UK lineages summary report

This report gives summaries of UK specific lineages sequenced by EDIN for week 2020-05-29. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-05-24. The analysis (eg time since last sample) is therefore undertaken from this date. 1049 sequences in the UK from the sequencing centre EDIN 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.

365 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: 10 are pending extinction, ie last seen three weeks ago. 5 lineages have gone quiet, ie haven't been seen this week. 4 lineages have reactivated. 4 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	Scotland	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK36	82 (100.0%)	Mar-21, May-04	82	B.1	20	0.0272
UK175	65 (100.0%)	Mar-22, May-04	65	B.1	20	0.0336
UK2464	41 (100.0%)	Mar-20, May-12	41	B.1.p11	12	0.1104
UK5	31 (100.0%)	Mar-18, May-18	31	B.1.1.1	6	0.3389
UK296	25 (100.0%)	Apr-08, May-13	25	B.1.5	11	0.1326
UK21	24 (100.0%)	Mar-18, May-23	24	B.1.40	1	2.8696

Lineage name	Scotland	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK53	24 (100.0%)	Apr-16, May-21	24	B.1.1.4	3	0.5072
UK44	21 (100.0%)	Mar-25, May-01	21	B	23	0.0804
UK461	19 (100.0%)	Apr-18, May-19	19	B.1.5	5	0.3444
UK150	19 (100.0%)	Mar-21, Apr-22	19	B.1.1.p12	32	0.0556

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.

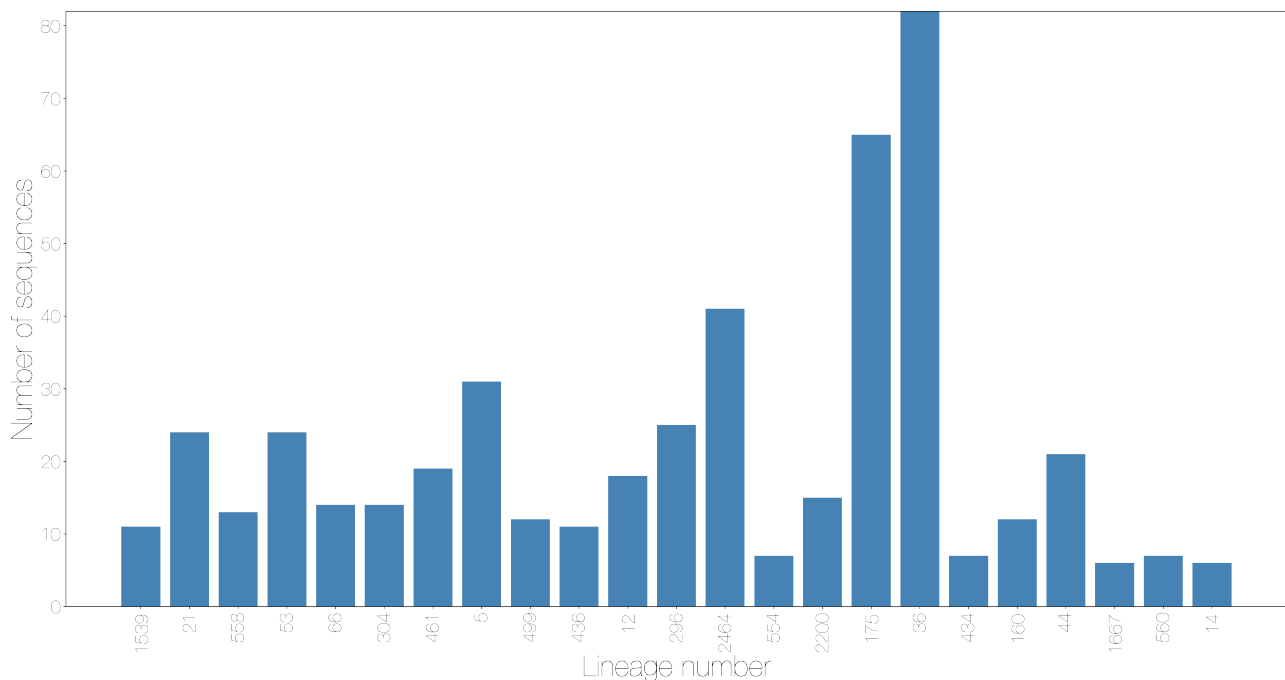The raw data for this bar chart are in the table above.

