

Lineages report for LIVE

This report gives summaries of UK specific lineages sequenced by LIVE for week 2020-09-13. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-08-31. The analysis (eg time since last sample) is therefore undertaken from this date. 3036 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 33 and the maximum is 1666

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

348 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 5 that remain: 2 are pending extinction, ie last seen three weeks ago. 1 has gone quiet, ie hasn't been seen this week. 1 has reactivated. 1 lineage has been continuously circulating.

The following table contains information about the ten largest lineages and the number of sequences the dataset. Information about other lineages 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 expected 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 2020-07-20

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	Global lineage	Total
UK5	574 (100.0%)	Mar-03, Aug-29	B.1.1.1, B.1.1.10, B.1.1	574 taxa
UK1951	185 (100.0%)	Mar-02, Jul-25	B.1.1.1, B.1.1	185 taxa
UK175	169 (100.0%)	Feb-28, Aug-17	B.1, B.1.76, B.1.5, B.1.82	169 taxa
UK1683	160 (100.0%)	Mar-08, Aug-03	B.1.1.1, B.1.1.10, B.1.1	160 taxa
UK1684	143 (100.0%)	Mar-14, Jun-20	B.1.1.1, B.1.1	143 taxa
UK107	71 (100.0%)	Mar-09, May-19	B.2, B.2.1	71 taxa
UK199	70 (100.0%)	Mar-11, Aug-20	B.1, B.1.5	70 taxa
UK336	61 (100.0%)	Mar-11, Jun-08	B.1.93, B.1	61 taxa
UK72	60 (100.0%)	Mar-06, Apr-14	B, B.2	60 taxa
UK109	53 (100.0%)	Mar-17, Jul-23	B.1, B.1.5	53 taxa

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.

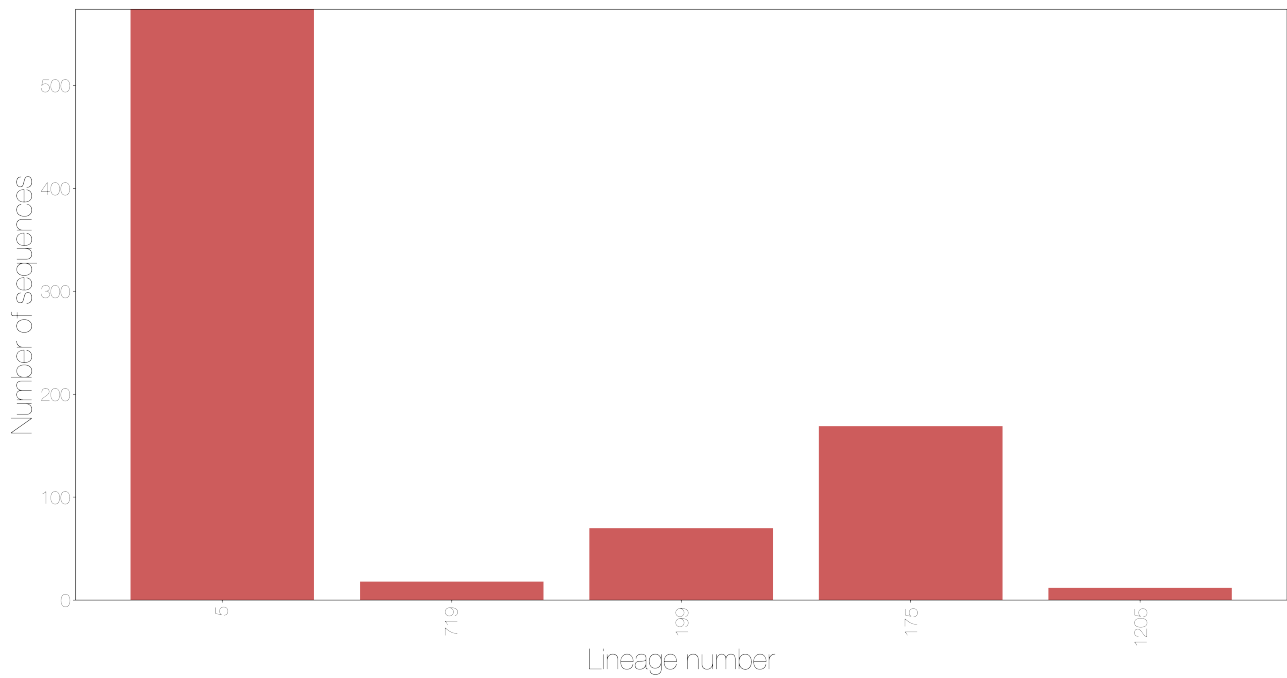


Figure 1: Number of sequences sampled in a lineage by country

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.

```
-----NameError
Traceback (most recent call last)<ipython-input-1-2620455843ef> in
<module>
      2     lag_dict, lags = dp.sequencing_centre_lags(taxa, sc_dict,
current_date, country)
      3 elif sequencing_centre != "":
----> 4     print("The lag for this sequencing centre is " +
str(lags[sequencing_centre]) + " days")
NameError: name 'lags' is not defined
```

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.

