Lineages report for PHEC

This report gives summaries of UK specific lineages sequenced by PHEC for week 2020-09-13. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-08-24. The analysis (eg time since last sample) is therefore undertaken from this date. 4570 sequences in the UK from the sequencing centre PHEC 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 59 and the maximum is 3131

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

466 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 2 that remain: 1 is pending extinction ie last seen three weeks ago. 1 has gone quiet, ie hasn't been seen this week.

The following table contains information about the ten largest lineages 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	787 (100.0%)	Feb-16, Aug-12	B.1.1.10, B.1.1	787 taxa
UK107	698 (100.0%)	Feb-09, May-11	B.2.1, B.2, B	698 taxa
UK175	231 (100.0%)	Feb-24, Jun-09	B.1.5, B.1.76, B.1, B.1.105	231 taxa
UK5676	132 (100.0%)	Feb-26, May-11	B.2.1, B.2, B.2.9	132 taxa
UK2916	126 (100.0%)	Feb-03, Apr-30	B.1	126 taxa
UK72	103 (100.0%)	Feb-05, Apr-11	B.2.1, B	103 taxa
UK9	90 (100.0%)	Mar-09, May-15	B.1.13, B.1	90 taxa
UK600	80 (100.0%)	Feb-29, May-06	B.1.1	80 taxa
UK2913	80 (100.0%)	Mar-07, Apr-20	B.1.5, B.1, B.1.11	80 taxa
UK917	53 (100.0%)	Mar-18, May-11	B.1.1	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.

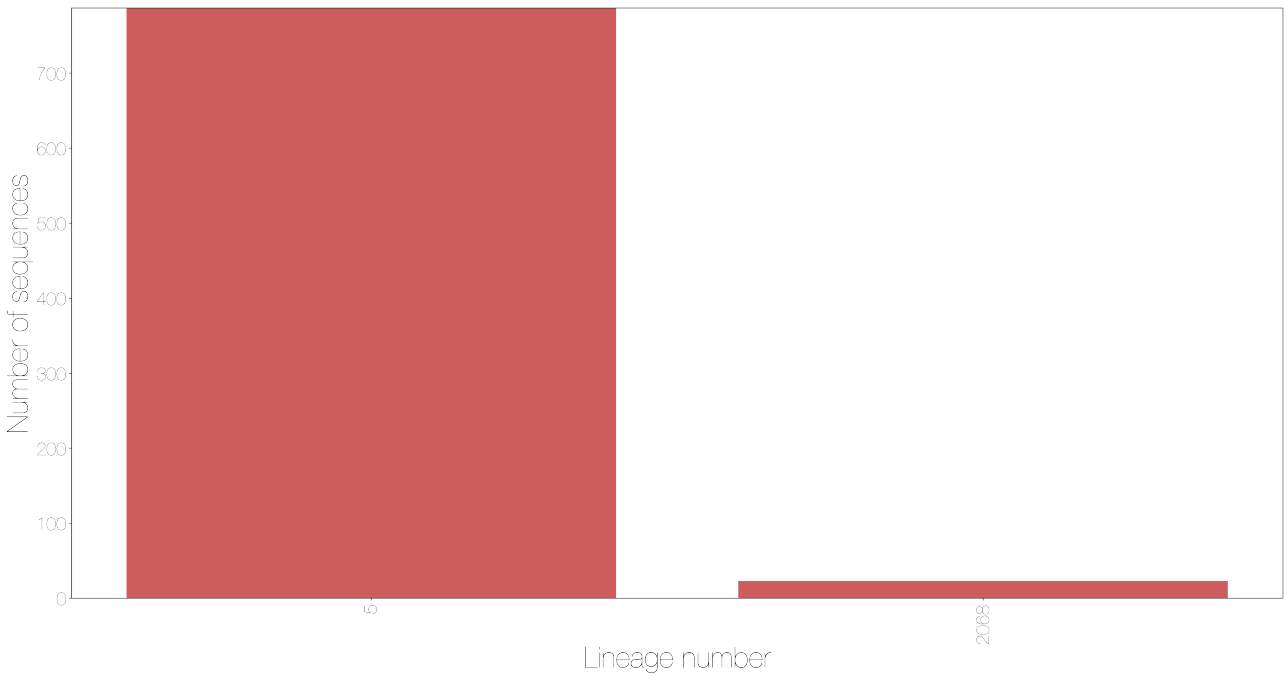

Figure 1: Number of sequences sampled in a lineage by country

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.

