Lineages report for PHEC

This report gives summaries of UK specific lineages sequenced by PHEC for week 2020-06-05. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-05-19. The analysis (eg time since last sample) is therefore undertaken from this date. 3918 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 2013 and the maximum is 2715

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

1889 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: 15 are pending extinction, ie last seen three weeks ago. 3 lineages have gone quiet, ie haven't been seen this week. 4 lineages have reactivated. 1 lineage has been continuously circulating.

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 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
UK2916	127 (100.0%)	Feb-03, Apr-30	127	B.1, B.1.p11	19	0.0363
UK5	100 (100.0%)	Mar-03, May-10	100	B.1.1.1	9	0.0763
UK9	84 (100.0%)	Mar-09, May-15	84	B.1.13	4	0.2018
UK107	68 (100.0%)	Mar-15, Apr-21	68	B.2.1, B.2, B.2.5	28	0.0197
UK2464	51 (100.0%)	Mar-09, May-04	51	B.1.p11	15	0.0747
UK77	48 (100.0%)	Mar-11, May-05	48	B.2, B.2.4	14	0.0836
UK4	42 (100.0%)	Feb-28, Mar-31	42	B	49	0.0159
UK2913	38 (100.0%)	Mar-10, Apr-15	38	B.1.p11	34	0.0286
UK339	34 (100.0%)	Feb-23, Apr-09	34	B.3	40	0.0348
UK63	32 (100.0%)	Mar-18, Apr-15	32	B.1.1	34	0.0266

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.

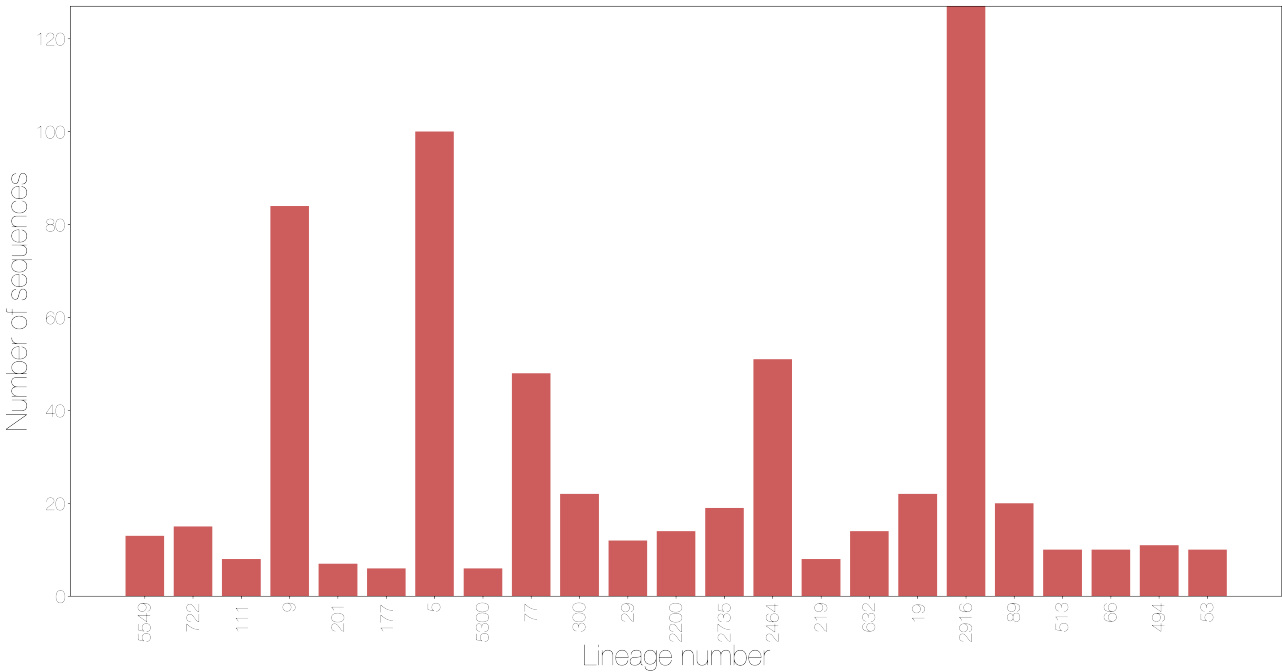The raw data for this bar chart are in the table above.