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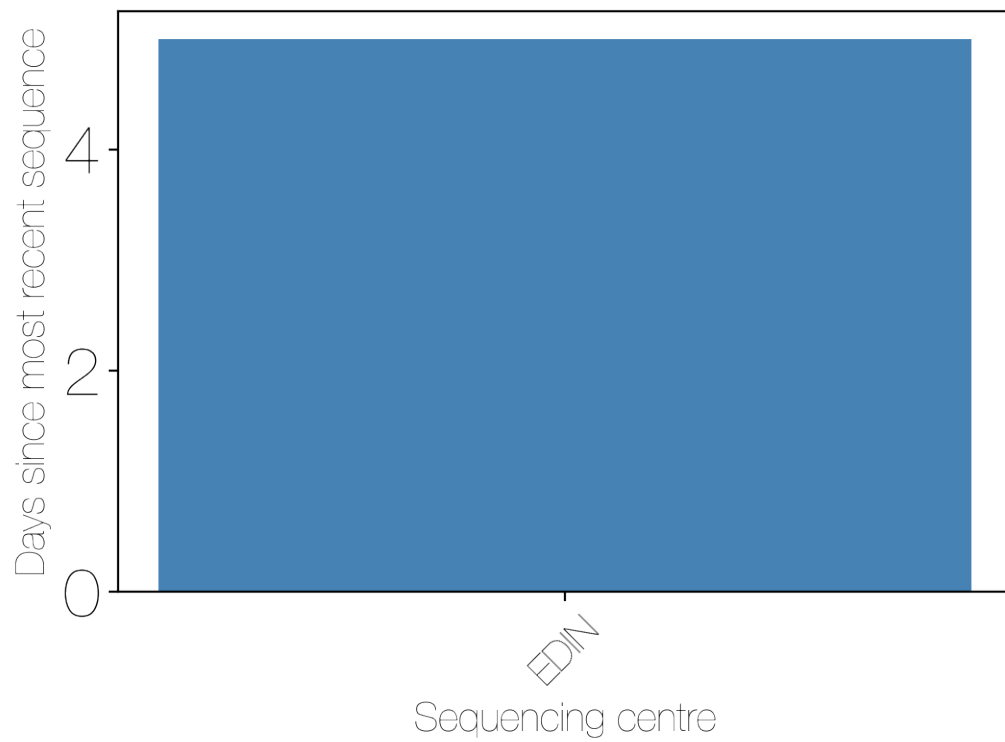
Figure 2: Lag since the most recent sequence from each sequencing centre to most current date