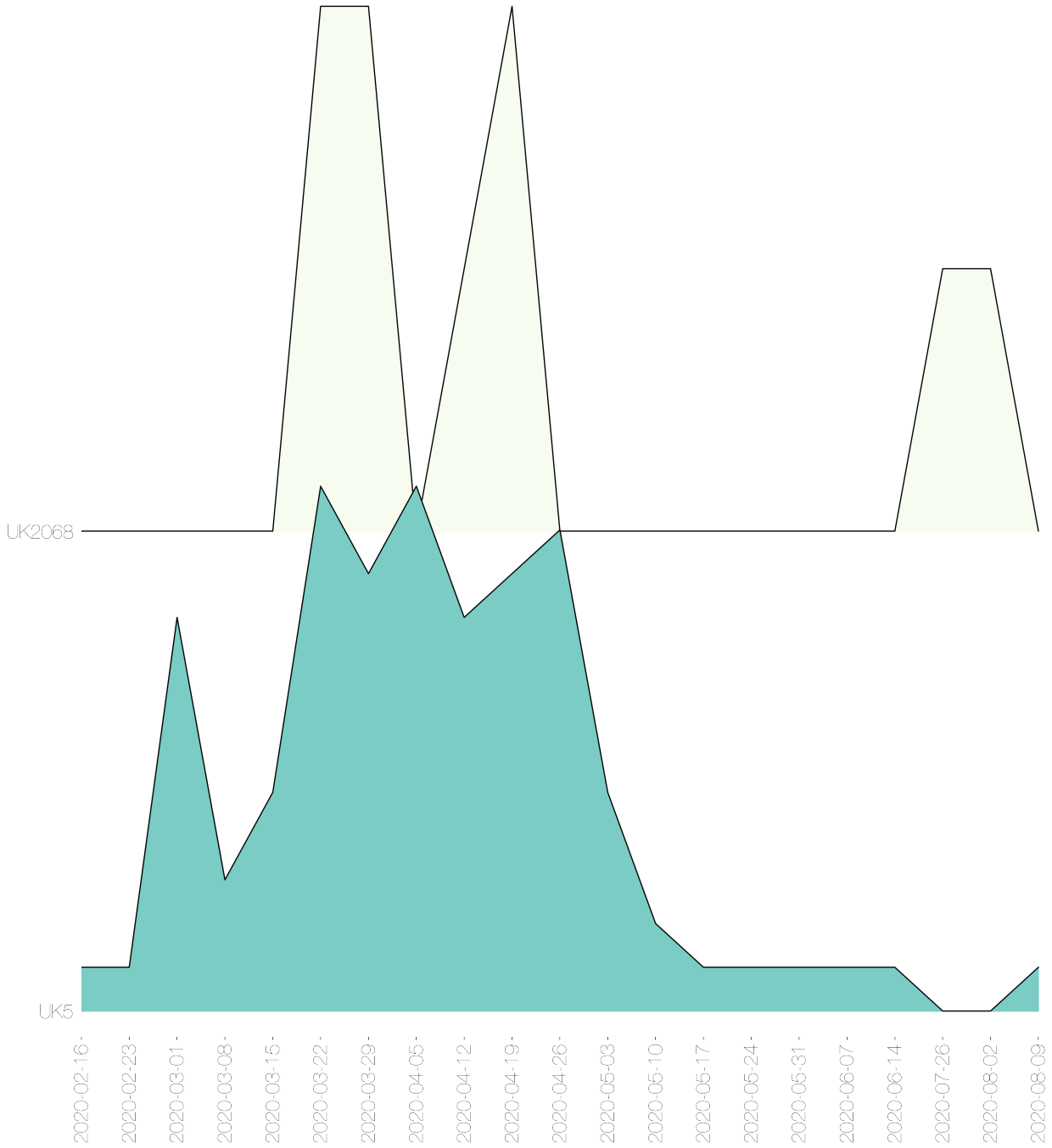

Figure 2: Lineages by number of adm2 regions present by epiweek

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.

