

Lineages report for PHWC

This report gives summaries of UK specific lineages sequenced by PHWC for week 2020-06-19. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-06-06. The analysis (eg time since last sample) is therefore undertaken from this date. 3953 sequences in the UK from the sequencing centre PHWC 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 111 and the maximum is 1398

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

170 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: 7 are pending extinction, ie last seen three weeks ago. 13 lineages have gone quiet, ie haven't been seen this week. 2 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 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	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK5	821 (100.0%)	Mar-01, Jun-04	821	B.1.1.p16, B.1.1.p11, B.1.1, B.1.1.1, B.1.1.16	2	0.0579
UK61	397 (100.0%)	Mar-08, May-27	397	B.3	10	0.0202
UK42	371 (100.0%)	Feb-27, May-27	371	B.1.67, B.1.71, B.1, B.1.p11, B.1.35	10	0.0243
UK158	301 (100.0%)	Mar-20, Jun-03	301	B.1.1, B.1.1.2	3	0.0833
UK3021	217 (100.0%)	Mar-29, Jun-06	217	B.1	0	active today
UK632	205 (100.0%)	Mar-25, Jun-04	205	B.1.1	2	0.174
UK167	115 (100.0%)	Jan-27, May-28	115	B.1	9	0.1189
UK5741	111 (100.0%)	Mar-17, Jun-02	111	B.1, B.1.44	4	0.175
UK2464	80 (100.0%)	Mar-26, May-11	80	B.1.p11	26	0.0224
UK394	79 (100.0%)	Mar-17, May-22	79	B.1.1, B.1.1.10	15	0.0564

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.

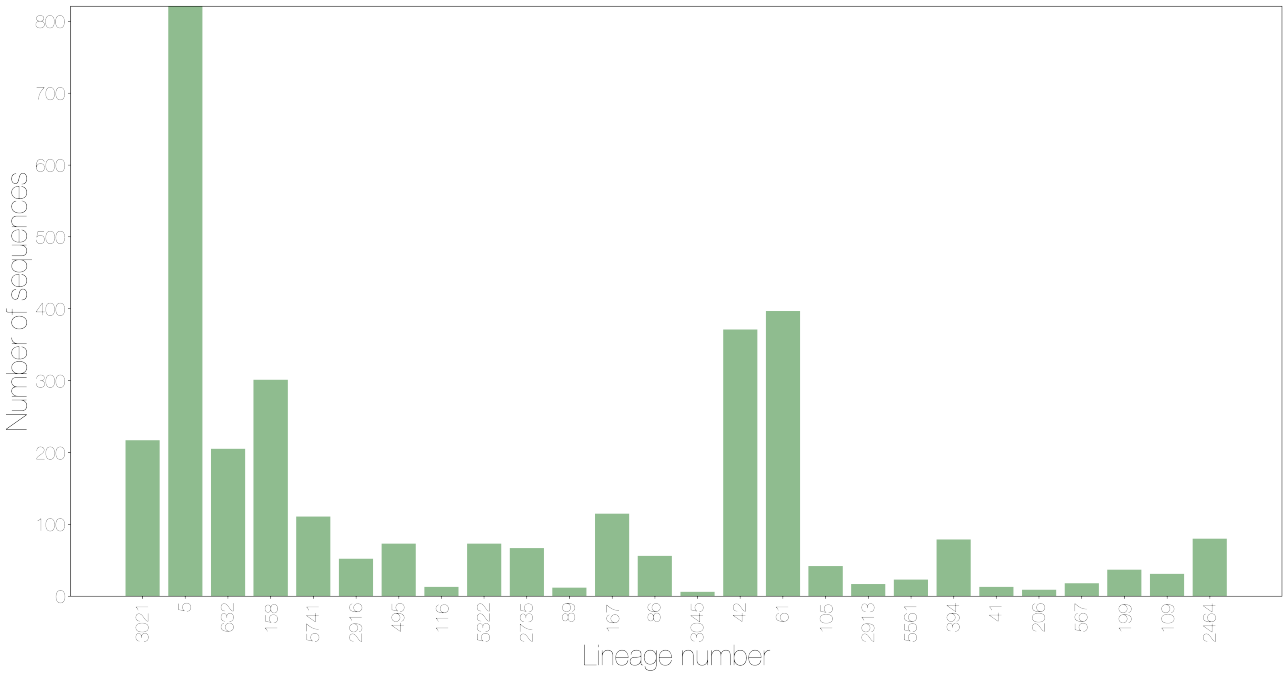


Figure 1: Number of sequences sampled in a lineage by country

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 13 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.

