UK lineages summary report

This report gives summaries of UK specific lineages sequenced by NORW for week 2020-05-29. 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. 677 sequences in the UK from the sequencing centre NORW 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.

116 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 24 that remain: 3 are pending extinction, ie last seen three weeks ago. 7 lineages have gone quiet, ie haven't been seen this week. 1 has reactivated. 13 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
UK200	62 (100.0%)	Apr-08, May-14	62	B.1.p11	3	0.1967
UK6	62 (100.0%)	Apr-08, May-13	62	B.1	4	0.1434
UK233	57 (100.0%)	Apr-08, May-13	57	B.1.1	4	0.1562
UK199	50 (100.0%)	Apr-08, May-17	50	B.1.5.5	0	active today
UK204	50 (100.0%)	Apr-08, May-12	50	B.1.1	5	0.1388
UK351	25 (100.0%)	Apr-13, May-17	25	B.1.1, B.1.1.10	0	active today

Lineage name	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK214	22 (100.0%)	Apr-08, May-13	22	B.1.1	4	0.4167
UK36	19 (100.0%)	Apr-15, May-12	19	B.1	5	0.3
UK135	18 (100.0%)	Apr-08, May-11	18	B.1.p11	6	0.3235
UK248	18 (100.0%)	Apr-08, May-11	18	B.1.1	6	0.3235

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.

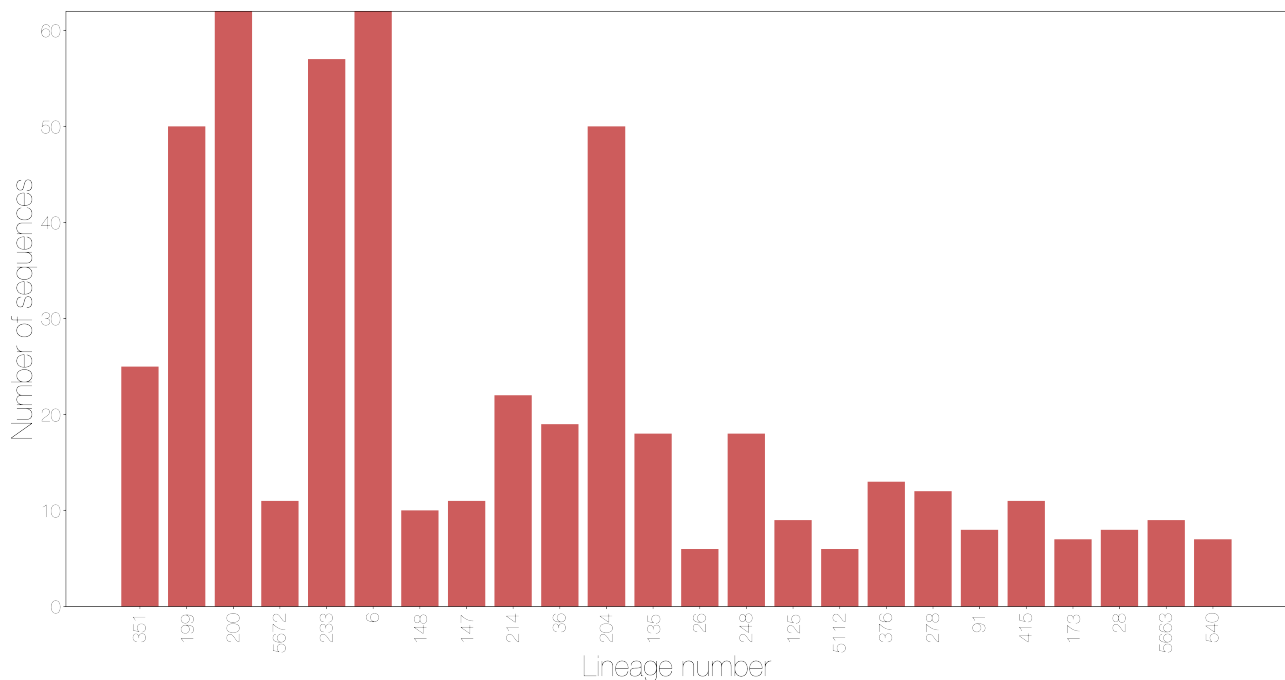The raw data for this bar chart are in the table above.

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.