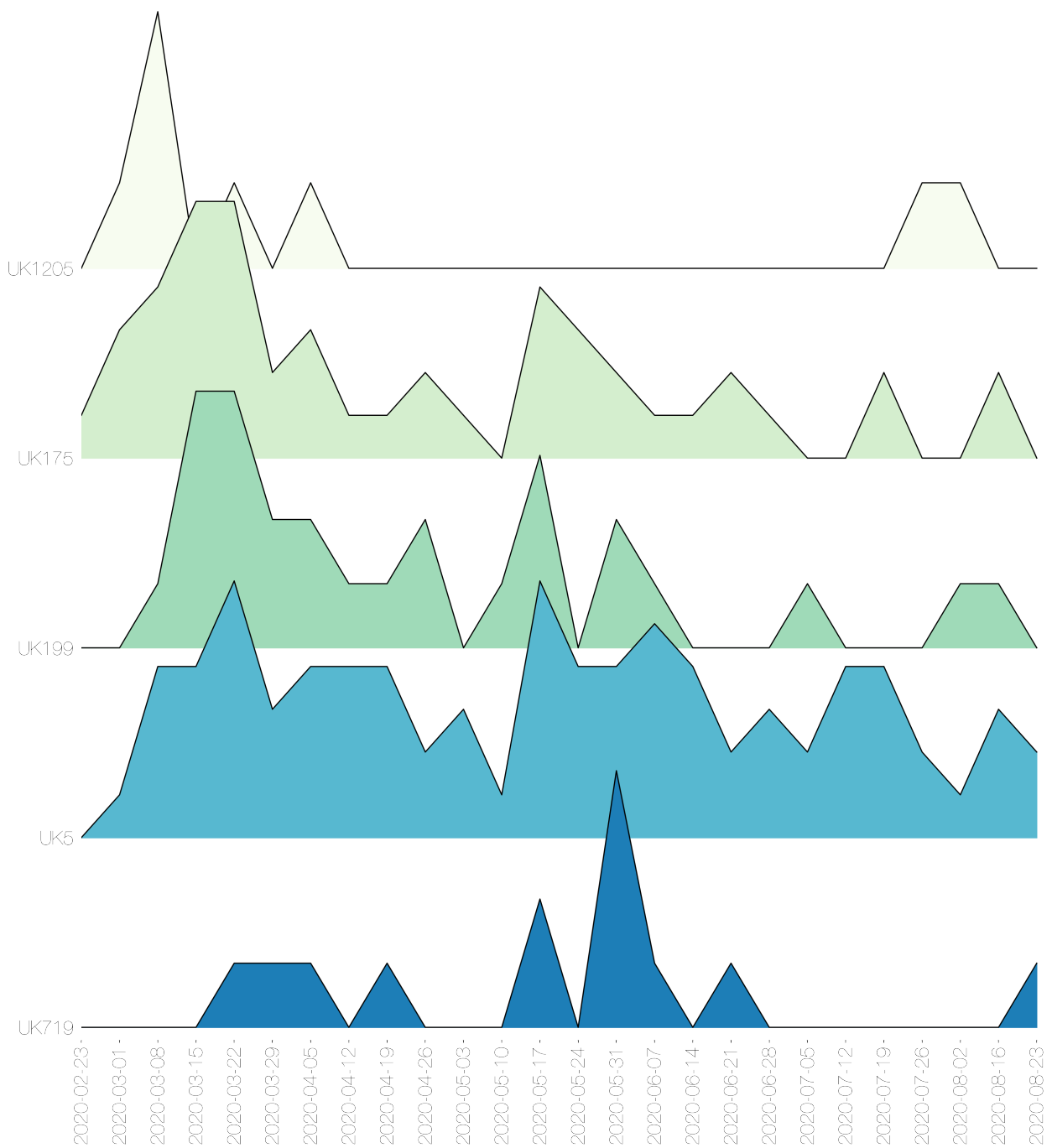


Figure 2: Lineages by number of adm2 regions present by epiweek

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.

