

Lineages report for PHEC

This report gives summaries of UK specific lineages sequenced by PHEC for week 2020-06-19. 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 180 and the maximum is 2693

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

313 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 26 that remain: 16 are pending extinction, ie last seen three weeks ago. 4 lineages have gone quiet, ie haven't been seen this week. 4 lineages have reactivated. 2 lineages have 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 2020-05-19

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
UK5	788 (100.0%)	Feb-23, May-19	788	B.1.1, B.1.1.13, B.1.1.1	0	active today
UK107	607 (100.0%)	Feb-09, May-07	607	B.2.5, B.2, B.2.1	12	0.0121
UK42	201 (100.0%)	Feb-03, May-13	201	B.1.5, B.1.72, B.1	6	0.0833
UK5676	121 (100.0%)	Feb-14, Apr-13	121	B.2	36	0.0137
UK2916	112 (100.0%)	Feb-03, Apr-14	112	B.1	35	0.0183
UK72	99 (100.0%)	Feb-05, May-01	99	B, B.2	18	0.0488
UK9	84 (100.0%)	Mar-09, May-15	84	B.1.13	4	0.2018
UK2913	75 (100.0%)	Mar-07, Apr-20	75	B.1, B.1.p11	29	0.0205
UK15	72 (100.0%)	Feb-27, May-06	72	B.1.1	13	0.0748
UK2464	50 (100.0%)	Mar-09, May-04	50	B.1.p11	15	0.0762

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.

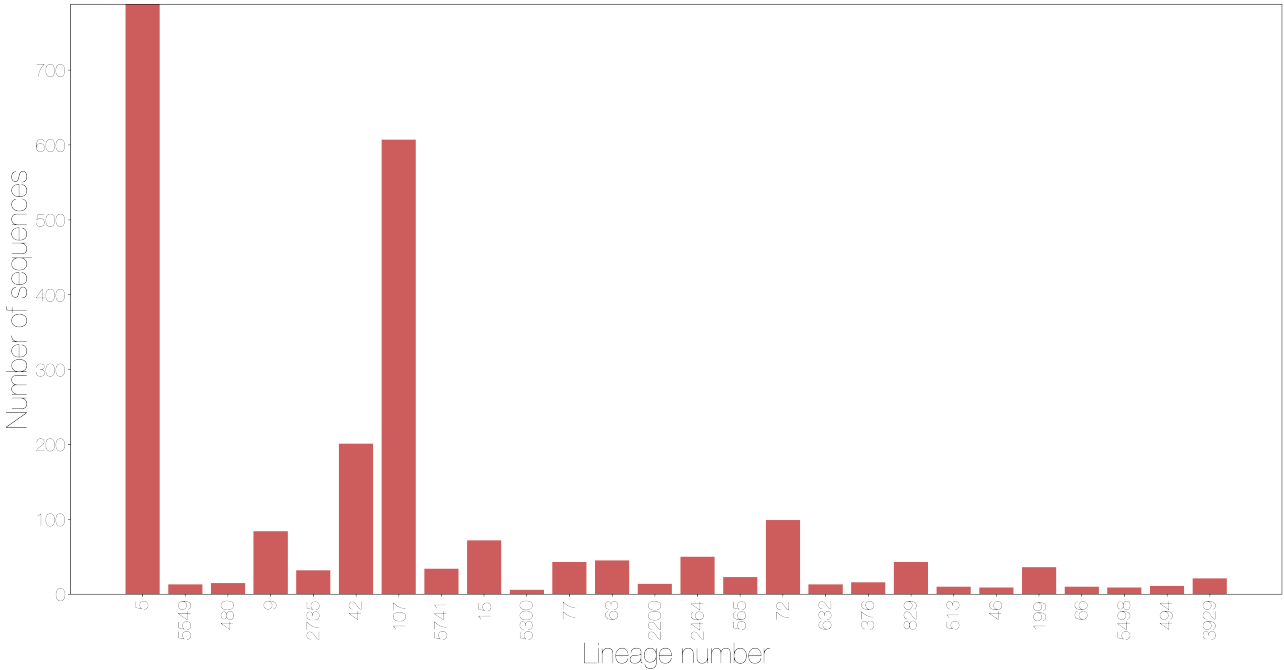


Figure 1: Number of sequences sampled in a lineage by country

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 31 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.