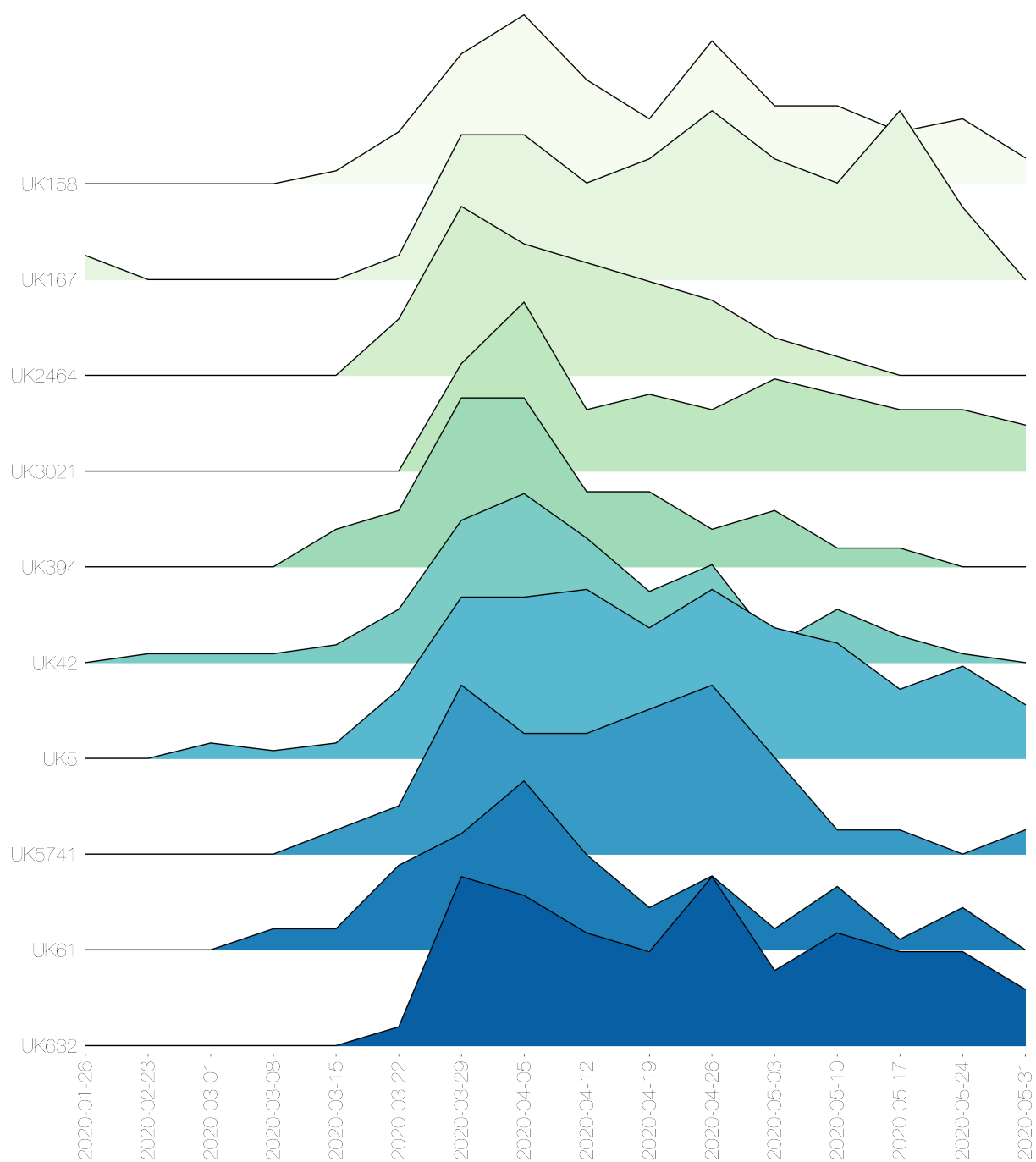


Figure 2: Lineages by number of adm2 regions present by epiweek

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.