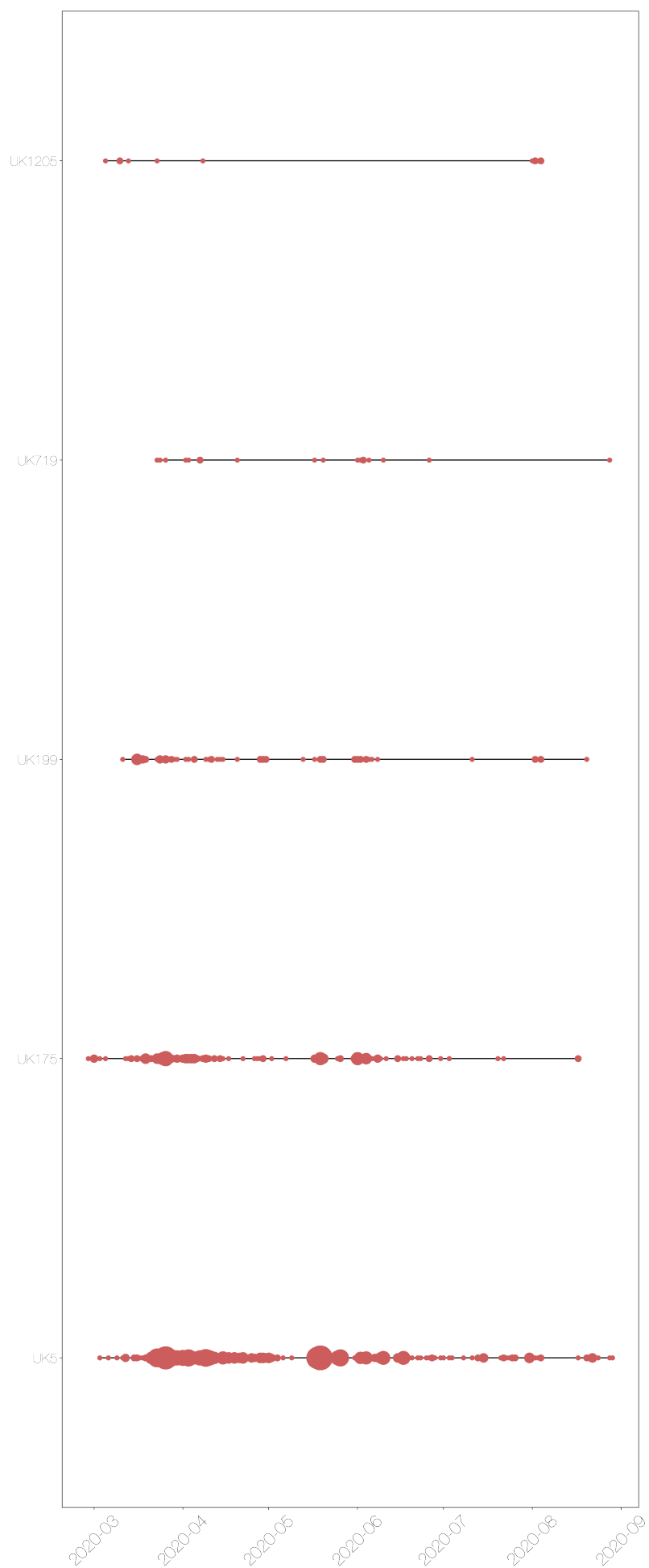


Figure 3: Timeline of lineages, sized by number of sequences from each country.