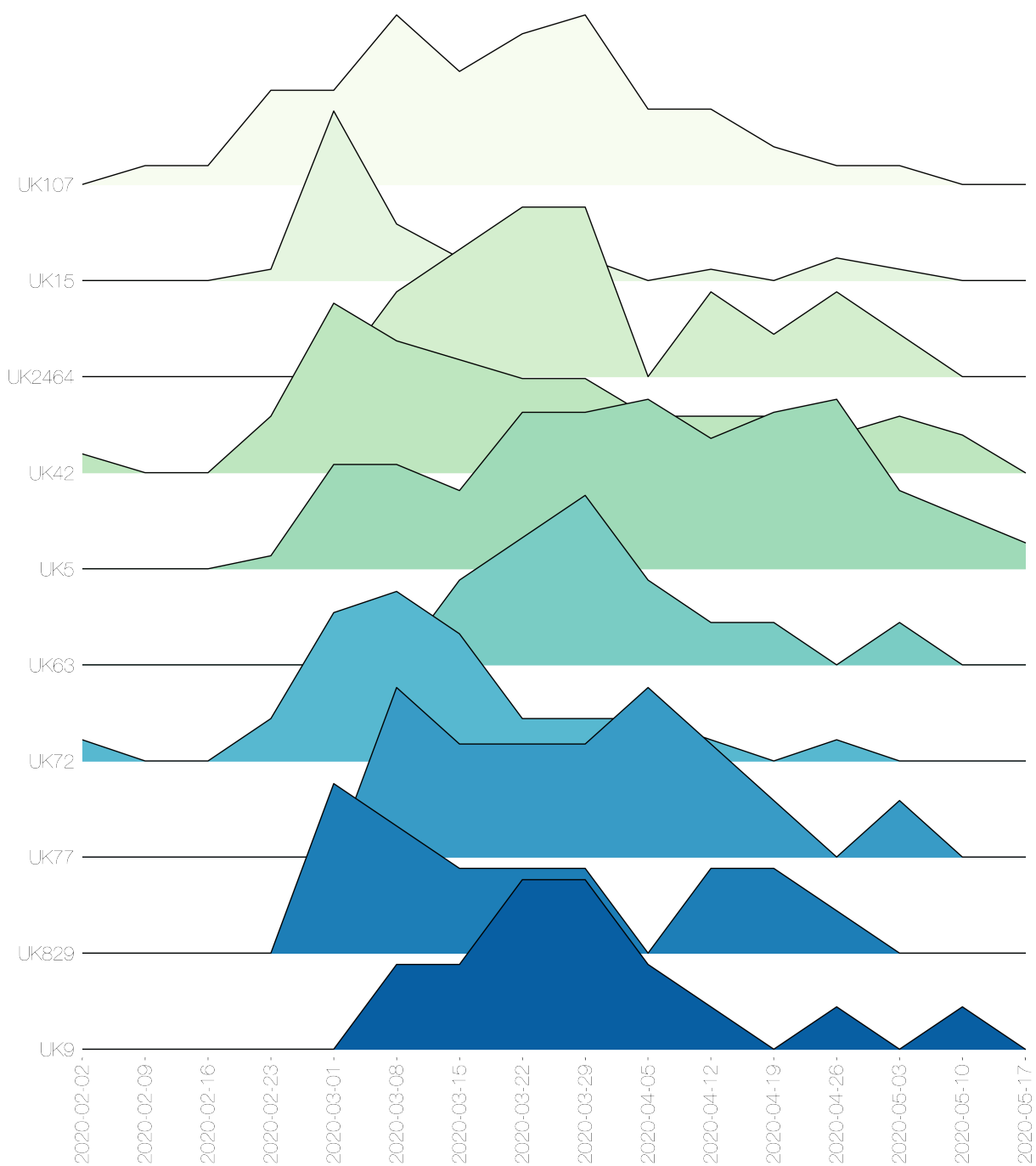


Figure 2: Lineages by number of adm2 regions present by epiweek

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.