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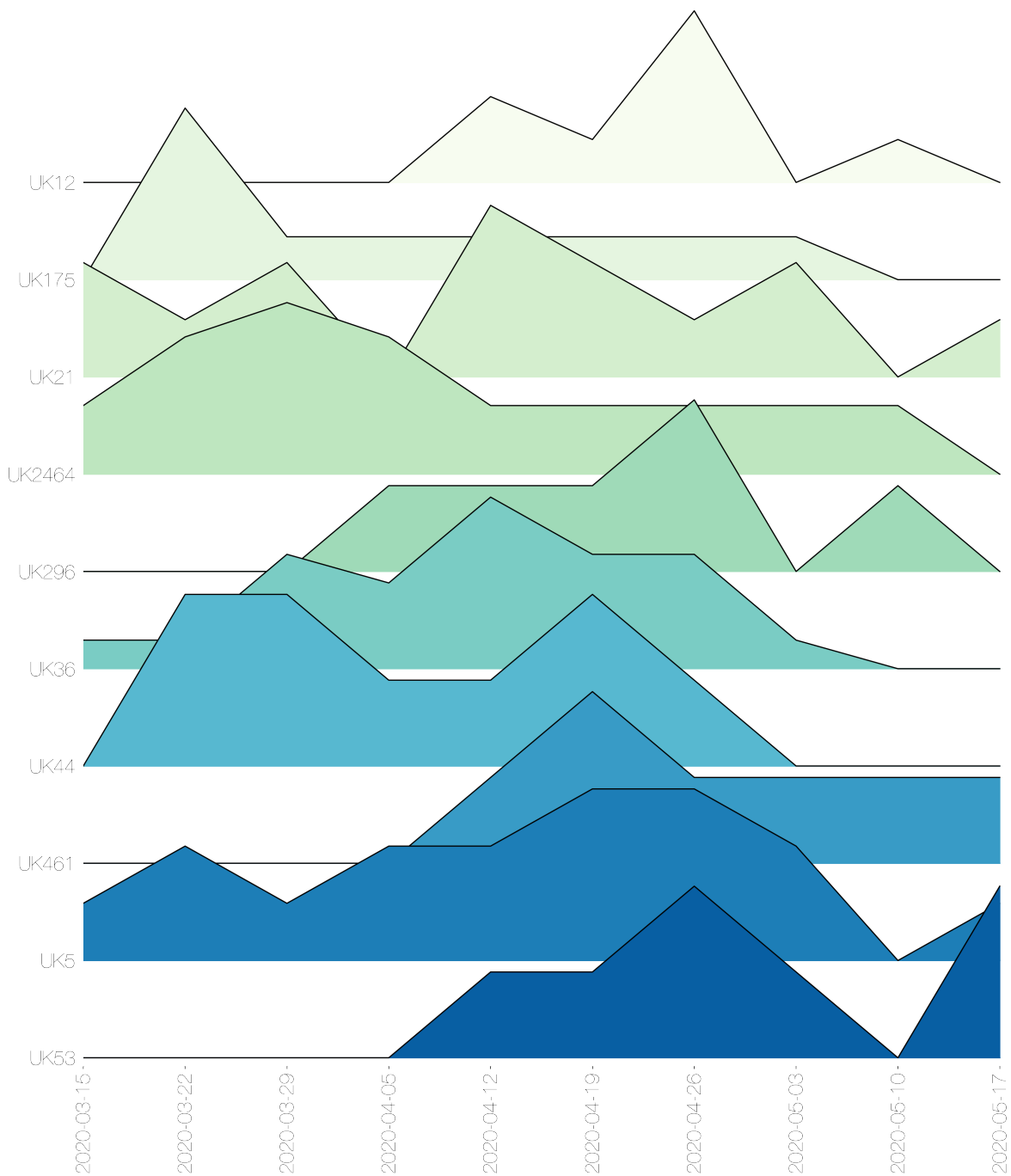
Figure 3: Lineages by number of adm2 regions present by epiweek