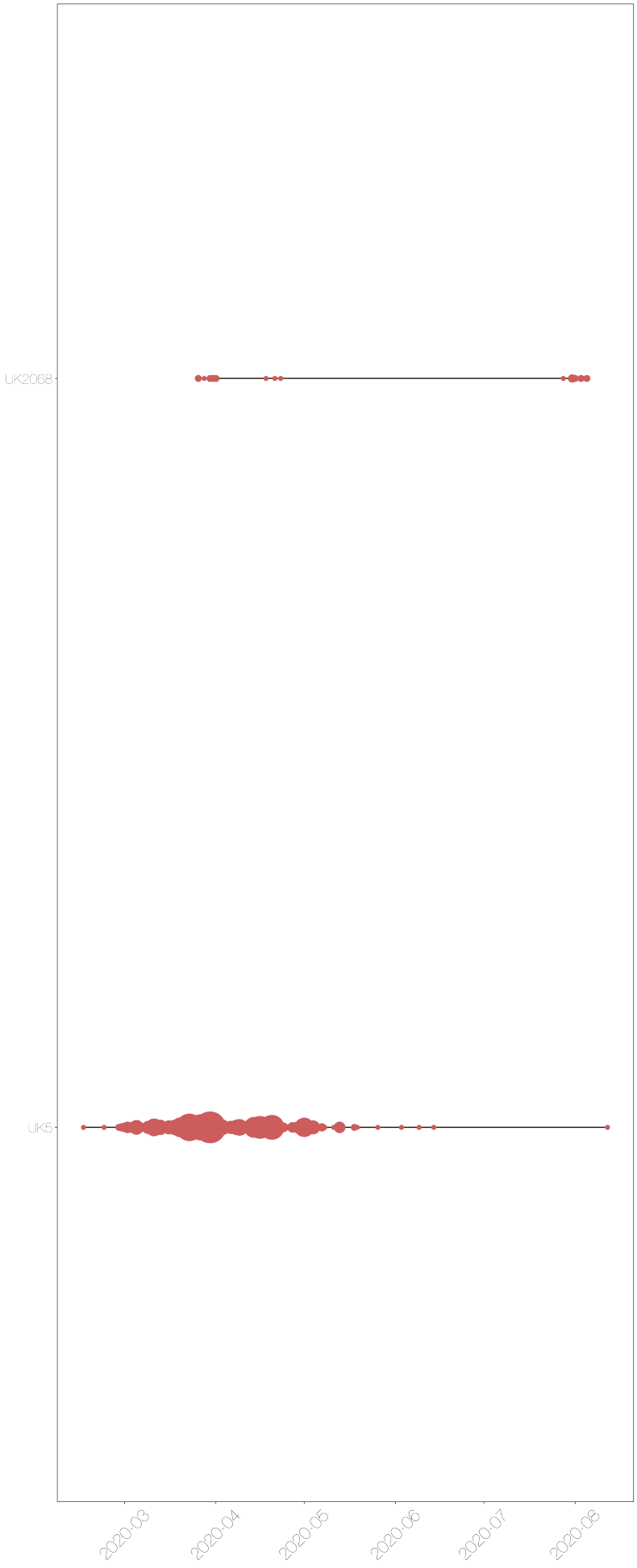

Figure 3: Timeline of lineages, sized by number of sequences from each country.

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

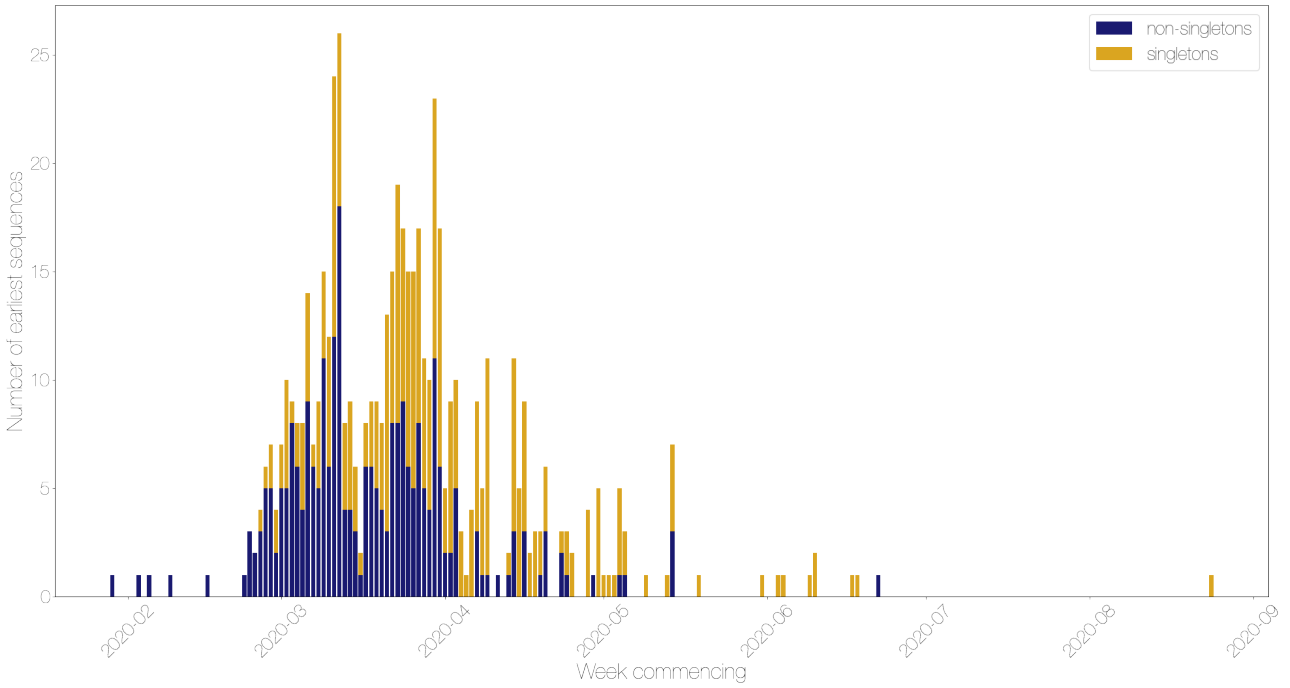

Figure 4: Lineage starts per week, split by singletons and non-singletons

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.