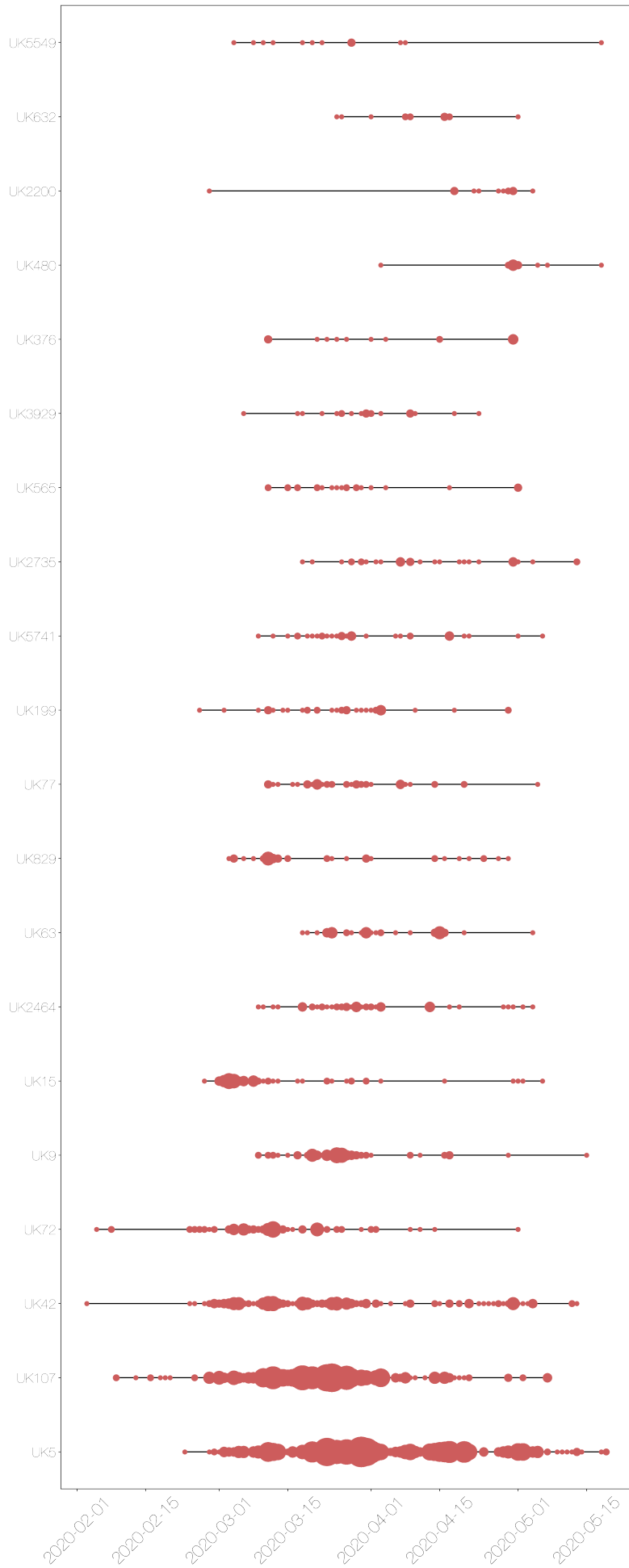


Figure 3: Timeline of lineages, sized by number of sequences from each country.