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 lag for this sequencing centre is 17 days`

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
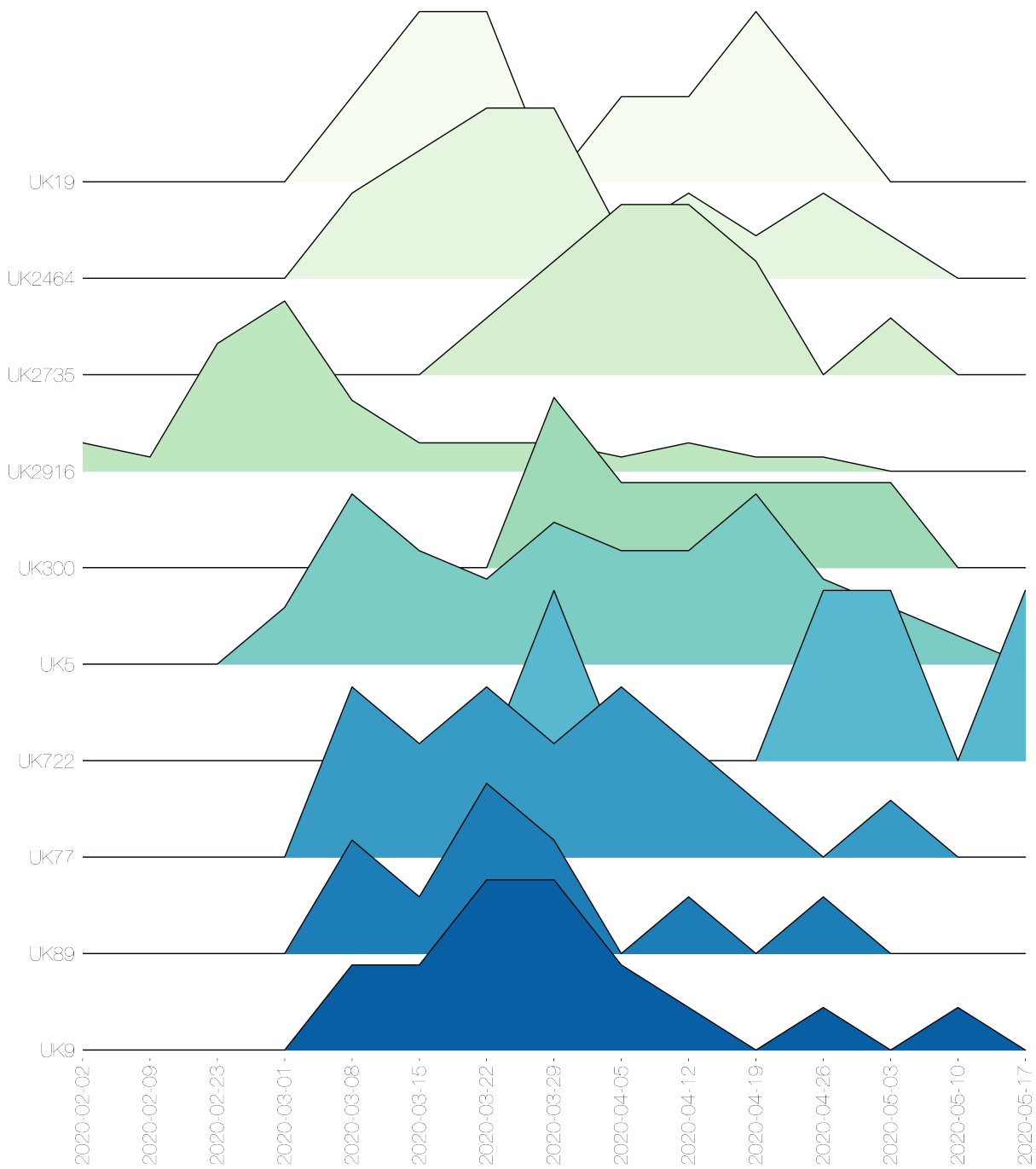
These lineages are shown on the timeline. Each line represents the length of the cluster, from oldest to most recent sampling date. The dots are sized by the number of sequences taken on that date, and again are colour coded by country. The raw data has been written to a summary file.
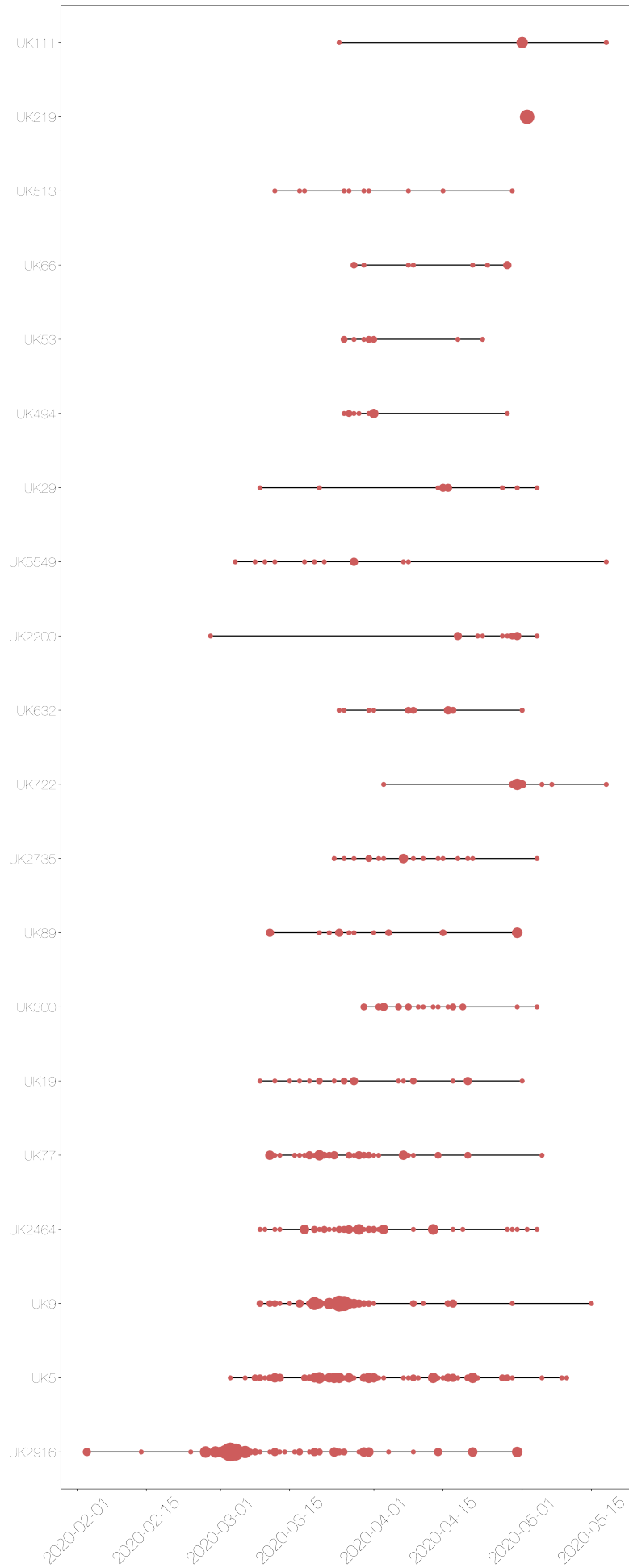
Figure 3: Timeline of lineages, sized by number of sequences from each country.