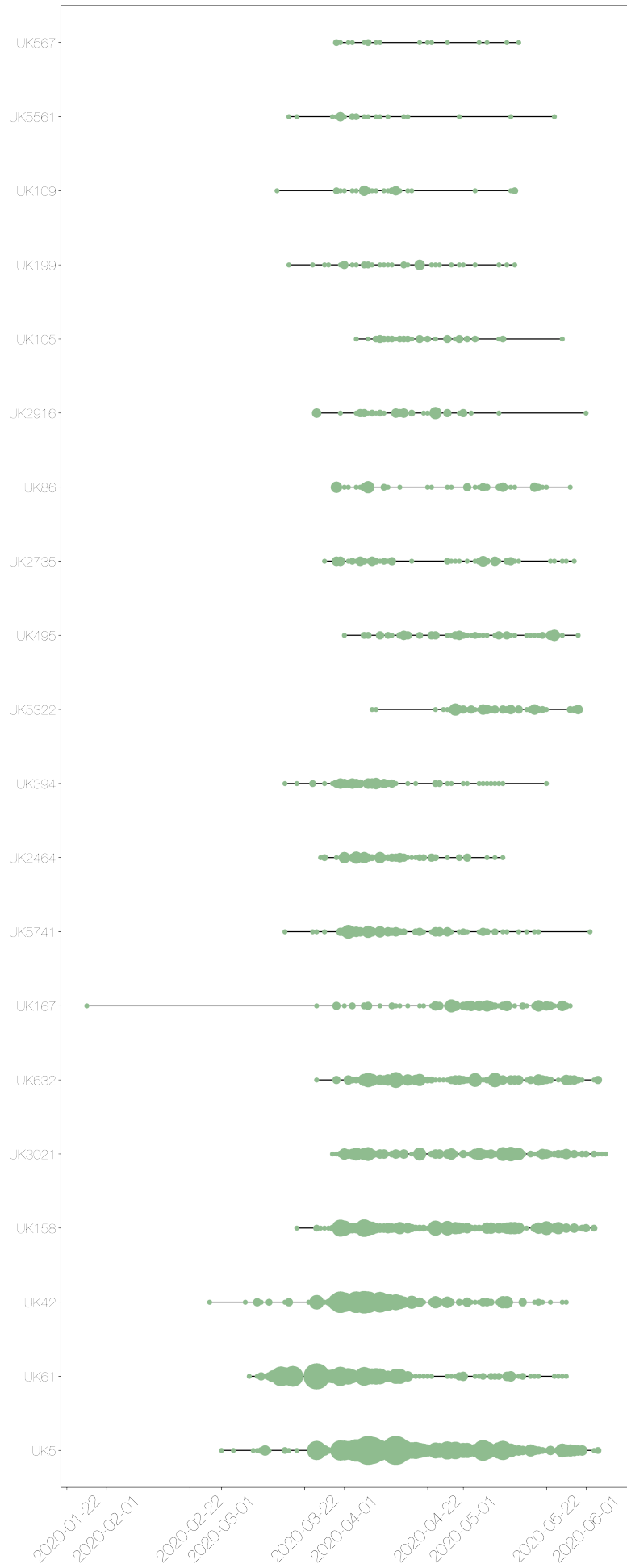


Figure 3: Timeline of lineages, sized by number of sequences from each country.