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

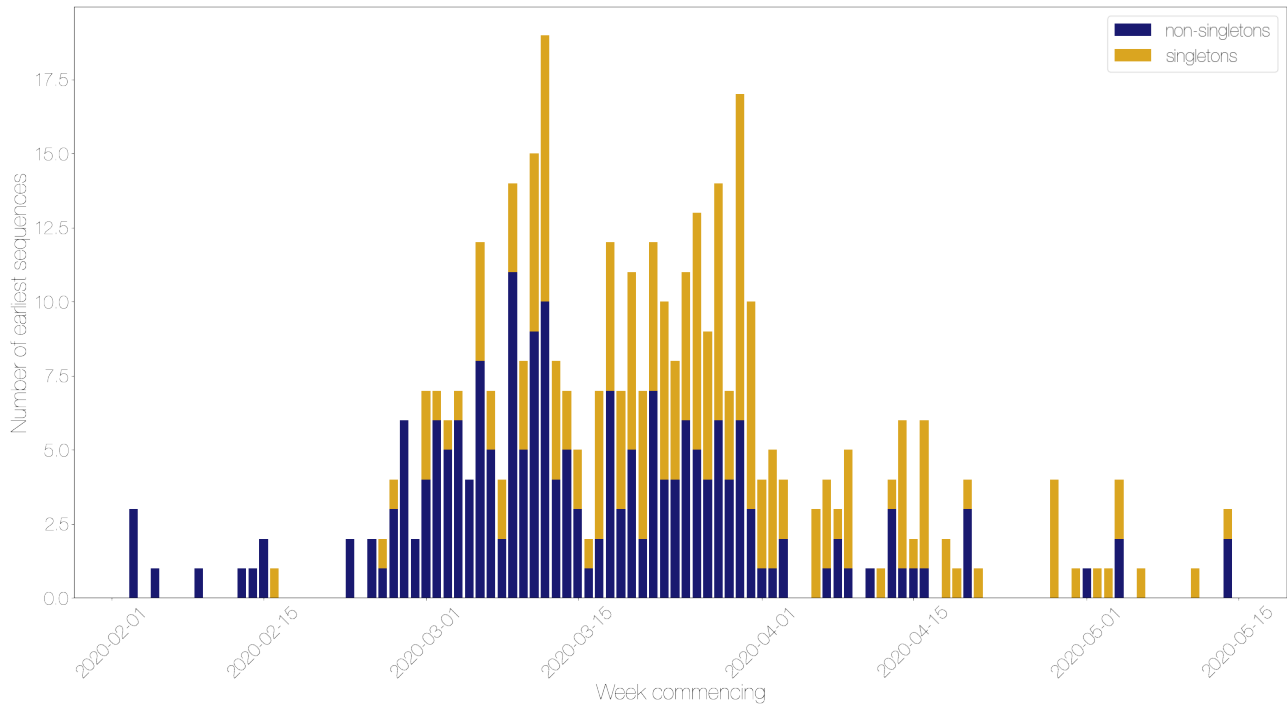


Figure 4: Lineage starts per week, split by singletons and non-singletons

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.