The date of first sequence in the cluster is shown in figure five for every cluster with date information.
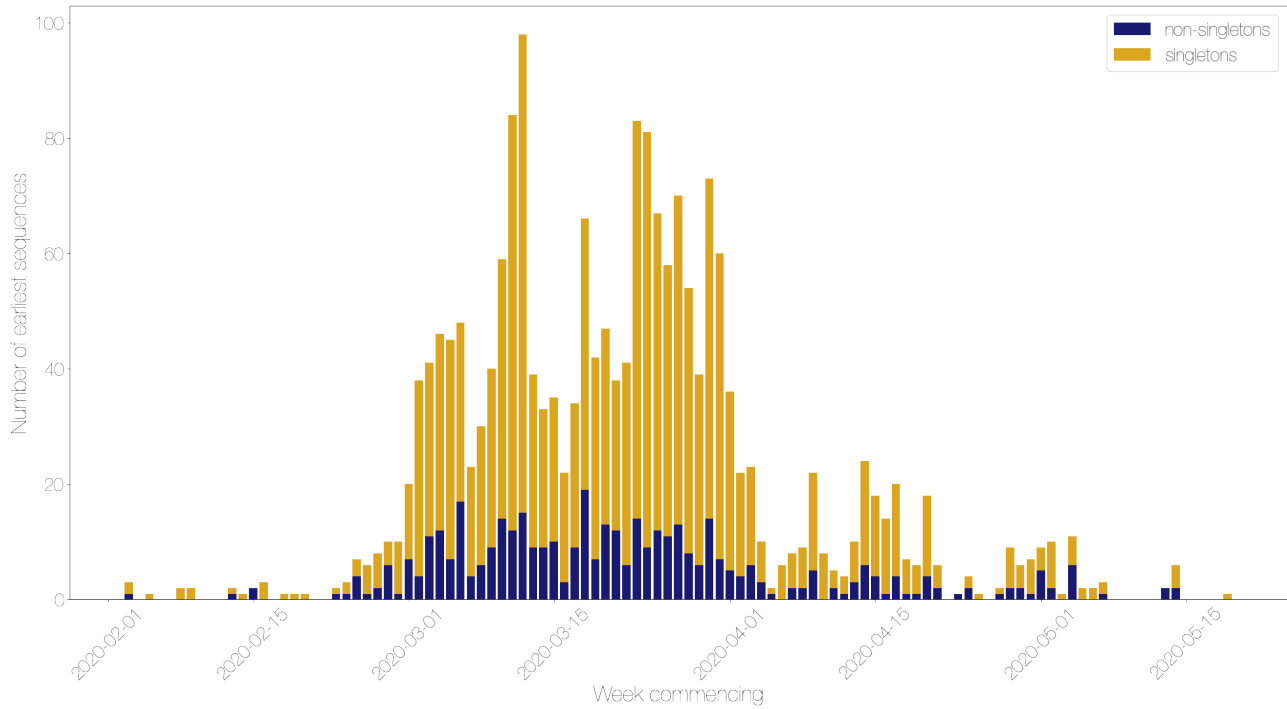
Figure 4: Lineage starts per week, split by singletons and non-singletons