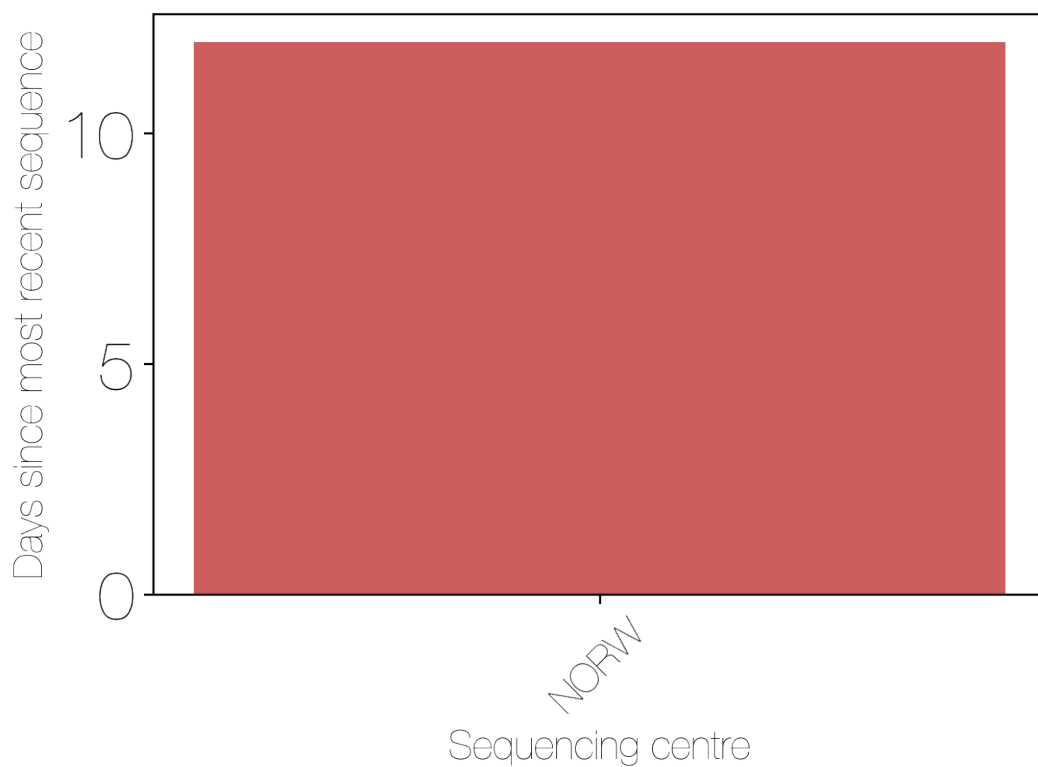Figure 2: Lag since the most recent sequence from each sequencing centre to most current date