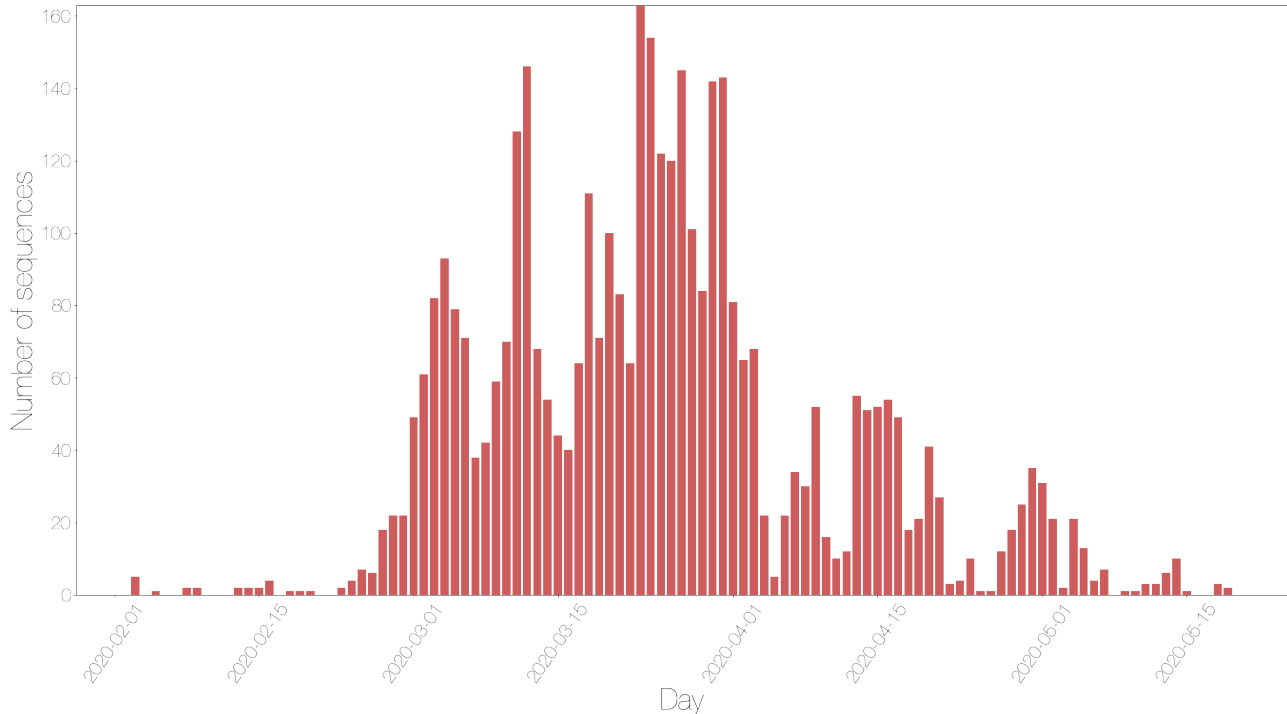


Figure 5: Sequences taken on each day by country

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.