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

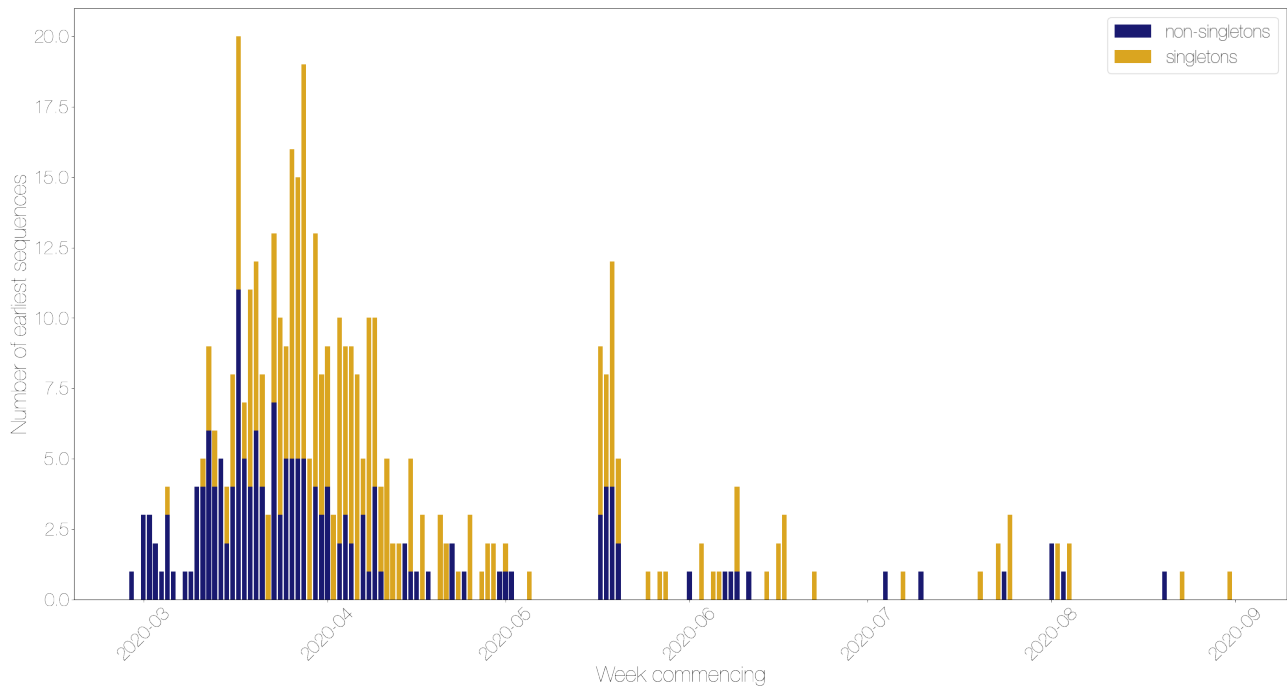


Figure 4: Lineage starts per week, split by singletons and non-singletons

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.

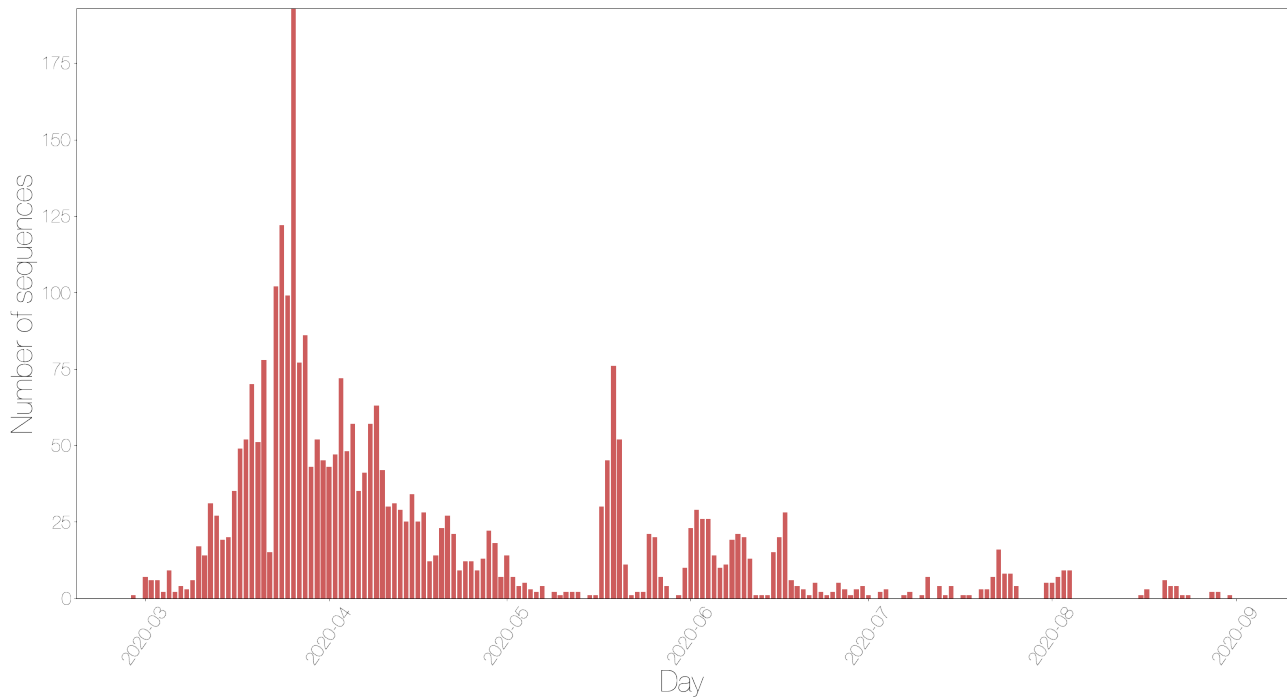


Figure 5: Sequences taken on each day by country

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 739 sequences without enough geographical information to map from this centre.