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

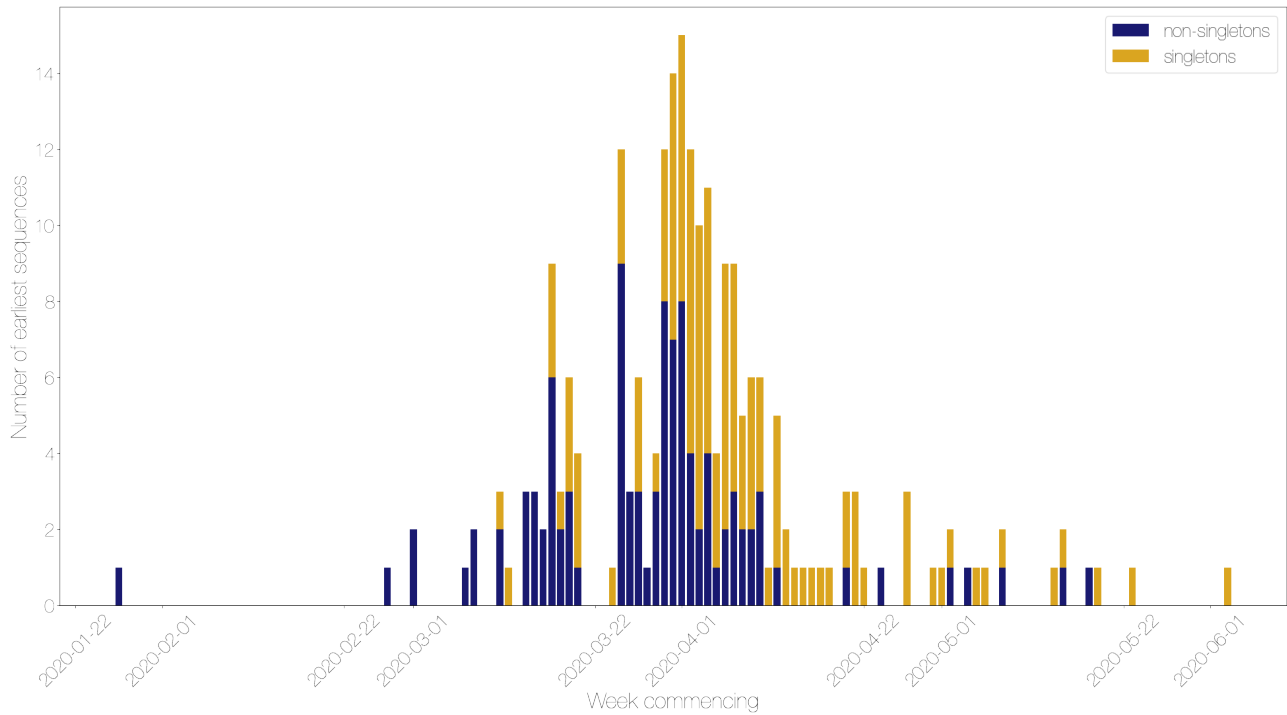


Figure 4: Lineage starts per week, split by singletons and non-singletons

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.

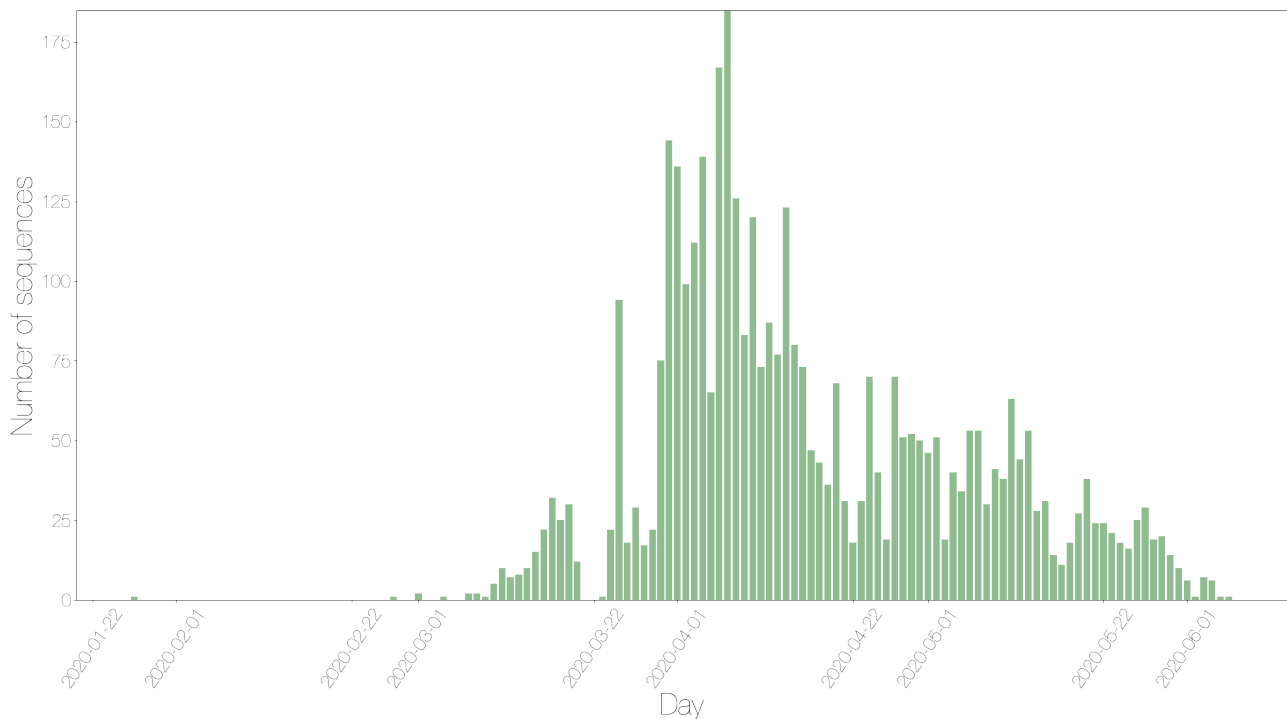


Figure 5: Sequences taken on each day by country

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 584 sequences without enough geographical information to map

from this centre.