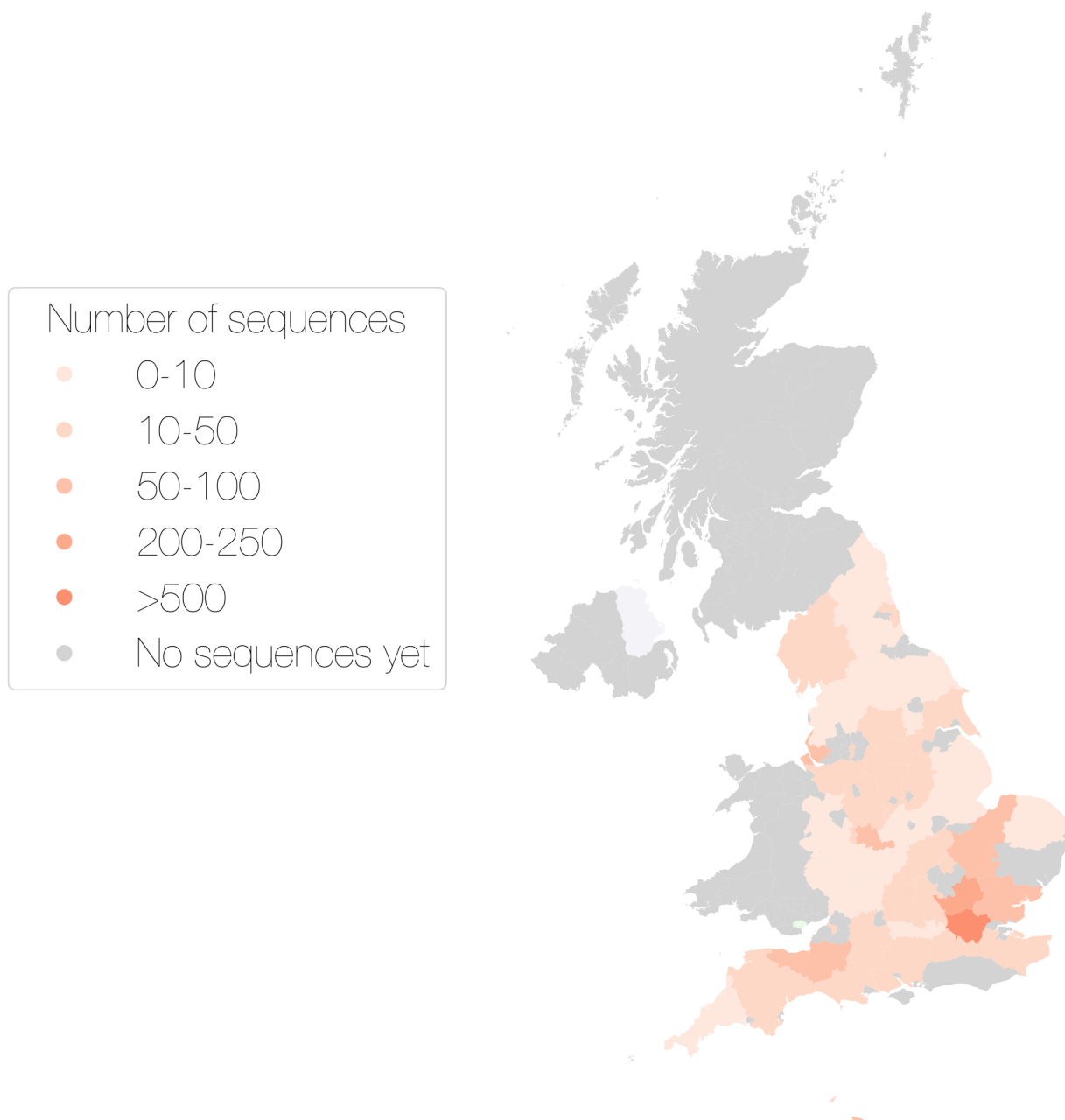


Figure 6: Map showing the number of sequences sampled by adm2 region

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
UK5	788 (100.0%)	Feb-23, May-19	788	B.1.1, B.1.1.13, B.1.1.1	0	active today
UK107	607 (100.0%)	Feb-09, May-07	607	B.2.5, B.2, B.2.1	12	0.0121
UK42	201 (100.0%)	Feb-03, May-13	201	B.1.5, B.1.72, B.1	6	0.0833
UK5676	121 (100.0%)	Feb-14, Apr-13	121	B.2	36	0.0137
UK2916	112 (100.0%)	Feb-03, Apr-14	112	B.1	35	0.0183
UK72	99 (100.0%)	Feb-05, May-01	99	B, B.2	18	0.0488
UK9	84 (100.0%)	Mar-09, May-15	84	B.1.13	4	0.2018
UK2913	75 (100.0%)	Mar-07, Apr-20	75	B.1, B.1.p11	29	0.0205
UK15	72 (100.0%)	Feb-27, May-06	72	B.1.1	13	0.0748
UK2464	50 (100.0%)	Mar-09, May-04	50	B.1.p11	15	0.0762
UK63	45 (100.0%)	Mar-18, May-04	45	B.1.1	15	0.0712
UK120	45 (100.0%)	Feb-03, Mar-31	45	B	49	0.0264
UK829	43 (100.0%)	Mar-03, Apr-29	43	B.2.5	20	0.0679
UK77	43 (100.0%)	Mar-11, May-05	43	B.2	14	0.0935
UK4	40 (100.0%)	Feb-28, Mar-31	40	B	49	0.0167
UK199	36 (100.0%)	Feb-26, Apr-29	36	B.1.5, B.1	20	0.09
UK339	34 (100.0%)	Feb-23, Apr-09	34	B.3	40	0.0348
UK5741	34 (100.0%)	Mar-09, May-06	34	B.1	13	0.1352
UK5561	33 (100.0%)	Feb-25, Apr-03	33	B.2.2	46	0.0258
UK2735	32 (100.0%)	Mar-18, May-13	32	B.1.1	6	0.3011
UK85	28 (100.0%)	Mar-09, Apr-17	28	B, B.3	32	0.0451
UK94	26 (100.0%)	Mar-12, Apr-19	26	B.2, B.2.1	30	0.0507
UK167	24 (100.0%)	Mar-06, Apr-12	24	B.1	37	0.0435
UK240	24 (100.0%)	Feb-25, Mar-31	24	B.2.1, B.2	49	0.0311
UK384	24 (100.0%)	Feb-28, Apr-14	24	B.2, B.2.1	35	0.0571
UK404	23 (100.0%)	Mar-01, Apr-12	23	B.1	37	0.0516