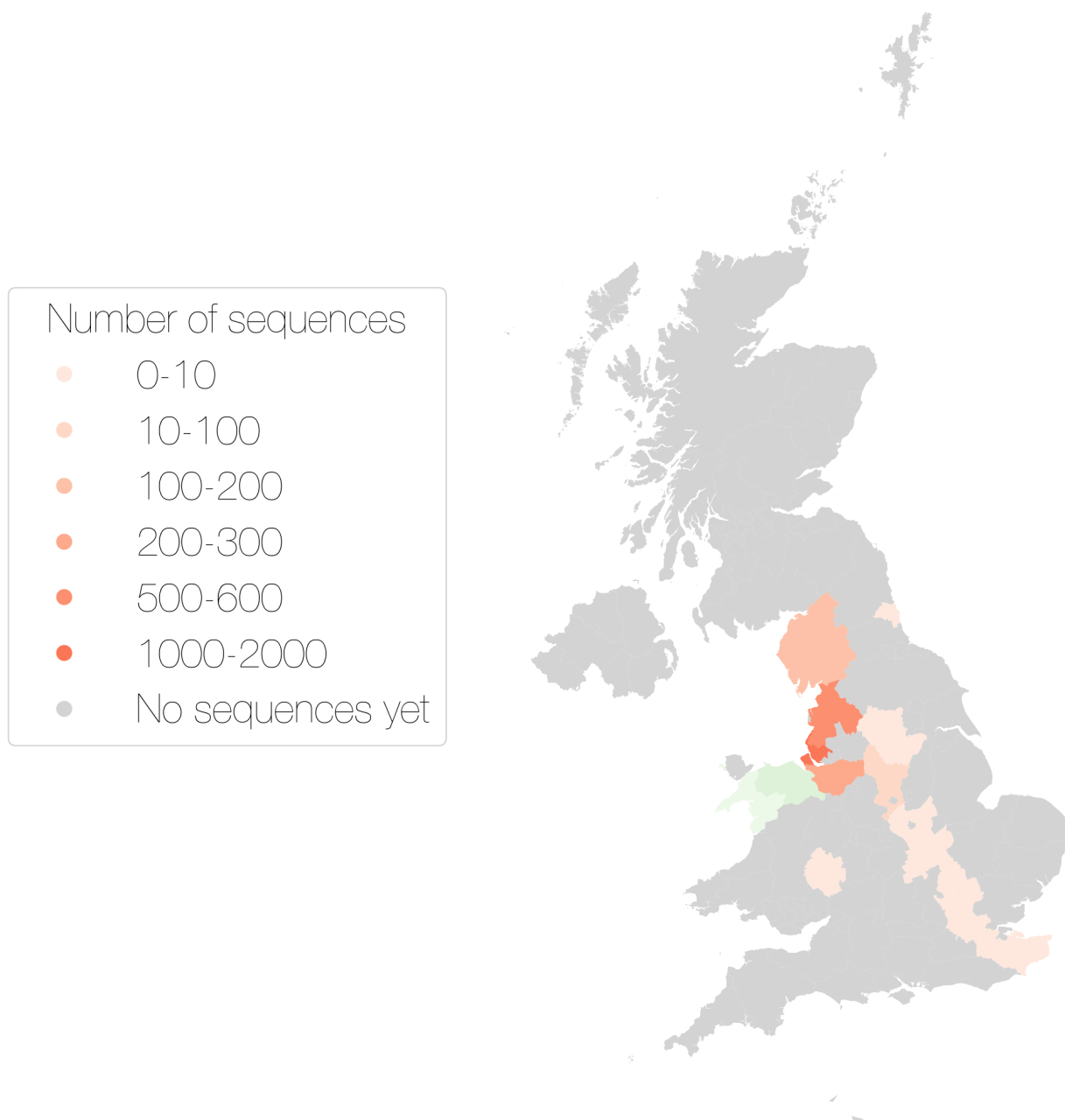


Figure 6: Map showing the number of sequences sampled by adm2 region

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	Global lineage	Total
UK5	574 (100.0%)	Mar-03, Aug-29	B.1.1.1, B.1.1.10, B.1.1	574 taxa
UK1951	185 (100.0%)	Mar-02, Jul-25	B.1.1.1, B.1.1	185 taxa
UK175	169 (100.0%)	Feb-28, Aug-17	B.1, B.1.76, B.1.5, B.1.82	169 taxa
UK1683	160 (100.0%)	Mar-08, Aug-03	B.1.1.1, B.1.1.10, B.1.1	160 taxa
UK1684	143 (100.0%)	Mar-14, Jun-20	B.1.1.1, B.1.1	143 taxa
UK107	71 (100.0%)	Mar-09, May-19	B.2, B.2.1	71 taxa
UK199	70 (100.0%)	Mar-11, Aug-20	B.1, B.1.5	70 taxa
UK336	61 (100.0%)	Mar-11, Jun-08	B.1.93, B.1	61 taxa
UK72	60 (100.0%)	Mar-06, Apr-14	B, B.2	60 taxa
UK109	53 (100.0%)	Mar-17, Jul-23	B.1, B.1.5	53 taxa
UK274	51 (100.0%)	Mar-16, Apr-18	B, B.3	51 taxa
UK600	47 (100.0%)	Mar-01, May-19	B.1, B.1.1	47 taxa
UK1060	41 (100.0%)	Mar-13, May-25	B.1.1	41 taxa
UK501	31 (100.0%)	Mar-20, Jun-02	B.1, B.1.86	31 taxa
UK1264	31 (100.0%)	Mar-18, Jun-22	B.1.1	31 taxa
UK4	30 (100.0%)	Mar-17, Apr-14	B	30 taxa
UK167	29 (100.0%)	Mar-11, Apr-06	B.1	29 taxa
UK1037	29 (100.0%)	Mar-12, Jul-04	B.1.1.30, B.1.1	29 taxa
UK2916	29 (100.0%)	Mar-01, Jul-24	B.1, B.1.98	29 taxa
UK1911	28 (100.0%)	Mar-26, Apr-29	B.1.1	28 taxa
UK1843	26 (100.0%)	Mar-21, Jun-07	B.1.1	26 taxa
UK5676	22 (100.0%)	Mar-10, Apr-08	B.2, B.2.2	22 taxa
UK1155	22 (100.0%)	Mar-05, Jun-09	B.1.1	22 taxa
UK384	20 (100.0%)	Mar-05, Apr-08	B.2.1	20 taxa
UK2007	19 (100.0%)	Apr-04, Jun-24	B.1.1	19 taxa
UK131	19 (100.0%)	Mar-18, Apr-17	B, B.15	19 taxa
UK348	18 (100.0%)	May-18, Jul-23	B.1	18 taxa
UK719	18 (100.0%)	Mar-23, Aug-28	B.1.1	18 taxa
UK345	18 (100.0%)	Mar-19, Apr-15	B.2, B.2.2	18 taxa
UK646	18 (100.0%)	Mar-13, May-20	B.1.1	18 taxa
UK31	17 (100.0%)	Mar-10, Apr-18	B.3	17 taxa
UK527	17 (100.0%)	Mar-16, May-20	B.1	17 taxa
UK201	17 (100.0%)	Apr-07, Jun-09	B, B.2, B.18	17 taxa
UK315	17 (100.0%)	Mar-12, Apr-12	B.2, B.2.2	17 taxa
UK1535	16 (100.0%)	Mar-03, Apr-21	B.1.1	16 taxa