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.
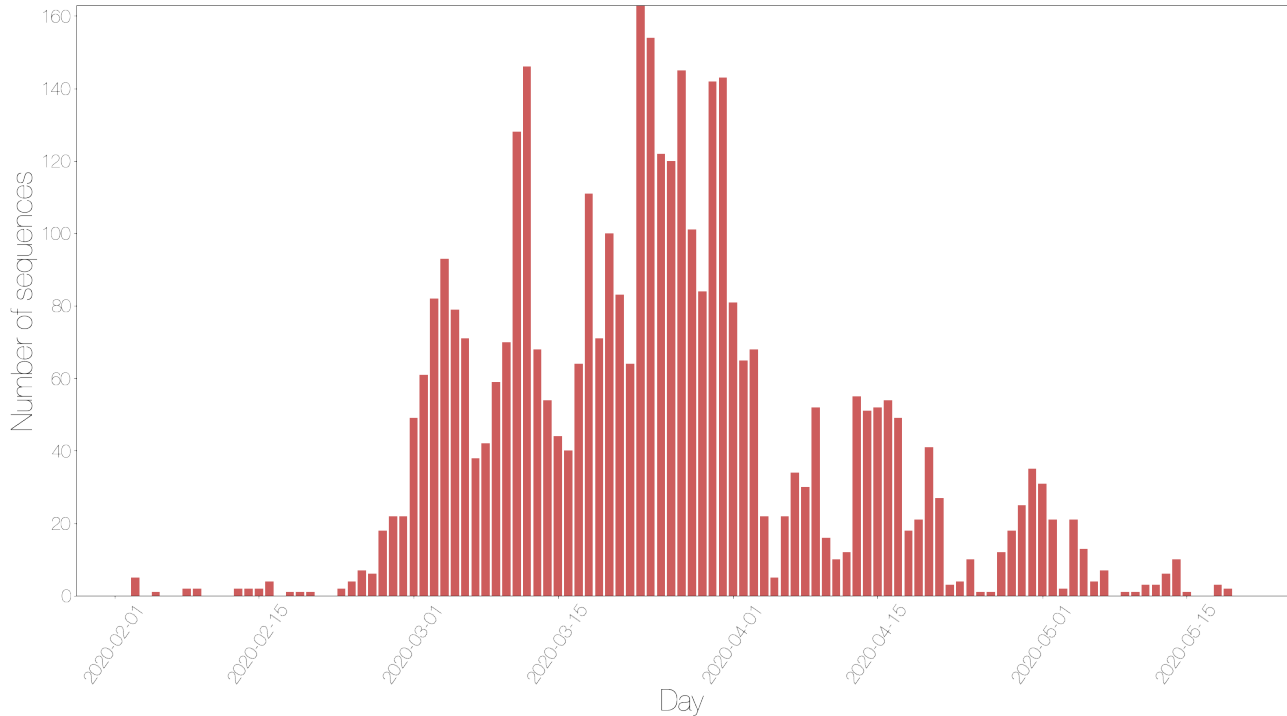
Figure 5: Sequences taken on each day by country

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 959 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.

