-04-27	8
2020-04-28	7
2020-04-29	5
2020-04-30	2
2020-05-01	12
2020-05-02	4
2020-05-03	4
2020-05-04	2
2020-05-05	2

Day	England
2020-05-06	1
2020-05-07	3
2020-05-09	1
2020-05-12	2
2020-05-13	1
2020-05-15	1

Table S6 Raw data for the map with the number of sequences assigned to each admin2 region.