UK9	15 (100.0%)	Mar-14, Apr-09	B.1, B.1.13	15 taxa
UK1157	14 (100.0%)	Mar-27, Apr-19	B.1.1.7, B.1.1	14 taxa
UK267	14 (100.0%)	Mar-12, Mar-28	B.2	14 taxa
UK1145	14 (100.0%)	Mar-23, Apr-19	B.1.1	14 taxa
UK183	13 (100.0%)	Mar-14, Apr-03	B.2, B	13 taxa
UK108	12 (100.0%)	May-17, Jun-17	B.1	12 taxa
UK1205	12 (100.0%)	Mar-05, Aug-04	B.1.1.1, B.1.1	12 taxa
UK1487	12 (100.0%)	Mar-19, Apr-15	B.1, B.1.5	12 taxa
UK1681	11 (100.0%)	Apr-14, Jun-07	B.1.1.1, B.1.1	11 taxa
UK6	11 (100.0%)	Mar-17, Apr-29	B.1.75, B.1, B.1.5	11 taxa
UK2200	11 (100.0%)	Mar-17, May-21	B.1, B.1.5	11 taxa
UK934	11 (100.0%)	Mar-25, Apr-13	B.1.1	11 taxa
UK1129	11 (100.0%)	Mar-27, Jul-23	B.1.1	11 taxa
UK1202	10 (100.0%)	May-18, Jun-15	B.1.1	10 taxa
UK1855	10 (100.0%)	Jul-04, Jul-26	B.1.1	10 taxa
UK390	9 (100.0%)	Mar-26, May-01	B.1	9 taxa
UK1332	9 (100.0%)	Mar-28, Apr-05	B.1	9 taxa
UK832	9 (100.0%)	Mar-26, Apr-26	A.5	9 taxa
UK387	9 (100.0%)	Mar-18, Apr-15	B.1, B.1.86	9 taxa
UK870	8 (100.0%)	Mar-16, Apr-23	B.1.1	8 taxa
UK40	8 (100.0%)	Mar-27, Apr-20	B.16, B	8 taxa

Lineage name	England	Date range	Global lineage	Total
UK946	8 (100.0%)	Mar-25, Apr-23	B.1	8 taxa
UK601	7 (100.0%)	Mar-25, May-04	B, B.10	7 taxa
UK1300	7 (100.0%)	May-20, Jun-04	B.1.1	7 taxa
UK55	7 (100.0%)	Mar-10, Apr-08	B.3	7 taxa
UK86	7 (100.0%)	Mar-27, Apr-16	B.1	7 taxa
UK2022	7 (100.0%)	Apr-05, Jul-26	B.1.1	7 taxa
UK2464	7 (100.0%)	Mar-20, Jun-10	B.1	7 taxa
UK1700	6 (100.0%)	Apr-01, Apr-14	B.1	6 taxa
UK829	6 (100.0%)	Mar-17, Apr-03	B.2	6 taxa
UK1852	6 (100.0%)	Mar-17, Apr-10	B.1.1.1, B.1.1	6 taxa
UK1266	6 (100.0%)	Apr-01, May-19	B.1.1	6 taxa
UK693	6 (100.0%)	Mar-14, Mar-26	A.2	6 taxa
UK236	6 (100.0%)	Mar-12, Mar-27	B.2, B.2.1	6 taxa
UK1347	6 (100.0%)	May-18, Jun-04	B.1.1	6 taxa
UK491	6 (100.0%)	Mar-17, Mar-28	B.2.4, B.2	6 taxa
UK705	6 (100.0%)	Mar-16, Mar-26	B.1	6 taxa
UK360	6 (100.0%)	Mar-14, Apr-05	B.2, B.2.2	6 taxa
UK619	6 (100.0%)	Mar-02, Apr-28	B.1.1	6 taxa
UK1228	6 (100.0%)	Mar-20, Mar-28	B.1.1	6 taxa