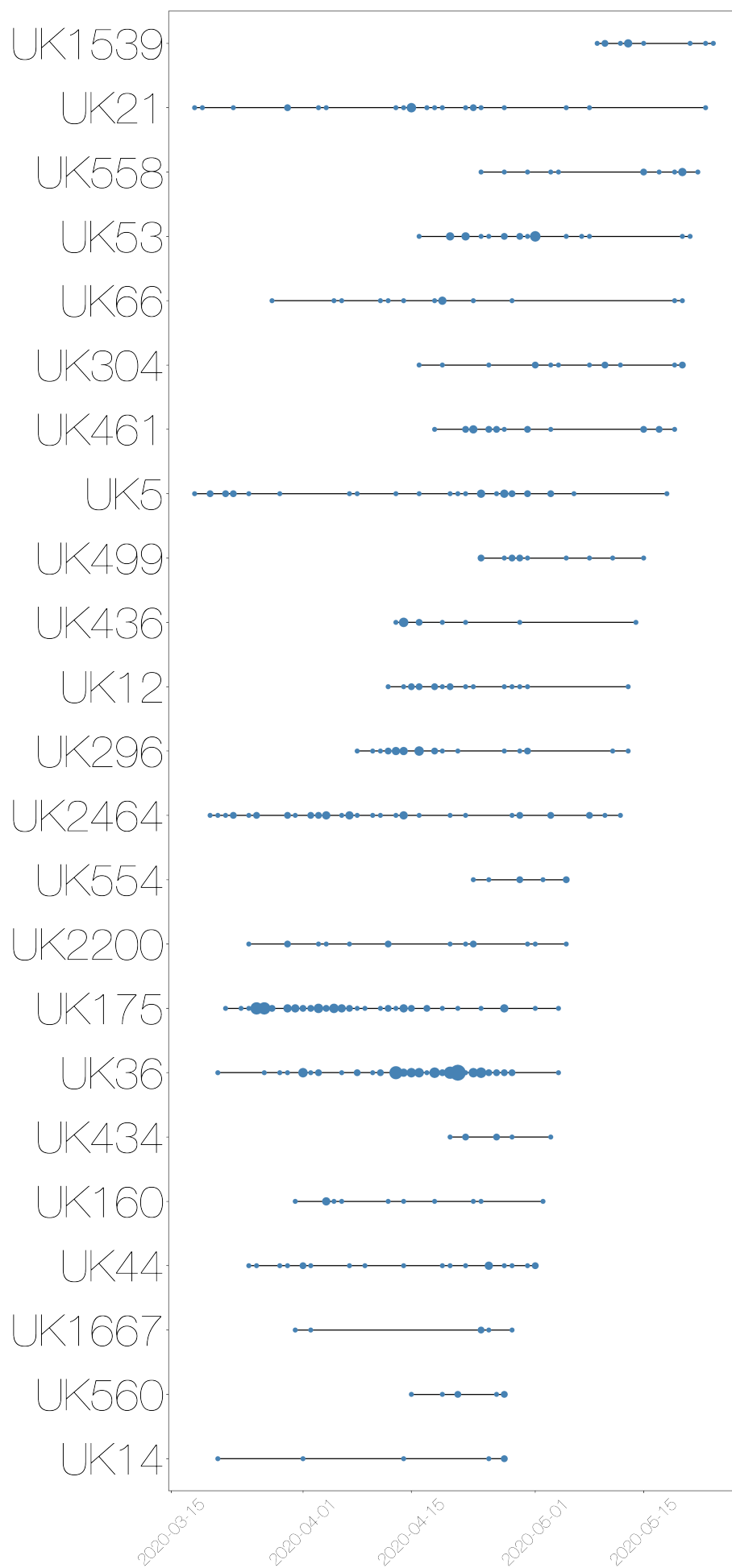
Figure 4: Timeline of lineages, sized by number of sequences from each country.