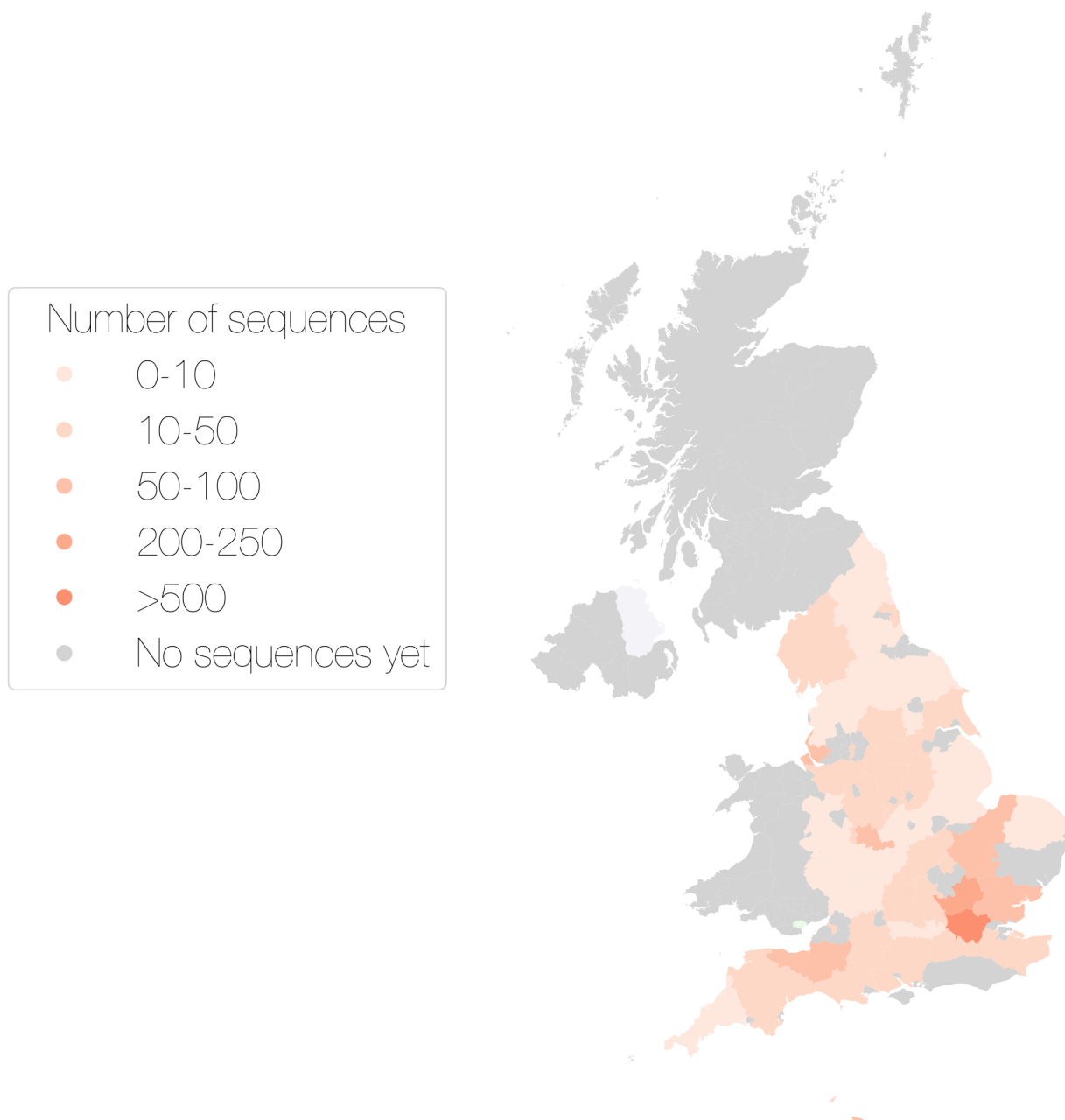| Lineage name	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK2916	127 (100.0%)	Feb-03, Apr-30	127	B.1, B.1.p11	19	0.0363
UK5	100 (100.0%)	Mar-03, May-10	100	B.1.1.1	9	0.0763
UK9	84 (100.0%)	Mar-09, May-15	84	B.1.13	4	0.2018
UK107	68 (100.0%)	Mar-15, Apr-21	68	B.2.1, B.2, B.2.5	28	0.0197
UK2464	51 (100.0%)	Mar-09, May-04	51	B.1.p11	15	0.0747
UK77	48 (100.0%)	Mar-11, May-05	48	B.2, B.2.4	14	0.0836
UK4	42 (100.0%)	Feb-28, Mar-31	42	B	49	0.0159
UK2913	38 (100.0%)	Mar-10, Apr-15	38	B.1.p11	34	0.0286
UK339	34 (100.0%)	Feb-23, Apr-09	34	B.3	40	0.0348
UK63	32 (100.0%)	Mar-18, Apr-15	32	B.1.1	34	0.0266
UK116	28 (100.0%)	Feb-25, Apr-01	28	B.2.1	48	0.0278
UK94	28 (100.0%)	Mar-12, Apr-19	28	B.2.1, B.2	30	0.0469
UK3	24 (100.0%)	Feb-24, Apr-05	24	B.1	44	0.0405
UK190	23 (100.0%)	Mar-01, Apr-12	23	B.1	37	0.0516
UK19	22 (100.0%)	Mar-09, May-01	22	B.1	18	0.1402
UK300	22 (100.0%)	Mar-30, May-04	22	B.1.1	15	0.1111
UK18	20 (100.0%)	Mar-11, Apr-10	20	B.1.1.7	39	0.0405
UK89	20 (100.0%)	Mar-11, Apr-30	20	B.1.1.9	19	0.1385
UK2735	19 (100.0%)	Mar-24, May-04	19	B.1.1	15	0.1519
UK37	17 (100.0%)	Mar-18, Apr-02	17	B.1.30, B.1	47	0.0199
UK241	16 (100.0%)	Mar-25, Apr-07	16	B.1.5.3	42	0.0206
UK117	15 (100.0%)	Feb-28, Mar-22	15	B.2.1	58	0.0283
UK274	15 (100.0%)	Mar-06, Apr-02	15	B, B.3	47	0.041
UK722	15 (100.0%)	Apr-03, May-18	15	B.1.1	1	3.2143
UK403	14 (100.0%)	Mar-23, Mar-31	14	B.1.1	49	0.0126
UK41	14 (100.0%)	Mar-01, Mar-26	14	B.1	54	0.0356