Table S2 Raw data for figure two showing lags between the most recent sequence and current date for each sequencing centre

pillar2: —> 2 lag_df = pd.DataFrame(lag_dict) 3 print(lag_df.to_markdown()) 4 else: 5 print("Table S2 is not appropriate for this report and so has been omitted.") NameError: name 'lag_dict' is not defined

Table S3 Raw data for figure three showing the number of admin2 regions a lineage is present in over time

Week commencing	UK5	UK175	UK199	UK719	UK1205
2020-02-23	0	1	0	0	0
2020-03-01	1	3	0	0	1
2020-03-08	4	4	1	0	3
2020-03-15	4	6	4	0	0
2020-03-22	6	6	4	1	1
2020-03-29	3	2	2	1	0
2020-04-05	4	3	2	1	1
2020-04-12	4	1	1	0	0
2020-04-19	4	1	1	1	0
2020-04-26	2	2	2	0	0
2020-05-03	3	1	0	0	0
2020-05-10	1	0	1	0	0
2020-05-17	6	4	3	2	0
2020-05-24	4	3	0	0	0
2020-05-31	4	2	2	4	0
2020-06-07	5	1	1	1	0
2020-06-14	4	1	0	0	0
2020-06-21	2	2	0	1	0
2020-06-28	3	1	0	0	0
2020-07-05	2	0	1	0	0
2020-07-12	4	0	0	0	0
2020-07-19	4	2	0	0	0
2020-07-26	2	0	0	0	1
2020-08-02	1	0	1	0	1
2020-08-16	3	2	1	0	0
2020-08-23	2	0	0	1	0

Table S4 is not appropriate for this report and so has been omitted.

Table S5 Raw data for figure five showing when lineages started per day, divided by singletons and non-singletons

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-02-28	0	1	1
2020-03-01	0	3	3
2020-03-02	0	3	3
2020-03-03	0	2	2
2020-03-04	0	1	1
2020-03-05	1	3	4
2020-03-06	0	1	1
2020-03-08	0	1	1
2020-03-09	0	1	1
2020-03-10	0	4	4
2020-03-11	1	4	5
2020-03-12	3	6	9
2020-03-13	2	4	6
2020-03-14	0	5	5
2020-03-15	2	2	4
2020-03-16	4	4	8
2020-03-17	9	11	20
2020-03-18	2	5	7
2020-03-19	7	4	11
2020-03-20	6	6	12
2020-03-21	4	4	8
2020-03-22	3	0	3
2020-03-23	6	7	13
2020-03-24	7	3	10
2020-03-25	4	5	9
2020-03-26	11	5	16
2020-03-27	10	5	15
2020-03-28	14	5	19
2020-03-29	5	0	5
2020-03-30	9	4	13
2020-03-31	5	3	8
2020-04-01	5	4	9
2020-04-02	3	0	3
2020-04-03	8	2	10
2020-04-04	6	3	9
2020-04-05	7	2	9
2020-04-06	8	0	8
2020-04-07	2	3	5
2020-04-08	9	1	10
2020-04-09	6	4	10
2020-04-10	3	1	4
2020-04-11	5	0	5
2020-04-12	2	0	2
2020-04-13	2	0	2
2020-04-14	0	2	2
2020-04-15	4	1	5
2020-04-16	0	1	1
2020-04-17	3	0	3
2020-04-18	0	1	1
2020-04-20	3	0	3
2020-04-21	2	0	2
2020-04-22	0	2	2
2020-04-23	1	0	1
2020-04-24	0	1	1
2020-04-25	3	0	3
2020-04-27	1	0	1
2020-04-28	2	0	2

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-04-29	2	0	2
2020-04-30	0	1	1
2020-05-01	1	1	2
2020-05-02	0	1	1
2020-05-05	1	0	1
2020-05-17	6	3	9
2020-05-18	4	4	8
2020-05-19	8	4	12
2020-05-20	3	2	5
2020-05-25	1	0	1
2020-05-27	1	0	1
2020-05-28	1	0	1
2020-06-01	0	1	1
2020-06-03	2	0	2
2020-06-05	1	0	1
2020-06-06	1	0	1
2020-06-07	0	1	1
2020-06-08	0	1	1
2020-06-09	3	1	4
2020-06-11	0	1	1
2020-06-14	1	0	1
2020-06-16	2	0	2
2020-06-17	3	0	3
2020-06-22	1	0	1
2020-07-04	0	1	1
2020-07-07	1	0	1
2020-07-10	0	1	1
2020-07-20	1	0	1
2020-07-23	2	0	2
2020-07-24	0	1	1
2020-07-25	3	0	3
2020-08-01	0	2	2
2020-08-02	2	0	2
2020-08-03	0	1	1
2020-08-04	2	0	2
2020-08-20	0	1	1
2020-08-23	1	0	1
2020-08-31	1	0	1

Table S6 Raw data for figure six showing the number of sequences taken over time.