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.
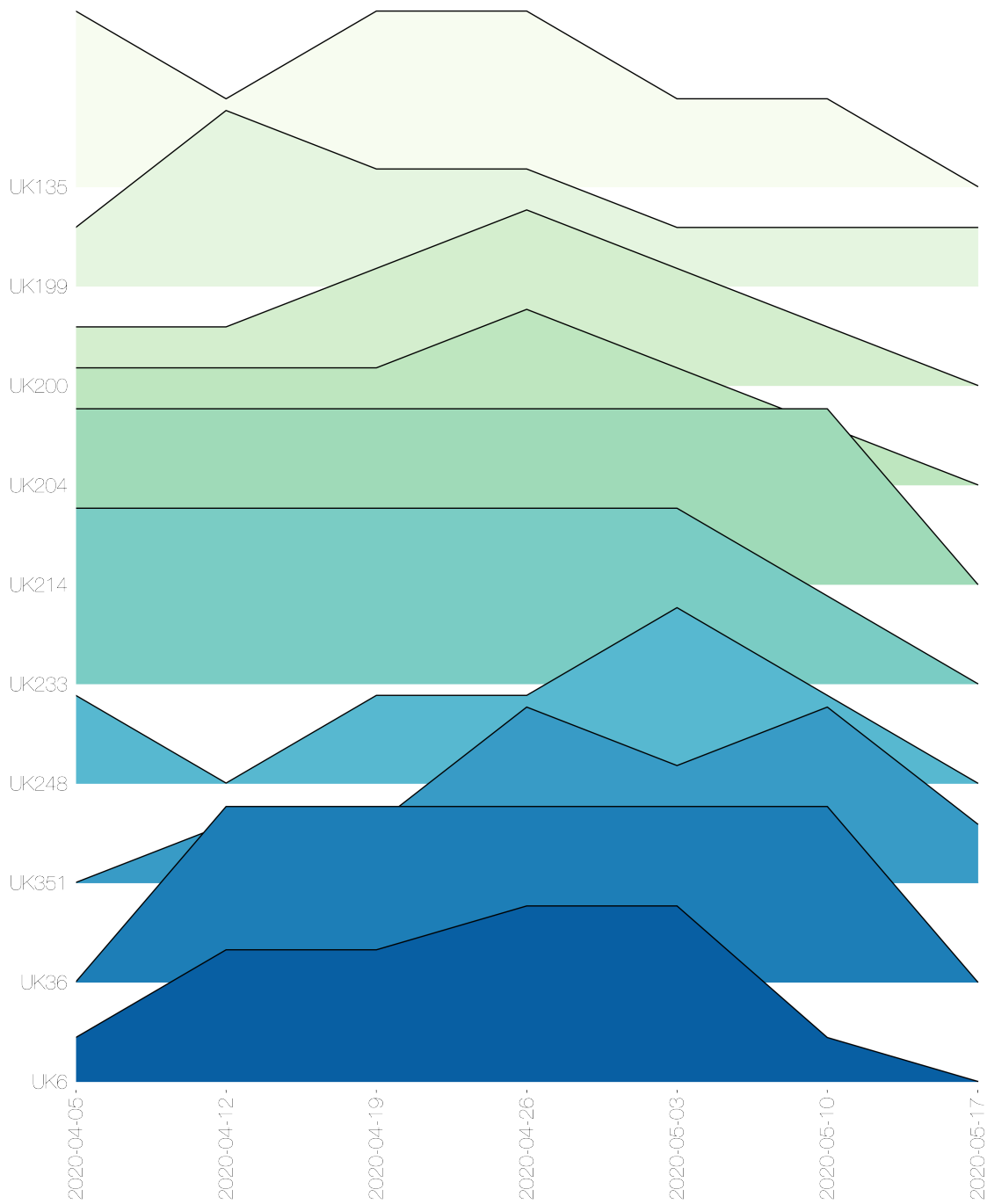
Figure 3: Lineages by number of adm2 regions present by epiweek
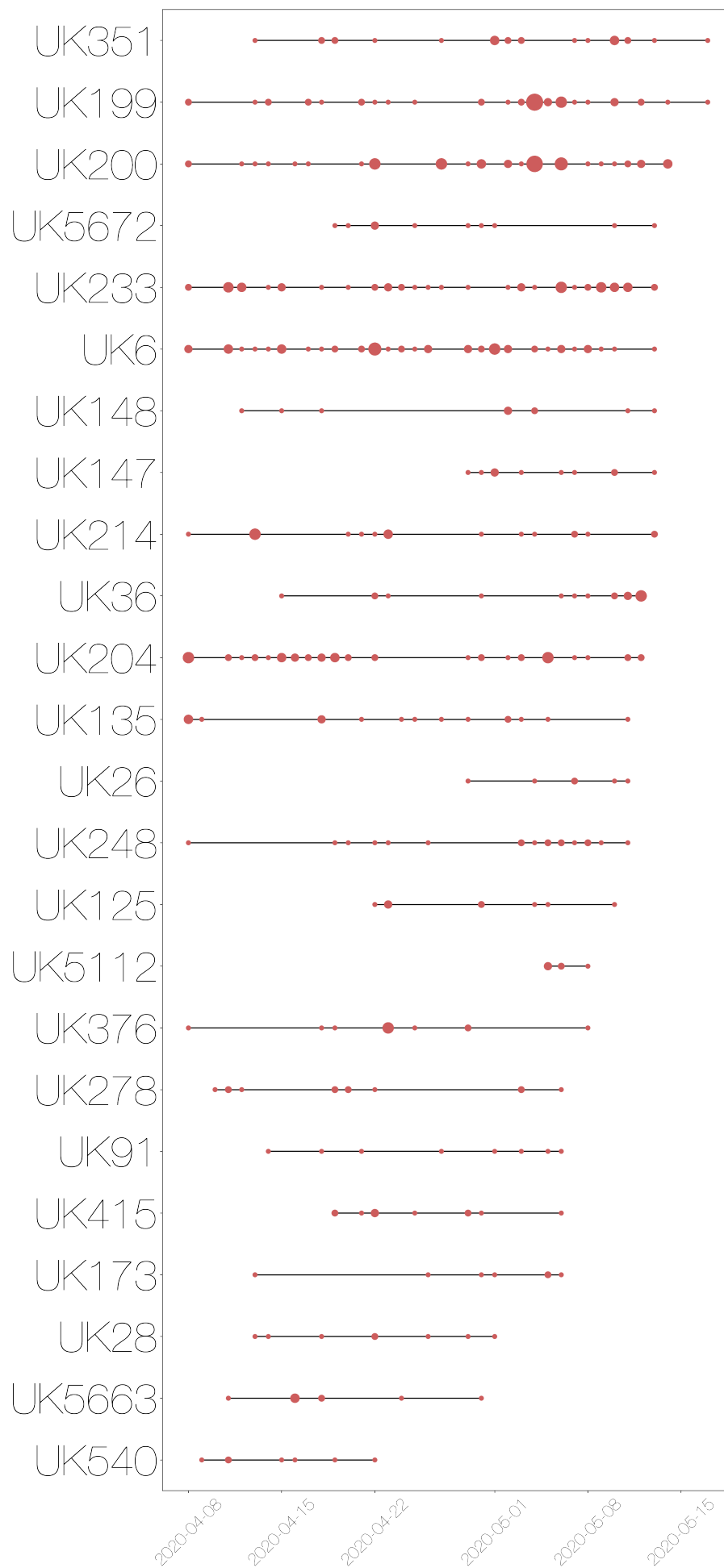
Figure 4: Timeline of lineages, sized by number of sequences from each country.