Lineage name	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK632	14 (100.0%)	Mar-25, May-01	14	B.1.1	18	0.1581
UK62	14 (100.0%)	Mar-12, Apr-19	14	B.3	30	0.0974
UK112	14 (100.0%)	Mar-15, Mar-31	14	B.1.1	49	0.0251
UK2200	14 (100.0%)	Feb-28, May-04	14	B.1.5, B.1.5.6	15	0.3385
UK12	13 (100.0%)	Mar-12, Apr-15	13	B.1.p11	34	0.0833
UK119	13 (100.0%)	Mar-11, Apr-16	13	B.2.5	33	0.0909
UK5549	13 (100.0%)	Mar-04, May-18	13	B.2.2	1	6.25
UK378	13 (100.0%)	Feb-15, Mar-05	13	B.1.1	75	0.0211
UK371	13 (100.0%)	Mar-12, Mar-30	13	B.1.1	50	0.03
UK34	13 (100.0%)	Feb-15, Apr-02	13	B.4	47	0.0833
UK29	12 (100.0%)	Mar-09, May-04	12	B.1.1	15	0.3394
UK694	12 (100.0%)	Mar-06, Mar-14	12	B	66	0.011
UK347	12 (100.0%)	Mar-13, Apr-02	12	B.1	47	0.0387
UK143	12 (100.0%)	Mar-14, Apr-16	12	B.2.1	33	0.0909
UK494	11 (100.0%)	Mar-26, Apr-28	11	B.1.p11	21	0.1571
UK1018	11 (100.0%)	Apr-20, Apr-21	11	B.1.1	28	0.0036
UK53	10 (100.0%)	Mar-26, Apr-23	10	B.1.1.4	26	0.1197
UK5715	10 (100.0%)	Feb-13, Mar-07	10	B.2	73	0.035
UK604	10 (100.0%)	Mar-09, Mar-12	10	B.1.1	68	0.0049
UK428	10 (100.0%)	Mar-20, Apr-01	10	B.2.1, B.2	48	0.0278
UK291	10 (100.0%)	Mar-13, Apr-03	10	B.2.1	46	0.0507
UK5180	10 (100.0%)	Apr-12, Apr-19	10	B.1.1.7	30	0.0259
UK3021	10 (100.0%)	Mar-12, Apr-15	10	B.1	34	0.1111
UK513	10 (100.0%)	Mar-12, Apr-29	10	B.1.p11	20	0.2667
UK66	10 (100.0%)	Mar-28, Apr-28	10	B.1.1.8	21	0.164
UK687	9 (100.0%)	Feb-28, Mar-08	9	B.2.1, B.2	72	0.0156
UK64	9 (100.0%)	Mar-12, Apr-17	9	B.1	32	0.1406
UK5741	9 (100.0%)	Mar-01, Apr-17	9	B.2, B.1	32	0.1836
UK46	8 (100.0%)	Mar-02, Mar-24	8	B.2.1	56	0.0561