Day	England
2020-02-28	1
2020-03-01	7
2020-03-02	6
2020-03-03	6
2020-03-04	2
2020-03-05	9
2020-03-06	2
2020-03-07	4
2020-03-08	3
2020-03-09	6
2020-03-10	17
2020-03-11	14
2020-03-12	31
2020-03-13	27
2020-03-14	19
2020-03-15	20

Day	England
2020-03-16	35
2020-03-17	49
2020-03-18	52
2020-03-19	70
2020-03-20	51
2020-03-21	78
2020-03-22	15
2020-03-23	102
2020-03-24	122
2020-03-25	99
2020-03-26	193
2020-03-27	77
2020-03-28	86
2020-03-29	43
2020-03-30	52
2020-03-31	45
2020-04-01	43
2020-04-02	47
2020-04-03	72
2020-04-04	48
2020-04-05	57
2020-04-06	35
2020-04-07	41
2020-04-08	57
2020-04-09	63
2020-04-10	42
2020-04-11	30
2020-04-12	31
2020-04-13	29
2020-04-14	25
2020-04-15	34
2020-04-16	25
2020-04-17	28
2020-04-18	12
2020-04-19	14
2020-04-20	23
2020-04-21	27
2020-04-22	21
2020-04-23	9
2020-04-24	12
2020-04-25	12
2020-04-26	9
2020-04-27	13
2020-04-28	22
2020-04-29	18
2020-04-30	7
2020-05-01	14
2020-05-02	7
2020-05-03	4
2020-05-04	5
2020-05-05	3
2020-05-06	2
2020-05-07	4
2020-05-09	2
2020-05-10	1
2020-05-11	2
2020-05-12	2
2020-05-13	2
2020-05-15	1

Day	England
2020-05-16	1
2020-05-17	30
2020-05-18	45
2020-05-19	76
2020-05-20	52
2020-05-21	11
2020-05-22	1
2020-05-23	2
2020-05-24	2
2020-05-25	21
2020-05-26	20
2020-05-27	7
2020-05-28	4
2020-05-30	1
2020-05-31	10
2020-06-01	23
2020-06-02	29
2020-06-03	26
2020-06-04	26
2020-06-05	14
2020-06-06	10
2020-06-07	11
2020-06-08	19
2020-06-09	21
2020-06-10	20
2020-06-11	13
2020-06-12	1
2020-06-13	1
2020-06-14	1
2020-06-15	15
2020-06-16	20
2020-06-17	28
2020-06-18	6
2020-06-19	4
2020-06-20	3
2020-06-21	1
2020-06-22	5
2020-06-23	2
2020-06-24	1
2020-06-25	2
2020-06-26	5
2020-06-27	3
2020-06-28	1
2020-06-29	3
2020-06-30	4
2020-07-01	1
2020-07-03	2
2020-07-04	3
2020-07-07	1
2020-07-08	2
2020-07-10	1
2020-07-11	7
2020-07-13	4
2020-07-14	1
2020-07-15	4
2020-07-17	1
2020-07-18	1
2020-07-20	3
2020-07-21	3

Day	England
2020-07-22	7
2020-07-23	16
2020-07-24	8
2020-07-25	8
2020-07-26	4
2020-07-31	5
2020-08-01	5
2020-08-02	7
2020-08-03	9
2020-08-04	9
2020-08-16	1
2020-08-17	3
2020-08-20	6
2020-08-21	4
2020-08-22	4
2020-08-23	1
2020-08-24	1
2020-08-28	2
2020-08-29	2
2020-08-31	1

Table S7 Raw data for the figure seven with the number of sequences assigned to each admin2 region.

Admin2	Country	Number of sequences	Sequence group
BEDFORDSHIRE	England	1	1-10
CHESHIRE	England	294	200-300
CLWYD	Wales	12	10-100
CUMBRIA	England	156	100-200
DERBYSHIRE	England	19	10-100
GREATER LONDON	England	3	1-10
GWYNEDD	Wales	3	1-10
HEREFORDSHIRE	England	1	1-10
HERTFORDSHIRE	England	1	1-10
KENT	England	1	1-10
LANCASHIRE	England	506	500-600
LEICESTERSHIRE	England	1	1-10
MERSEYSIDE	England	1263	1000-2000
NORTHAMPTONSHIRE	England	1	1-10
SOUTH YORKSHIRE	England	1	1-10
TYNE AND WEAR	England	1	1-10
WEST YORKSHIRE	England	2	1-10