Lineage name	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK5180	23 (100.0%)	Mar-07, Apr-19	23	B.1.1.7	30	0.0652
UK565	23 (100.0%)	Mar-11, May-01	23	B.1.1	18	0.1288
UK370	21 (100.0%)	Mar-06, Apr-09	21	B.1.1.10	40	0.0425
UK3929	21 (100.0%)	Mar-06, Apr-23	21	B.1.1.4, B.1.1, B.1.1.3	26	0.0923
UK275	20 (100.0%)	Mar-09, Apr-21	20	B.1.13	28	0.0808
UK41	20 (100.0%)	Feb-29, Mar-26	20	B.1	54	0.0253
UK517	20 (100.0%)	Mar-02, Apr-15	20	B.1.1	34	0.0681
UK18	20 (100.0%)	Mar-11, Apr-10	20	B.1.1.7	39	0.0405
UK31	18 (100.0%)	Mar-11, Apr-19	18	B.3	30	0.0765
UK371	18 (100.0%)	Mar-12, Apr-09	18	B.1.1	40	0.0412
UK37	17 (100.0%)	Mar-18, Apr-02	17	B.1.30, B.1	47	0.0199
UK376	16 (100.0%)	Mar-11, Apr-30	16	B.1.1.9	19	0.1754
UK241	16 (100.0%)	Mar-25, Apr-07	16	B.1.5.3	42	0.0206
UK34	15 (100.0%)	Feb-15, Apr-02	15	B.4	47	0.0714
UK480	15 (100.0%)	Apr-03, May-18	15	B.1.1	1	3.2143
UK2200	14 (100.0%)	Feb-28, May-04	14	B.1.5.6, B.1.5	15	0.3385
UK632	13 (100.0%)	Mar-25, May-01	13	B.1.1	18	0.1713
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
UK119	13 (100.0%)	Mar-11, Apr-16	13	B.2.5	33	0.0909
UK274	13 (100.0%)	Mar-06, Apr-02	13	B.3	47	0.0479
UK2906	13 (100.0%)	Mar-06, Mar-25	13	B.1	55	0.0288
UK12	12 (100.0%)	Mar-12, Mar-31	12	B.1.p11	49	0.0353
UK494	11 (100.0%)	Mar-26, Apr-28	11	B.1.p11	21	0.1571
UK604	10 (100.0%)	Mar-09, Mar-12	10	B.1.1	68	0.0049
UK5715	10 (100.0%)	Feb-13, Mar-07	10	B.2	73	0.035
UK3021	10 (100.0%)	Mar-12, Apr-15	10	B.1	34	0.1111
UK491	10 (100.0%)	Mar-03, Apr-14	10	B, B.2, B.2.1	35	0.1333
UK64	10 (100.0%)	Mar-12, Apr-17	10	B.1	32	0.125

Lineage name	England	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK788	10 (100.0%)	Feb-28, Mar-05	10	B.4	75	0.0089
UK66	10 (100.0%)	Mar-28, Apr-28	10	B.1.1.8	21	0.164
UK513	10 (100.0%)	Mar-12, Apr-29	10	B.1.p11	20	0.2667
UK276	10 (100.0%)	Mar-10, Apr-15	10	B.1.1	34	0.1176
UK501	9 (100.0%)	Mar-02, Mar-31	9	B.1	49	0.074
UK5498	9 (100.0%)	Mar-06, Apr-28	9	B, B.2	21	0.3155
UK46	9 (100.0%)	Mar-02, Apr-29	9	B.2.1	20	0.3625
UK756	8 (100.0%)	Feb-27, Mar-05	8	B.1.1	75	0.0133
UK242	8 (100.0%)	Mar-30, Apr-20	8	B.1.5	29	0.1034
UK4237	8 (100.0%)	Apr-14, Apr-18	8	B.1.1	31	0.0184
UK739	8 (100.0%)	Mar-01, Mar-08	8	B.4	72	0.0139
UK22	7 (100.0%)	Mar-02, Mar-18	7	B	62	0.043
UK38	7 (100.0%)	Mar-04, Apr-20	7	B.2.1	29	0.2701
UK6	7 (100.0%)	Mar-06, Apr-20	7	B.1	29	0.2586
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
UK799	6 (100.0%)	Mar-01, Mar-07	6	B.1	73	0.0164
UK654	6 (100.0%)	Feb-27, Mar-08	6	B.2.5	72	0.0278
UK497	6 (100.0%)	Mar-29, Apr-09	6	A.2	40	0.055
UK178	6 (100.0%)	Mar-14, Apr-04	6	B.1.1	45	0.0933