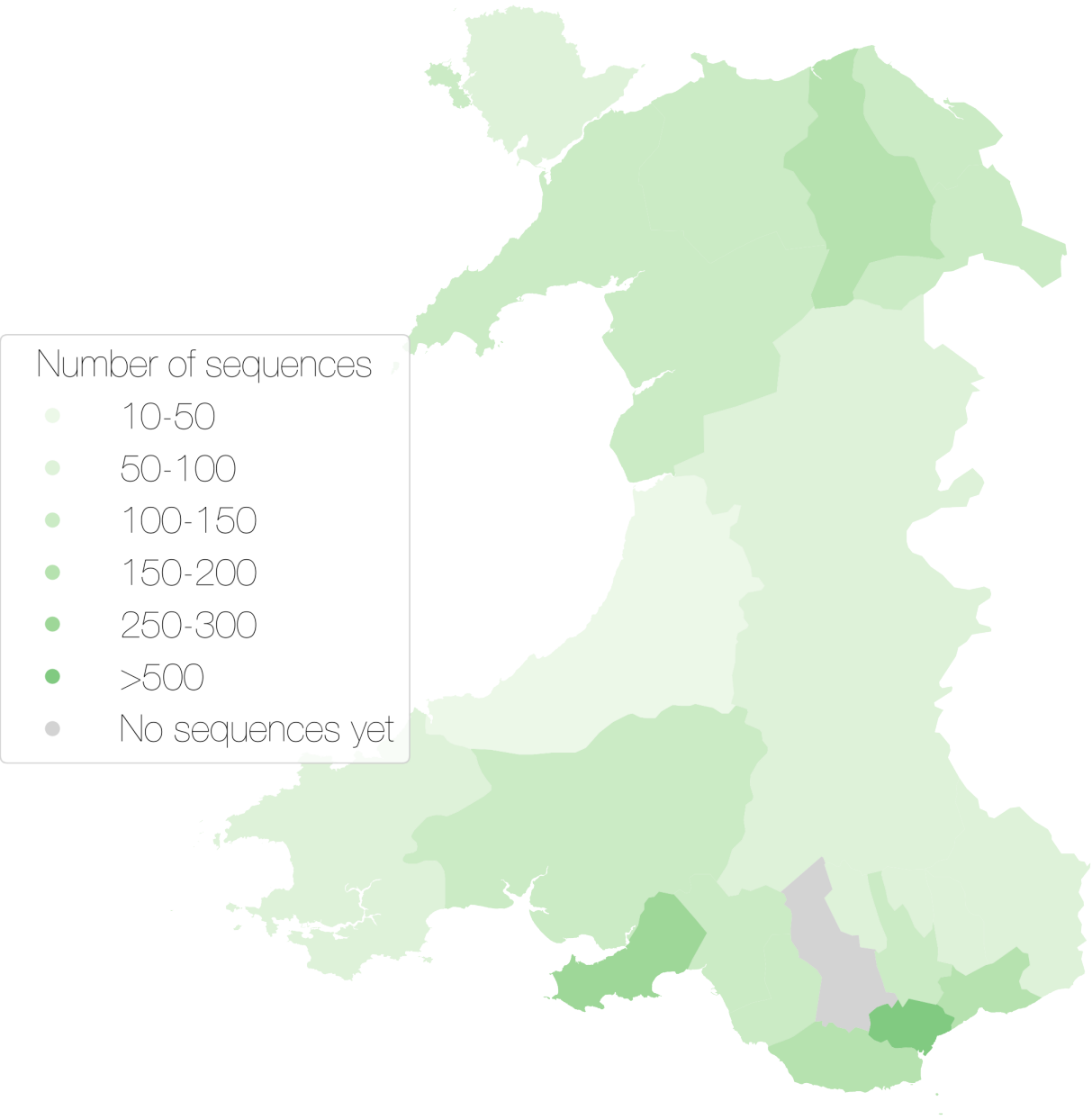


Figure 6: Map showing the number of sequences sampled by adm2 region

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	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK5	821 (100.0%)	Mar-01, Jun-04	821	B.1.1.p16, B.1.1.p11, B.1.1, B.1.1.1, B.1.1.16	2	0.0579
UK61	397 (100.0%)	Mar-08, May-27	397	B.3	10	0.0202
UK42	371 (100.0%)	Feb-27, May-27	371	B.1.67, B.1.71, B.1, B.1.p11, B.1.35	10	0.0243
UK158	301 (100.0%)	Mar-20, Jun-03	301	B.1.1, B.1.1.2	3	0.0833
UK3021	217 (100.0%)	Mar-29, Jun-06	217	B.1	0	active today
UK632	205 (100.0%)	Mar-25, Jun-04	205	B.1.1	2	0.174
UK167	115 (100.0%)	Jan-27, May-28	115	B.1	9	0.1189
UK5741	111 (100.0%)	Mar-17, Jun-02	111	B.1, B.1.44	4	0.175
UK2464	80 (100.0%)	Mar-26, May-11	80	B.1.p11	26	0.0224
UK394	79 (100.0%)	Mar-17, May-22	79	B.1.1, B.1.1.10	15	0.0564
UK5322	73 (100.0%)	Apr-08, May-30	73	B.1.1	7	0.1032
UK495	73 (100.0%)	Apr-01, May-30	73	B.1.p11	7	0.1171
UK2735	67 (100.0%)	Mar-27, May-29	67	B.1.1	8	0.1193
UK5676	57 (100.0%)	Mar-15, May-01	57	B.2, B.3	36	0.0233
UK86	56 (100.0%)	Mar-30, May-28	56	B.1	9	0.1192
UK107	55 (100.0%)	Mar-14, Apr-23	55	B.2.1	44	0.0168
UK370	54 (100.0%)	Mar-19, Apr-27	54	B.1.1.10	40	0.0184
UK2916	52 (100.0%)	Mar-25, Jun-01	52	B.1	5	0.2667
UK105	42 (100.0%)	Apr-04, May-26	42	B.1.p11	11	0.1153