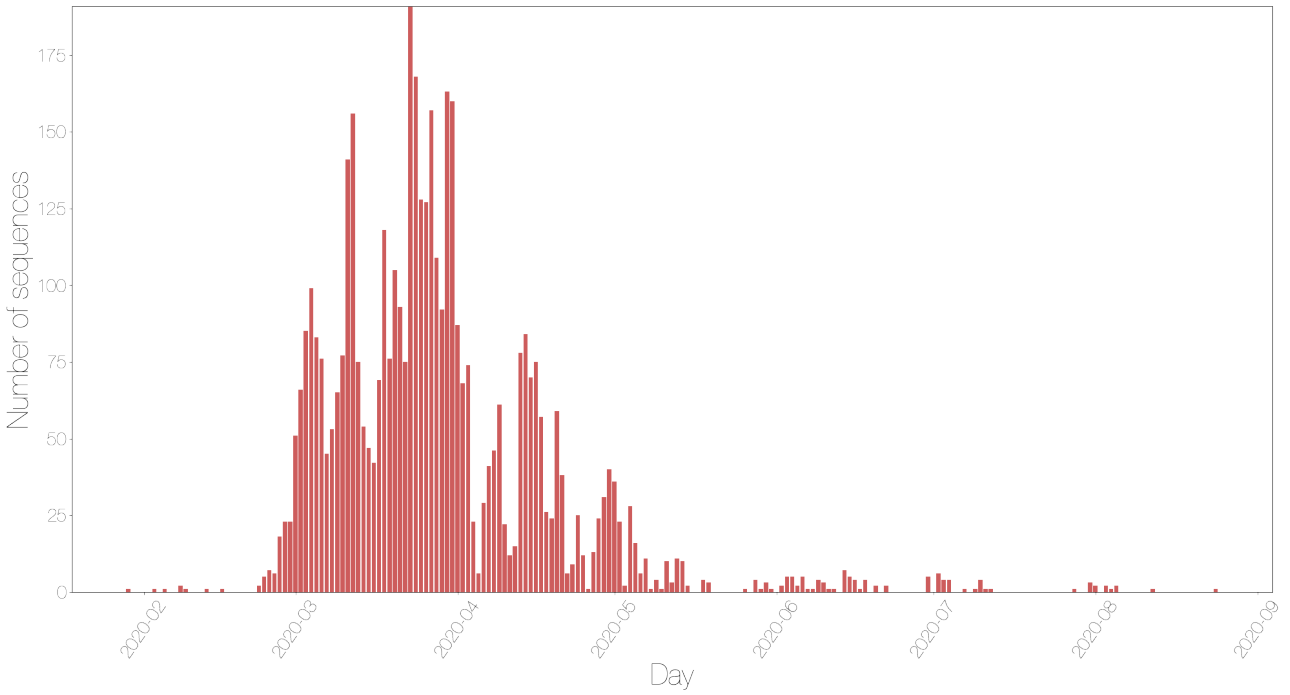

Figure 5: Sequences taken on each day by country

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 1100 sequences without enough geographical information to
map from this centre.
```

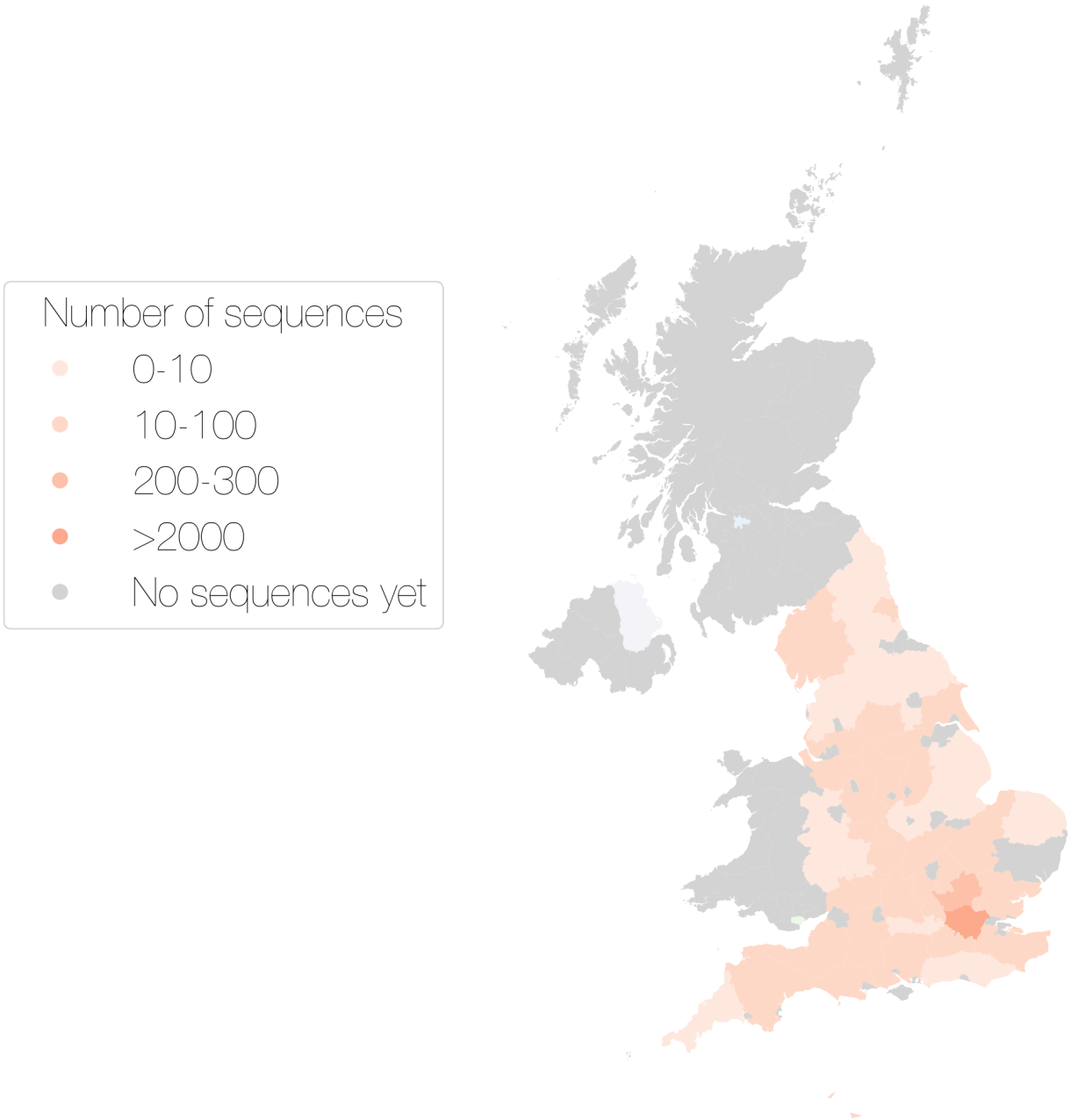

Figure 6: Map showing the number of sequences sampled by adm2 region

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	787 (100.0%)	Feb-16, Aug-12	B.1.1.10, B.1.1	787 taxa
UK107	698 (100.0%)	Feb-09, May-11	B.2.1, B.2, B	698 taxa
UK175	231 (100.0%)	Feb-24, Jun-09	B.1.5, B.1.76, B.1, B.1.105	231 taxa
UK5676	132 (100.0%)	Feb-26, May-11	B.2.1, B.2, B.2.9	132 taxa
UK2916	126 (100.0%)	Feb-03, Apr-30	B.1	126 taxa
UK72	103 (100.0%)	Feb-05, Apr-11	B.2.1, B	103 taxa
UK9	90 (100.0%)	Mar-09, May-15	B.1.13, B.1	90 taxa
UK600	80 (100.0%)	Feb-29, May-06	B.1.1	80 taxa
UK2913	80 (100.0%)	Mar-07, Apr-20	B.1.5, B.1, B.1.11	80 taxa
UK917	53 (100.0%)	Mar-18, May-11	B.1.1	53 taxa
UK2464	48 (100.0%)	Mar-09, May-04	B.1	48 taxa
UK120	43 (100.0%)	Feb-27, Mar-27	B	43 taxa
UK829	43 (100.0%)	Mar-03, Apr-29	B.2.1, B.2	43 taxa
UK5741	41 (100.0%)	Mar-12, May-11	B.1	41 taxa
UK1951	41 (100.0%)	Mar-09, May-09	B.1.1.1	41 taxa
UK199	39 (100.0%)	Feb-26, Jun-18	B.1.5, B.1	39 taxa
UK4	38 (100.0%)	Feb-28, Mar-31	B.2, B	38 taxa
UK491	37 (100.0%)	Mar-11, May-19	B.2.1, B.2, B.2.4, B	37 taxa
UK215	34 (100.0%)	Mar-12, Apr-19	B.2.1	34 taxa
UK945	33 (100.0%)	Mar-18, May-29	B.1.1	33 taxa
UK1205	31 (100.0%)	Mar-03, Jul-04	B.1.1.1	31 taxa
UK315	30 (100.0%)	Feb-25, Apr-19	B.2.1, B.2.2	30 taxa
UK5498	29 (100.0%)	Mar-06, Jun-20	B.2	29 taxa
UK59	27 (100.0%)	Feb-23, Apr-07	B.3	27 taxa
UK1274	27 (100.0%)	Jun-22, Jul-12	B.1.1	27 taxa
UK384	26 (100.0%)	Feb-28, Apr-20	B.2.1	26 taxa
UK1157	25 (100.0%)	Mar-07, Apr-25	B.1.1.7, B.1.1	25 taxa
UK2200	25 (100.0%)	Feb-28, Jun-20	B.1.5, B.1.35, B.1	25 taxa
UK1060	24 (100.0%)	Mar-03, May-11	B.1.1	24 taxa
UK55	23 (100.0%)	Mar-10, Apr-17	B.3	23 taxa