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	31

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

Week commencing	UK5	UK107	UK42	UK72	UK9	UK15	UK2464	UK63	UK829	UK77
2020-02-02	0	0	1	1	0	0	0	0	0	0
2020-02-09	0	1	0	0	0	0	0	0	0	0
2020-02-16	0	1	0	0	0	0	0	0	0	0
2020-02-23	1	5	3	2	0	1	0	0	0	0
2020-03-01	8	5	9	7	0	15	0	0	4	0
2020-03-08	8	9	7	8	2	5	2	0	3	3
2020-03-15	6	6	6	6	2	2	3	2	2	2
2020-03-22	12	8	5	2	4	2	4	3	2	2
2020-03-29	12	9	5	2	4	2	4	4	2	2
2020-04-05	13	4	3	2	2	0	0	2	0	3
2020-04-12	10	4	3	1	1	1	2	1	2	2
2020-04-19	12	2	3	0	0	0	1	1	2	1
2020-04-26	13	1	2	1	1	2	2	0	1	0
2020-05-03	6	1	3	0	0	1	1	1	0	1
2020-05-10	4	0	2	0	1	0	0	0	0	0
2020-05-17	2	0	0	0	0	0	0	0	0	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
2020-02-03	0	3	3
2020-02-05	0	1	1
2020-02-09	0	1	1
2020-02-13	0	1	1
2020-02-14	0	1	1
2020-02-15	0	2	2
2020-02-16	1	0	1
2020-02-23	0	2	2
2020-02-25	0	2	2
2020-02-26	1	1	2
2020-02-27	1	3	4
2020-02-28	0	6	6
2020-02-29	0	2	2
2020-03-01	3	4	7
2020-03-02	1	6	7
2020-03-03	1	5	6
2020-03-04	1	6	7
2020-03-05	0	4	4
2020-03-06	4	8	12
2020-03-07	2	5	7
2020-03-08	2	2	4
2020-03-09	3	11	14
2020-03-10	3	5	8
2020-03-11	6	9	15
2020-03-12	9	10	19
2020-03-13	4	4	8
2020-03-14	2	5	7
2020-03-15	2	3	5
2020-03-16	1	1	2
2020-03-17	5	2	7
2020-03-18	5	7	12
2020-03-19	4	3	7
2020-03-20	6	5	11
2020-03-21	5	2	7
2020-03-22	5	7	12
2020-03-23	6	4	10
2020-03-24	4	4	8
2020-03-25	5	6	11
2020-03-26	8	5	13
2020-03-27	5	4	9
2020-03-28	8	6	14
2020-03-29	3	4	7
2020-03-30	11	6	17
2020-03-31	7	3	10
2020-04-01	3	1	4
2020-04-02	4	1	5
2020-04-03	2	2	4
2020-04-06	3	0	3
2020-04-07	3	1	4
2020-04-08	1	2	3
2020-04-09	4	1	5
2020-04-11	0	1	1
2020-04-12	1	0	1
2020-04-13	1	3	4
2020-04-14	5	1	6
2020-04-15	1	1	2
2020-04-16	5	1	6

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-04-18	2	0	2
2020-04-19	1	0	1
2020-04-20	1	3	4
2020-04-21	1	0	1
2020-04-28	4	0	4
2020-04-30	1	0	1
2020-05-01	0	1	1
2020-05-02	1	0	1
2020-05-03	1	0	1
2020-05-04	2	2	4
2020-05-06	1	0	1
2020-05-11	1	0	1
2020-05-14	1	2	3

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