Figure 5: Lineage starts per week, split by singletons and non-singletons

Figure 6: Sequences taken on each day by country
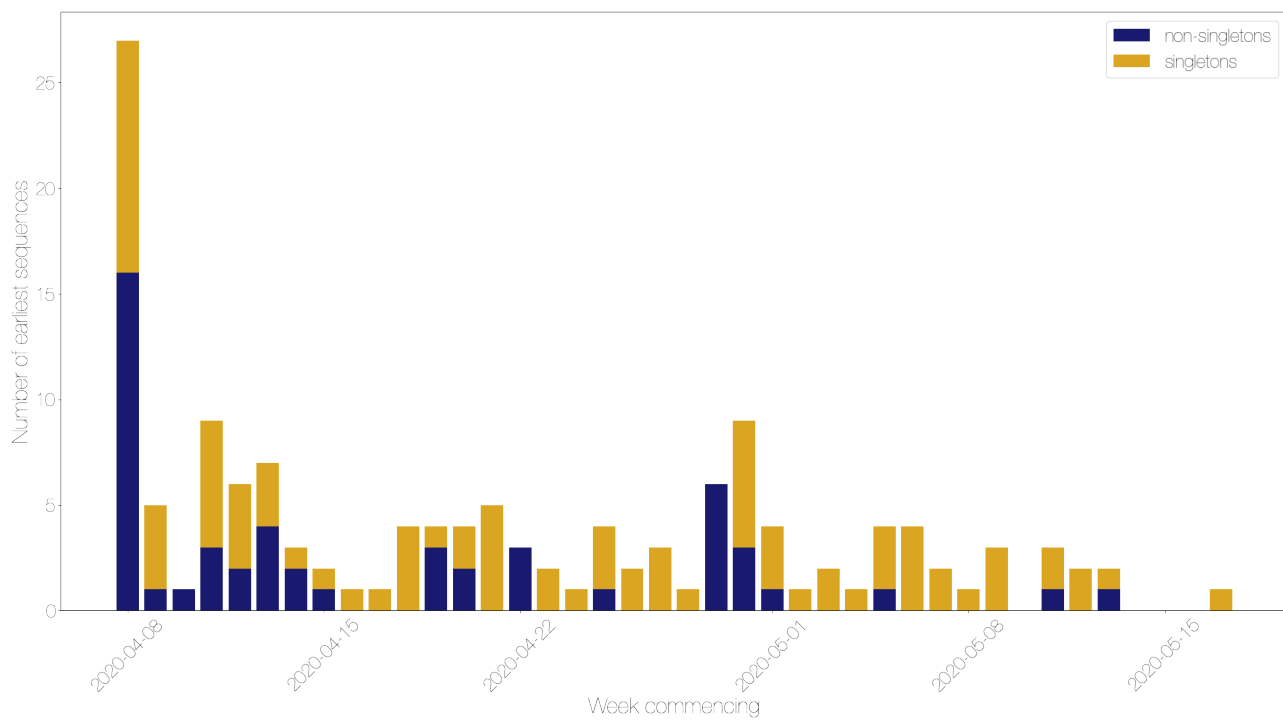
Number of sequen

- 0-10
- 10-50
- 100-150
- 400-500
- No sequenc

There are 20 sequences without enough geographical information to map from this centre.