UK2068	23 (100.0%)	Mar-26, Aug-05	B.1.1.4, B.1.1	23 taxa
UK619	23 (100.0%)	Mar-02, Apr-15	B.1.1	23 taxa
UK1158	23 (100.0%)	Mar-11, Apr-10	B.1.1.7, B.1.1	23 taxa
UK515	22 (100.0%)	Mar-01, Apr-12	B.1	22 taxa
UK275	21 (100.0%)	Mar-09, Apr-21	B.1.13, B.1	21 taxa
UK902	20 (100.0%)	Mar-11, Apr-30	B.1.1	20 taxa
UK1126	20 (100.0%)	Mar-11, May-01	B.1.1	20 taxa
UK167	19 (100.0%)	Mar-12, Apr-07	B.1, B.1.117	19 taxa
UK402	19 (100.0%)	Mar-01, Mar-26	B.1	19 taxa
UK267	18 (100.0%)	Feb-25, May-05	B.2.1, B.2, B	18 taxa
UK31	18 (100.0%)	Mar-12, Apr-19	B.3	18 taxa
UK37	17 (100.0%)	Mar-18, Apr-02	B.1, B.1.30	17 taxa
UK360	17 (100.0%)	Mar-04, May-18	B.2.1, B.2.2	17 taxa
UK678	16 (100.0%)	Mar-12, Apr-09	B.1.1	16 taxa
UK1683	16 (100.0%)	Mar-18, Apr-28	B.1.1.1	16 taxa
UK12	16 (100.0%)	Mar-11, Apr-15	B.1, B.1.11, B.1.88	16 taxa
UK241	16 (100.0%)	Mar-25, Apr-07	B.1.5.3, B.1.5	16 taxa
UK274	15 (100.0%)	Mar-06, Apr-02	B.3	15 taxa
UK46	15 (100.0%)	Mar-02, Apr-29	B.2.1, B.2	15 taxa
UK119	15 (100.0%)	Mar-11, Apr-16	B.2, B	15 taxa
UK2906	14 (100.0%)	Mar-06, Apr-08	B.1	14 taxa
UK1535	13 (100.0%)	Mar-02, May-01	B.1.1.3, B.1.1	13 taxa
UK899	13 (100.0%)	Mar-12, Apr-29	B.1, B.1.11	13 taxa
UK740	13 (100.0%)	Mar-28, Apr-28	B.1.1	13 taxa
UK1282	13 (100.0%)	Mar-15, Mar-31	B.1.1	13 taxa
UK34	12 (100.0%)	Feb-27, Mar-06	B.2.1, B.4	12 taxa
UK1153	12 (100.0%)	Mar-24, May-04	B.1.1	12 taxa
UK387	12 (100.0%)	Mar-09, May-18	B.1, B.1.11	12 taxa
UK968	12 (100.0%)	Apr-14, Jun-10	B.1.1	12 taxa

Lineage name	England	Date range	Global lineage	Total
UK494	12 (100.0%)	Mar-26, Apr-28	B.1.5, B.1.104	12 taxa
UK1263	11 (100.0%)	Mar-09, Apr-14	B.1.1	11 taxa
UK893	11 (100.0%)	Mar-12, Apr-17	B.1.5, B.1	11 taxa
UK788	10 (100.0%)	Feb-28, Mar-05	B.4	10 taxa
UK699	10 (100.0%)	Mar-12, Apr-15	B.1.5.9	10 taxa
UK1230	9 (100.0%)	Mar-24, May-02	B.1.1	9 taxa
UK558	9 (100.0%)	Feb-29, Mar-07	B.2	9 taxa
UK874	8 (100.0%)	Feb-27, Mar-05	B.1.1	8 taxa
UK1488	8 (100.0%)	Mar-30, Apr-20	B.1.5	8 taxa
UK231	8 (100.0%)	Apr-24, May-13	B.1	8 taxa