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

There are 70 sequences without enough geographical information to map from this centre.

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.

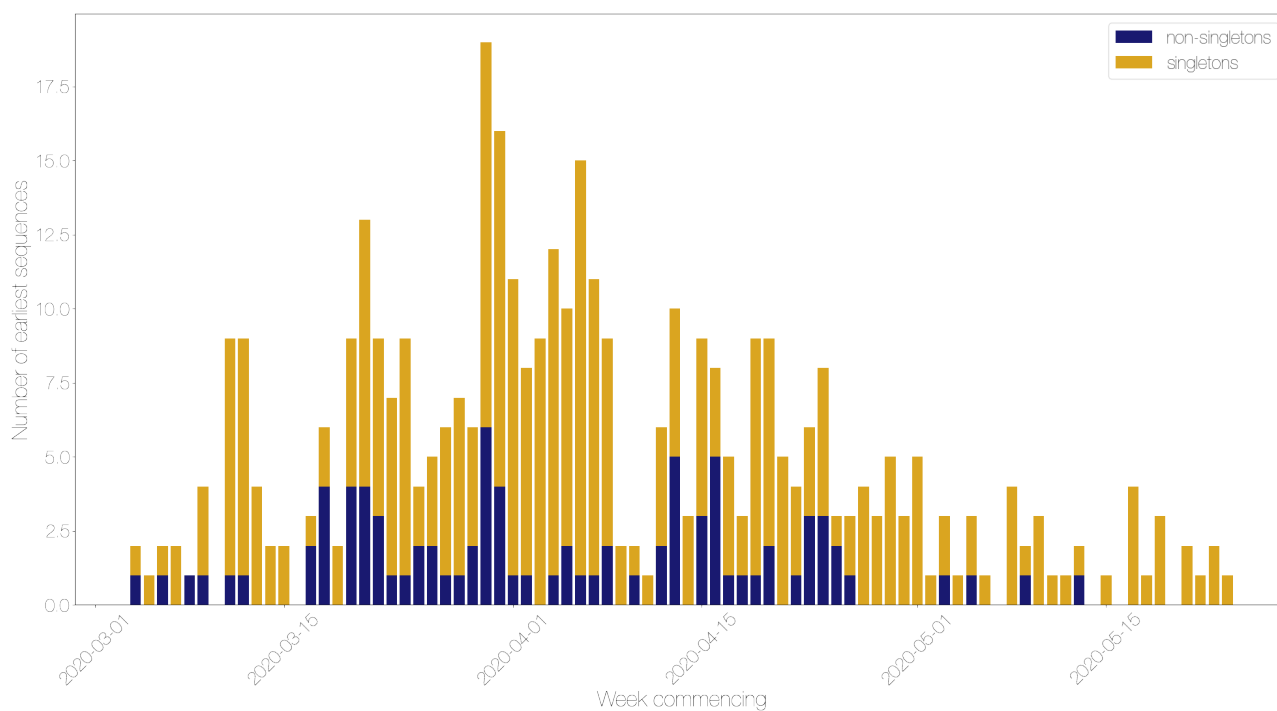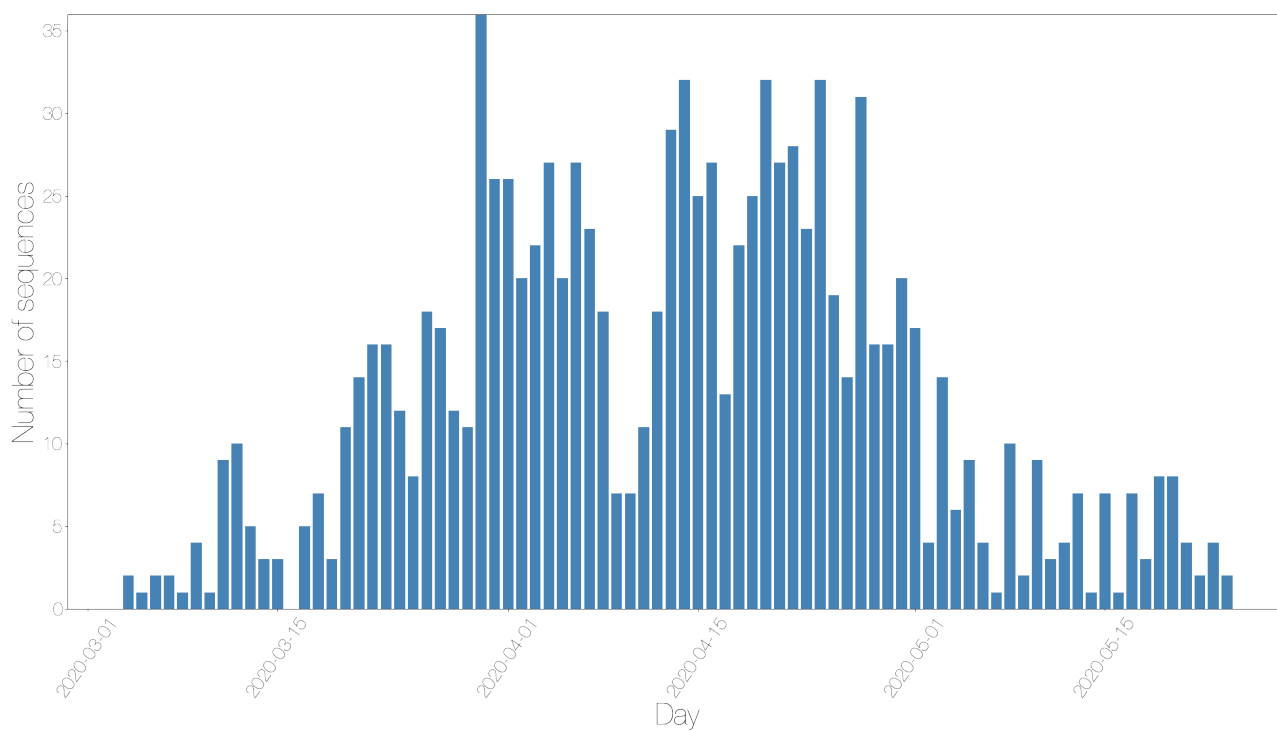| Lineage name	Scotland	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK36	82 (100.0%)	Mar-21, May-04	82	B.1	20	0.0272
UK175	65 (100.0%)	Mar-22, May-04	65	B.1	20	0.0336
UK2464	41 (100.0%)	Mar-20, May-12	41	B.1.p11	12	0.1104
UK5	31 (100.0%)	Mar-18, May-18	31	B.1.1.1	6	0.3389
UK296	25 (100.0%)	Apr-08, May-13	25	B.1.5	11	0.1326
UK21	24 (100.0%)	Mar-18, May-23	24	B.1.40	1	2.8696
UK53	24 (100.0%)	Apr-16, May-21	24	B.1.1.4	3	0.5072