Lineage name	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK739	8 (100.0%)	Mar-01, Mar-08	8	B.4	72	0.0139
UK756	8 (100.0%)	Feb-27, Mar-05	8	B.1.1	75	0.0133
UK195	8 (100.0%)	Mar-29, Apr-18	8	B.1.1	31	0.0922
UK788	8 (100.0%)	Feb-28, Mar-05	8	B.4	75	0.0114
UK1013	8 (100.0%)	Apr-15, Apr-16	8	B.1.1	33	0.0043
UK242	8 (100.0%)	Mar-30, Apr-20	8	B.1.5	29	0.1034
UK111	8 (100.0%)	Mar-25, May-18	8	B.1.1	1	7.7143
UK689	8 (100.0%)	Mar-05, Mar-28	8	B.2.1, B.2	52	0.0632
UK219	8 (100.0%)	May-02, May-02	8	B.1.1	17	0
UK2013	7 (100.0%)	Mar-15, Apr-21	7	B.1	28	0.2202
UK38	7 (100.0%)	Mar-04, Apr-20	7	B.2.1	29	0.2701
UK22	7 (100.0%)	Mar-02, Mar-18	7	B	62	0.043
UK8	7 (100.0%)	Mar-03, Mar-12	7	B	68	0.0221
UK733	7 (100.0%)	Mar-10, Mar-18	7	B.2.1	62	0.0215
UK201	7 (100.0%)	Apr-24, May-13	7	B.1	6	0.5278
UK177	6 (100.0%)	Apr-03, May-12	6	B.1.1	7	1.1143
UK5300	6 (100.0%)	May-04, May-06	6	B.1.1	13	0.0308
UK131	6 (100.0%)	Mar-11, Apr-08	6	B.15	41	0.1366
UK4237	6 (100.0%)	Apr-15, Apr-15	6	B.1.1	34	0
UK223	6 (100.0%)	Mar-10, Mar-27	6	B.2.1	53	0.0642
UK799	6 (100.0%)	Mar-01, Mar-07	6	B.1	73	0.0164
UK103	6 (100.0%)	Mar-20, Apr-19	6	B.1.1	30	0.2
UK171	6 (100.0%)	Mar-13, Mar-27	6	B.2.1, B.2	53	0.0528
UK5780	6 (100.0%)	Mar-14, Mar-29	6	B.2.1, B.2	51	0.0588
UK654	6 (100.0%)	Feb-27, Mar-08	6	B.2.5	72	0.0278
UK178	6 (100.0%)	Mar-14, Apr-04	6	B.1.1	45	0.0933
UK335	6 (100.0%)	Mar-25, Mar-31	6	B.2.1	49	0.0245
UK289	6 (100.0%)	Mar-25, Apr-16	6	B.2.1	33	0.1333
UK317	6 (100.0%)	Mar-26, Apr-16	6	B.3	33	0.1273

Lineage name	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK857	6 (100.0%)	Mar-24, Mar-29	6	B.2.1	51	0.0196

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

	Centre	Lag in days
0	PHEC	17

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

Week commencing	UK2916	UK5	UK9	UK2464	UK77	UK19	UK300	UK89	UK2735	UK722
2020-02-02	2	0	0	0	0	0	0	0	0	0
2020-02-09	1	0	0	0	0	0	0	0	0	0
2020-02-23	9	0	0	0	0	0	0	0	0	0
2020-03-01	12	2	0	0	0	0	0	0	0	0
2020-03-08	5	6	2	2	3	1	0	2	0	0
2020-03-15	2	4	2	3	2	2	0	1	0	0
2020-03-22	2	3	4	4	3	2	0	3	1	0
2020-03-29	2	5	4	4	2	0	2	2	2	1
2020-04-05	1	4	2	1	3	1	1	0	3	0
2020-04-12	2	4	1	2	2	1	1	1	3	0
2020-04-19	1	6	0	1	1	2	1	0	2	0
2020-04-26	1	3	1	2	0	1	1	1	0	1
2020-05-03	0	2	0	1	1	0	1	0	1	1
2020-05-10	0	1	1	0	0	0	0	0	0	0
2020-05-17	0	0	0	0	0	0	0	0	0	1

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
2020-02-03	2	1	3
2020-02-05	1	0	1
2020-02-08	2	0	2
2020-02-09	2	0	2
2020-02-13	1	1	2
2020-02-14	1	0	1
2020-02-15	0	2	2
2020-02-16	3	0	3
2020-02-18	1	0	1
2020-02-19	1	0	1
2020-02-20	1	0	1
2020-02-23	1	1	2
2020-02-24	2	1	3
2020-02-25	3	4	7
2020-02-26	5	1	6
2020-02-27	6	2	8
2020-02-28	4	6	10
2020-02-29	9	1	10
2020-03-01	13	7	20
2020-03-02	34	4	38
2020-03-03	30	11	41
2020-03-04	34	12	46
2020-03-05	38	7	45
2020-03-06	31	17	48
2020-03-07	19	4	23
2020-03-08	24	6	30
2020-03-09	31	9	40
2020-03-10	45	14	59
2020-03-11	72	12	84
2020-03-12	83	15	98
2020-03-13	30	9	39
2020-03-14	24	9	33
2020-03-15	25	10	35
2020-03-16	19	3	22
2020-03-17	25	9	34
2020-03-18	47	19	66
2020-03-19	35	7	42
2020-03-20	34	13	47
2020-03-21	26	12	38
2020-03-22	35	6	41
2020-03-23	69	14	83
2020-03-24	72	9	81
2020-03-25	55	12	67
2020-03-26	47	11	58
2020-03-27	57	13	70
2020-03-28	46	8	54
2020-03-29	33	6	39
2020-03-30	59	14	73
2020-03-31	53	7	60
2020-04-01	31	5	36
2020-04-02	18	4	22
2020-04-03	17	6	23
2020-04-04	7	3	10
2020-04-05	1	1	2
2020-04-06	6	0	6
2020-04-07	6	2	8
2020-04-08	7	2	9