UK739	8 (100.0%)	Mar-01, Mar-08	B.4	8 taxa
UK1037	8 (100.0%)	Mar-22, May-01	B.1.1	8 taxa
UK597	8 (100.0%)	Apr-08, Jun-03	B.1	8 taxa
UK1278	8 (100.0%)	Apr-03, May-12	B.1.1	8 taxa
UK799	7 (100.0%)	Mar-01, Mar-07	B.1	7 taxa
UK6	7 (100.0%)	Mar-06, Apr-20	B.1.75, B.1	7 taxa
UK22	7 (100.0%)	Mar-02, Mar-18	B	7 taxa
UK1141	7 (100.0%)	Mar-14, Apr-10	B.1.1	7 taxa
UK501	7 (100.0%)	Mar-07, Mar-31	B.1	7 taxa
UK561	7 (100.0%)	Mar-01, Mar-30	B.1	7 taxa
UK1266	7 (100.0%)	Mar-24, May-06	B.1.1	7 taxa
UK1603	6 (100.0%)	Mar-30, Apr-19	B.1.1	6 taxa
UK334	6 (100.0%)	Mar-13, Apr-09	B.3	6 taxa
UK1155	6 (100.0%)	Mar-05, Apr-14	B.1.1	6 taxa
UK797	6 (100.0%)	Mar-25, Apr-08	B.1.11	6 taxa
UK629	6 (100.0%)	Mar-25, Apr-02	B.1	6 taxa
UK654	6 (100.0%)	Feb-27, Mar-08	B.2	6 taxa
UK131	6 (100.0%)	Mar-11, Apr-08	B, B.15	6 taxa
UK134	6 (100.0%)	Mar-04, Apr-13	B.1.5, B.1	6 taxa
UK1215	6 (100.0%)	Mar-27, Apr-28	B.1.1.1, B.1.1	6 taxa
UK1120	6 (100.0%)	Apr-13, Apr-14	B.1.1	6 taxa
UK38	6 (100.0%)	Mar-04, Apr-20	B.2.1	6 taxa
UK743	6 (100.0%)	Feb-24, Jun-14	B.1.102	6 taxa
UK499	6 (100.0%)	Mar-05, Mar-31	B.2.1, B.2.6	6 taxa
UK786	6 (100.0%)	Mar-07, Apr-17	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

NameError Traceback (most recent call last) in 1 if not 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	UK2068
2020-02-16	1	0
2020-02-23	1	0
2020-03-01	9	0
2020-03-08	3	0
2020-03-15	5	0
2020-03-22	12	2
2020-03-29	10	2
2020-04-05	12	0
2020-04-12	9	1
2020-04-19	10	2
2020-04-26	11	0
2020-05-03	5	0
2020-05-10	2	0

Week commencing	UK5	UK2068
2020-05-17	1	0
2020-05-24	1	0
2020-05-31	1	0
2020-06-07	1	0
2020-06-14	1	0
2020-07-26	0	1
2020-08-02	0	1
2020-08-09	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
2020-01-29	0	1	1
2020-02-03	0	1	1
2020-02-05	0	1	1
2020-02-09	0	1	1
2020-02-16	0	1	1
2020-02-23	0	1	1
2020-02-24	0	3	3
2020-02-25	0	2	2
2020-02-26	1	3	4
2020-02-27	1	5	6
2020-02-28	2	5	7
2020-02-29	2	2	4
2020-03-01	2	5	7
2020-03-02	5	5	10
2020-03-03	1	8	9