There are some sequences with locations that are not matched to real Admin2 regions, some manual curation required.

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	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK200	62 (100.0%)	Apr-08, May-14	62	B.1.p11	3	0.1967
UK6	62 (100.0%)	Apr-08, May-13	62	B.1	4	0.1434
UK233	57 (100.0%)	Apr-08, May-13	57	B.1.1	4	0.1562
UK199	50 (100.0%)	Apr-08, May-17	50	B.1.5.5	0	active today
UK204	50 (100.0%)	Apr-08, May-12	50	B.1.1	5	0.1388
UK351	25 (100.0%)	Apr-13, May-17	25	B.1.1, B.1.1.10	0	active today
UK214	22 (100.0%)	Apr-08, May-13	22	B.1.1	4	0.4167
UK36	19 (100.0%)	Apr-15, May-12	19	B.1	5	0.3
UK135	18 (100.0%)	Apr-08, May-11	18	B.1.p11	6	0.3235
UK248	18 (100.0%)	Apr-08, May-11	18	B.1.1	6	0.3235
UK376	13 (100.0%)	Apr-08, May-08	13	B.1.1	9	0.2778
UK278	12 (100.0%)	Apr-10, May-06	12	B.1.1	11	0.2149
UK147	11 (100.0%)	Apr-29, May-13	11	B.1.1	4	0.35
UK5672	11 (100.0%)	Apr-19, May-13	11	B.2	4	0.6
UK415	11 (100.0%)	Apr-19, May-06	11	B.1	11	0.1545
UK148	10 (100.0%)	Apr-12, May-13	10	B.1.1	4	0.8611
UK125	9 (100.0%)	Apr-22, May-10	9	B.1.1	7	0.3214
UK5663	9 (100.0%)	Apr-11, Apr-30	9	B.2	17	0.1397
UK91	8 (100.0%)	Apr-14, May-06	8	B.1.1	11	0.2857
UK28	8 (100.0%)	Apr-13, May-01	8	B.1.1.10	16	0.1607
UK173	7 (100.0%)	Apr-13, May-06	7	B	11	0.3485

Lineage name	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK540	7 (100.0%)	Apr-09, Apr-22	7	B.1.1, B.1.1.p15	25	0.0867
UK26	6 (100.0%)	Apr-29, May-11	6	B.1.1.3	6	0.4
UK5112	6 (100.0%)	May-05, May-08	6	B.1	9	0.0667

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

Week commencing	UK6	UK200	UK233	UK204	UK199	UK351	UK214	UK36	UK248	UK135
2020-04-05	1	1	2	2	1	0	1	0	1	2
2020-04-12	3	1	2	2	3	1	1	1	0	1
2020-04-19	3	2	2	2	2	1	1	1	1	2
2020-04-26	4	3	2	3	2	3	1	1	1	2
2020-05-03	4	2	2	2	1	2	1	1	2	1
2020-05-10	1	1	1	1	1	3	1	1	1	1
2020-05-17	0	0	0	0	1	1	0	0	0	0

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

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

Day	England
2020-04-08	41
2020-04-09	6
2020-04-10	1
2020-04-11	25
2020-04-12	16
2020-04-13	20
2020-04-14	11
2020-04-15	15
2020-04-16	10
2020-04-17	8
2020-04-18	22
2020-04-19	19
2020-04-20	13
2020-04-21	14
2020-04-22	41
2020-04-23	23
2020-04-24	7
2020-04-25	12
2020-04-26	9
2020-04-27	13
2020-04-28	1
2020-04-29	20
2020-04-30	30
2020-05-01	30
2020-05-02	19
2020-05-03	23
2020-05-04	37
2020-05-05	24
2020-05-06	41
2020-05-07	14
2020-05-08	18
2020-05-09	12
2020-05-10	22
2020-05-11	20
2020-05-12	19
2020-05-13	11
2020-05-14	5
2020-05-17	5

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

Admin2	Country	Number of sequences	Sequence group
CAMBRIDGESHIRE	England	30	10-50
ESSEX	England	19	10-50
GREATER LONDON	England	2	1-10
LINCOLNSHIRE	England	11	10-50
MANCHESTER	England	1	1-10
NORFOLK	England	480	400-500
SUFFOLK	England	112	100-150
WILTSHIRE	England	1	1-10

Admin2	Country	Number of sequences	Sequence group