UK44	21 (100.0%)	Mar-25, May-01	21	B	23	0.0804
UK461	19 (100.0%)	Apr-18, May-19	19	B.1.5	5	0.3444
UK150	19 (100.0%)	Mar-21, Apr-22	19	B.1.1.p12	32	0.0556
UK12	18 (100.0%)	Apr-12, May-13	18	B.1.p11	11	0.1658
UK2200	15 (100.0%)	Mar-25, May-05	15	B.1.5, B.1.5.6	19	0.1541
UK66	14 (100.0%)	Mar-28, May-20	14	B.1.1.8	4	1.0192
UK43	14 (100.0%)	Mar-26, Apr-18	14	A.5	36	0.0491
UK156	14 (100.0%)	Mar-18, Apr-18	14	B.1.71	36	0.0662
UK304	14 (100.0%)	Apr-16, May-20	14	B.1.1.14	4	0.6538
UK558	13 (100.0%)	Apr-24, May-22	13	B.1.5	2	1.1667
UK160	12 (100.0%)	Mar-31, May-02	12	B.1.1	22	0.1322
UK499	12 (100.0%)	Apr-24, May-15	12	B.1.5	9	0.2121
UK562	11 (100.0%)	Mar-27, Apr-25	11	B.1	29	0.1
UK436	11 (100.0%)	Apr-13, May-14	11	B.1.5	10	0.31