2020-03-04	2	6	8
2020-03-05	4	4	8
2020-03-06	5	9	14
2020-03-07	1	6	7
2020-03-08	4	5	9
2020-03-09	4	11	15
2020-03-10	6	6	12
2020-03-11	12	12	24
2020-03-12	8	18	26
2020-03-13	4	4	8
2020-03-14	5	4	9
2020-03-15	3	3	6
2020-03-16	1	1	2
2020-03-17	2	6	8
2020-03-18	3	6	9
2020-03-19	4	5	9
2020-03-20	4	4	8
2020-03-21	10	3	13
2020-03-22	7	8	15
2020-03-23	11	8	19
2020-03-24	8	9	17
2020-03-25	9	6	15
2020-03-26	10	5	15
2020-03-27	9	8	17
2020-03-28	6	5	11
2020-03-29	6	4	10
2020-03-30	12	11	23
2020-03-31	11	6	17
2020-04-01	3	2	5

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

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

Day	England
2020-01-29	1
2020-02-03	1
2020-02-05	1
2020-02-08	2
2020-02-09	1
2020-02-13	1
2020-02-16	1
2020-02-23	2
2020-02-24	5
2020-02-25	7
2020-02-26	6
2020-02-27	18
2020-02-28	23

Day	England
2020-02-29	23
2020-03-01	51
2020-03-02	66
2020-03-03	85
2020-03-04	99
2020-03-05	83
2020-03-06	76
2020-03-07	45
2020-03-08	53
2020-03-09	65
2020-03-10	77
2020-03-11	141
2020-03-12	156
2020-03-13	75
2020-03-14	54
2020-03-15	47
2020-03-16	42
2020-03-17	69
2020-03-18	118
2020-03-19	76
2020-03-20	105
2020-03-21	93
2020-03-22	75
2020-03-23	191
2020-03-24	168
2020-03-25	128
2020-03-26	127
2020-03-27	157
2020-03-28	109
2020-03-29	92
2020-03-30	163
2020-03-31	160
2020-04-01	87
2020-04-02	68
2020-04-03	74
2020-04-04	23
2020-04-05	6
2020-04-06	29
2020-04-07	41
2020-04-08	46
2020-04-09	61
2020-04-10	22
2020-04-11	12
2020-04-12	15
2020-04-13	78
2020-04-14	84
2020-04-15	70
2020-04-16	75
2020-04-17	57
2020-04-18	26
2020-04-19	24
2020-04-20	59
2020-04-21	38
2020-04-22	6
2020-04-23	9
2020-04-24	25
2020-04-25	12
2020-04-26	1
2020-04-27	13

Day	England
2020-04-28	24
2020-04-29	31
2020-04-30	40
2020-05-01	36
2020-05-02	23
2020-05-03	2
2020-05-04	28
2020-05-05	16
2020-05-06	6
2020-05-07	11
2020-05-08	1
2020-05-09	4
2020-05-10	1
2020-05-11	10
2020-05-12	3
2020-05-13	11
2020-05-14	10
2020-05-15	2
2020-05-18	4
2020-05-19	3
2020-05-26	1
2020-05-28	4