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

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

Day	England
2020-02-03	5
2020-02-05	1
2020-02-08	2
2020-02-09	2
2020-02-13	2
2020-02-14	2
2020-02-15	2
2020-02-16	4
2020-02-18	1
2020-02-19	1
2020-02-20	1
2020-02-23	2
2020-02-24	4
2020-02-25	7
2020-02-26	6
2020-02-27	18
2020-02-28	22
2020-02-29	22
2020-03-01	49
2020-03-02	61
2020-03-03	82
2020-03-04	93
2020-03-05	79
2020-03-06	71
2020-03-07	38
2020-03-08	42
2020-03-09	59
2020-03-10	70
2020-03-11	128
2020-03-12	146
2020-03-13	68
2020-03-14	54
2020-03-15	44
2020-03-16	40
2020-03-17	64
2020-03-18	111
2020-03-19	71
2020-03-20	100
2020-03-21	83
2020-03-22	64
2020-03-23	163
2020-03-24	154
2020-03-25	122
2020-03-26	120
2020-03-27	145
2020-03-28	101
2020-03-29	84
2020-03-30	142
2020-03-31	143
2020-04-01	81
2020-04-02	65
2020-04-03	68
2020-04-04	22
2020-04-05	5
2020-04-06	22
2020-04-07	34
2020-04-08	30

Day	England
2020-04-09	52
2020-04-10	16
2020-04-11	10
2020-04-12	12
2020-04-13	55
2020-04-14	51
2020-04-15	52
2020-04-16	54
2020-04-17	49
2020-04-18	18
2020-04-19	21
2020-04-20	41
2020-04-21	27
2020-04-22	3
2020-04-23	4
2020-04-24	10
2020-04-25	1
2020-04-26	1
2020-04-27	12
2020-04-28	18
2020-04-29	25
2020-04-30	35
2020-05-01	31
2020-05-02	21
2020-05-03	2
2020-05-04	21
2020-05-05	13
2020-05-06	4
2020-05-07	7
2020-05-09	1
2020-05-10	1
2020-05-11	3
2020-05-12	3
2020-05-13	6
2020-05-14	10
2020-05-15	1
2020-05-18	3
2020-05-19	2

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	11	10-50
BERKSHIRE	England	6	1-10
BRISTOL	England	16	10-50
BUCKINGHAMSHIRE	England	20	10-50
CAMBRIDGESHIRE	England	80	50-100
CARDIFF	Wales	1	1-10
CHESHIRE	England	12	10-50
CORNWALL	England	2	1-10
CUMBRIA	England	10	10-50
DERBYSHIRE	England	11	10-50
DEVON	England	22	10-50
DORSET	England	15	10-50
DURHAM	England	1	1-10
EAST RIDING OF YORKSHIRE	England	25	10-50
ESSEX	England	52	50-100
GLOUCESTERSHIRE	England	9	1-10
GREATER LONDON	England	1750	>500
GUERNSEY	Channel_islands	41	10-50
HAMPSHIRE	England	35	10-50
HEREFORDSHIRE	England	1	1-10
HERTFORDSHIRE	England	246	200-250
JERSEY	Channel_islands	77	50-100
KENT	England	26	10-50
LANCASHIRE	England	8	1-10
LEICESTERSHIRE	England	5	1-10
LINCOLNSHIRE	England	5	1-10
MANCHESTER	England	29	10-50
MERSEYSIDE	England	58	50-100
NORFOLK	England	2	1-10
NORTH YORKSHIRE	England	5	1-10
NORTHAMPTONSHIRE	England	11	10-50
NORTHUMBERLAND	England	1	1-10
NOTTINGHAMSHIRE	England	10	10-50
OXFORDSHIRE	England	24	10-50
SHROPSHIRE	England	1	1-10
SOMERSET	England	73	50-100
SOUTH YORKSHIRE	England	44	10-50
STAFFORDSHIRE	England	28	10-50
SURREY	England	41	10-50
TYNE AND WEAR	England	37	10-50
WARWICKSHIRE	England	9	1-10
WEST MIDLANDS	England	50	50-100
WEST YORKSHIRE	England	20	10-50
WILTSHIRE	England	12	10-50
WORCESTERSHIRE	England	7	1-10