Lineage name	Scotland	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK1539	11 (100.0%)	May-09, May-24	11	B.1.5	0	active today
UK225	11 (100.0%)	Mar-17, Apr-05	11	B.2	49	0.0388
UK414	10 (100.0%)	Apr-05, Apr-22	10	B.1.5	32	0.059
UK1548	9 (100.0%)	Apr-13, Apr-24	9	B.1.5, B.1.5.5	30	0.0458
UK560	7 (100.0%)	Apr-15, Apr-27	7	B.1.5	27	0.0741
UK554	7 (100.0%)	Apr-23, May-05	7	B.1.5	19	0.1053
UK434	7 (100.0%)	Apr-20, May-03	7	B.1.5	21	0.1032
UK14	6 (100.0%)	Mar-21, Apr-27	6	B	27	0.2741
UK555	6 (100.0%)	Apr-13, Apr-25	6	B.1.5	29	0.0828
UK1667	6 (100.0%)	Mar-31, Apr-28	6	B.1.p9	26	0.2154
UK133	6 (100.0%)	Mar-22, Apr-25	6	B.1	29	0.2345

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

Week commencing	UK36	UK175	UK2464	UK5	UK296	UK53	UK21	UK44	UK461	UK12
2020-03-15	1	0	2	1	0	0	2	0	0	0
2020-03-22	1	4	4	2	0	0	1	2	0	0
2020-03-29	4	1	5	1	0	0	2	2	0	0
2020-04-05	3	1	4	2	1	0	0	1	0	0
2020-04-12	6	1	2	2	1	1	3	1	1	2
2020-04-19	4	1	2	3	1	1	2	2	2	1
2020-04-26	4	1	2	3	2	2	1	1	1	4
2020-05-03	1	1	2	2	0	1	2	0	1	0
2020-05-10	0	0	2	0	1	0	0	0	1	1
2020-05-17	0	0	0	1	0	2	1	0	1	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
2020-03-04	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	2	4	6
2020-03-19	2	0	2
2020-03-20	5	4	9
2020-03-21	9	4	13
2020-03-22	6	3	9
2020-03-23	6	1	7
2020-03-24	8	1	9
2020-03-25	2	2	4
2020-03-26	3	2	5
2020-03-27	5	1	6
2020-03-28	6	1	7
2020-03-29	4	2	6
2020-03-30	13	6	19
2020-03-31	12	4	16
2020-04-01	10	1	11
2020-04-02	7	1	8
2020-04-03	9	0	9
2020-04-04	11	1	12
2020-04-05	8	2	10
2020-04-06	14	1	15
2020-04-07	10	1	11
2020-04-08	7	2	9
2020-04-09	2	0	2
2020-04-10	1	1	2
2020-04-11	1	0	1
2020-04-12	4	2	6
2020-04-13	5	5	10
2020-04-14	3	0	3
2020-04-15	6	3	9
2020-04-16	3	5	8
2020-04-17	4	1	5
2020-04-18	2	1	3
2020-04-19	8	1	9
2020-04-20	7	2	9

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-04-21	5	0	5
2020-04-22	3	1	4
2020-04-23	3	3	6
2020-04-24	5	3	8
2020-04-25	1	2	3
2020-04-26	2	1	3
2020-04-27	4	0	4
2020-04-28	3	0	3
2020-04-29	5	0	5
2020-04-30	3	0	3
2020-05-01	5	0	5
2020-05-02	1	0	1
2020-05-03	2	1	3
2020-05-04	1	0	1
2020-05-05	2	1	3
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	4	0	4
2020-05-18	1	0	1
2020-05-19	3	0	3
2020-05-21	2	0	2
2020-05-22	1	0	1
2020-05-23	2	0	2
2020-05-24	1	0	1

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

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	18
2020-03-27	17
2020-03-28	12
2020-03-29	11
2020-03-30	36
2020-03-31	26
2020-04-01	26
2020-04-02	20
2020-04-03	22
2020-04-04	27
2020-04-05	20
2020-04-06	27
2020-04-07	23
2020-04-08	18
2020-04-09	7
2020-04-10	7
2020-04-11	11
2020-04-12	18
2020-04-13	29
2020-04-14	32
2020-04-15	25
2020-04-16	27
2020-04-17	13
2020-04-18	22
2020-04-19	25
2020-04-20	32

Day	Scotland
2020-04-21	27
2020-04-22	28
2020-04-23	23
2020-04-24	32
2020-04-25	19
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	7
2020-05-18	3
2020-05-19	8
2020-05-20	8
2020-05-21	4
2020-05-22	2
2020-05-23	4
2020-05-24	2

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

Admin2	Country	Number of sequences	Sequence group
ABERDEEN	Scotland	1	1-10
ABERDEENSHIRE	Scotland	5	1-10
ANGUS	Scotland	13	10-50
CLACKMANNANSHIRE	Scotland	2	1-10
DUMFRIES AND GALLOWAY	Scotland	2	1-10
DUNDEE	Scotland	82	50-100
EAST LOTHIAN	Scotland	54	50-100
EDINBURGH	Scotland	402	400-500
FALKIRK	Scotland	4	1-10
FIFE	Scotland	42	10-50
GLASGOW	Scotland	1	1-10
MIDLOTHIAN	Scotland	127	100-150
NORTHUMBERLAND	England	1	1-10
PERTHSHIRE AND KINROSS	Scotland	18	10-50
SCOTTISH BORDERS	Scotland	128	100-150
SOUTH LANARKSHIRE	Scotland	4	1-10
WEST LOTHIAN	Scotland	92	50-100