2020-05-29	1
2020-05-30	3
2020-05-31	1
2020-06-02	2
2020-06-03	5
2020-06-04	5
2020-06-05	2
2020-06-06	5
2020-06-07	1
2020-06-08	1
2020-06-09	4
2020-06-10	3
2020-06-11	1
2020-06-12	1
2020-06-14	7
2020-06-15	5
2020-06-16	4
2020-06-17	1
2020-06-18	4
2020-06-20	2
2020-06-22	2
2020-06-30	5
2020-07-02	6
2020-07-03	4
2020-07-04	4
2020-07-07	1
2020-07-09	1
2020-07-10	4
2020-07-11	1
2020-07-12	1
2020-07-28	1
2020-07-31	3
2020-08-01	2
2020-08-03	2
2020-08-04	1
2020-08-05	2
2020-08-12	1

Day	England
2020-08-24	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
ANTRIM	Northern Ireland	1	1-10
BEDFORDSHIRE	England	13	10-100
BERKSHIRE	England	6	1-10
BRISTOL	England	18	10-100
BUCKINGHAMSHIRE	England	22	10-100
CAMBRIDGESHIRE	England	94	10-100
CARDIFF	Wales	1	1-10
CHESHIRE	England	15	10-100
CORNWALL	England	2	1-10
CUMBRIA	England	12	10-100
DERBYSHIRE	England	12	10-100
DEVON	England	22	10-100
DORSET	England	16	10-100
DURHAM	England	2	1-10
EAST RIDING OF YORKSHIRE	England	28	10-100
ESSEX	England	92	10-100
GLASGOW	Scotland	1	1-10
GLOUCESTERSHIRE	England	12	10-100
GREATER LONDON	England	2035	>2000
GREATER MANCHESTER	England	32	10-100
GUERNSEY	Channel_islands	43	10-100
HAMPSHIRE	England	42	10-100
HEREFORDSHIRE	England	1	1-10
HERTFORDSHIRE	England	281	200-300
JERSEY	Channel_islands	95	10-100
KENT	England	40	10-100
LANCASHIRE	England	8	1-10
LEICESTERSHIRE	England	7	1-10
LINCOLNSHIRE	England	5	1-10
MERSEYSIDE	England	63	10-100
NORFOLK	England	2	1-10
NORTH YORKSHIRE	England	5	1-10
NORTHAMPTONSHIRE	England	16	10-100
NORTHUMBERLAND	England	1	1-10
NOTTINGHAMSHIRE	England	10	10-100
OXFORDSHIRE	England	27	10-100
SHROPSHIRE	England	1	1-10
SOMERSET	England	83	10-100
SOUTH YORKSHIRE	England	46	10-100
STAFFORDSHIRE	England	34	10-100
SURREY	England	58	10-100
SUSSEX	England	8	1-10
TYNE AND WEAR	England	36	10-100
WARWICKSHIRE	England	12	10-100
WEST MIDLANDS	England	56	10-100
WEST YORKSHIRE	England	34	10-100
WILTSHIRE	England	13	10-100
WORCESTERSHIRE	England	7	1-10