UK199	37 (100.0%)	Mar-18, May-14	37	B.1, B.1.5	23	0.0688
UK2200	33 (100.0%)	Mar-15, Apr-29	33	B.1.5.6, B.1.5	38	0.037
UK109	31 (100.0%)	Mar-15, May-14	31	B.1.5	23	0.087
UK187	29 (100.0%)	Mar-30, Apr-30	29	B.1	37	0.0299
UK5561	23 (100.0%)	Mar-18, May-24	23	B.2.2	13	0.2343
UK179	23 (100.0%)	Mar-17, May-07	23	B.1.1.p11, B.1.1	30	0.0773
UK567	18 (100.0%)	Mar-30, May-15	18	B.2.2	22	0.123

Lineage name	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK2913	17 (100.0%)	Mar-16, May-24	17	B.1, B.1.p11	13	0.3317
UK198	15 (100.0%)	Mar-26, May-07	15	B.1.5	30	0.1
UK339	15 (100.0%)	Mar-14, Apr-14	15	B.3	53	0.0418
UK425	14 (100.0%)	Mar-28, May-05	14	B.1.1	32	0.0913
UK72	14 (100.0%)	Mar-11, Apr-17	14	B	50	0.0569
UK202	13 (100.0%)	Apr-24, May-09	13	B.1.1	28	0.0446
UK41	13 (100.0%)	Apr-10, May-21	13	B.1	16	0.2135
UK116	13 (100.0%)	May-08, May-30	13	B.1	7	0.2619
UK89	12 (100.0%)	Apr-10, May-28	12	B.1.1.9	9	0.4848
UK317	12 (100.0%)	Mar-19, Apr-20	12	B.3	47	0.0619
UK64	12 (100.0%)	Mar-25, May-05	12	B.1	32	0.1165
UK607	12 (100.0%)	Mar-11, Apr-24	12	B	43	0.093
UK15	11 (100.0%)	Mar-17, Apr-13	11	B.1.1	54	0.05
UK801	10 (100.0%)	Apr-05, May-05	10	B.1	32	0.1042
UK696	9 (100.0%)	Apr-10, May-01	9	B.1, B.1.5	36	0.0729
UK206	9 (100.0%)	Apr-02, May-20	9	B.1	17	0.3529
UK633	8 (100.0%)	Apr-03, Apr-28	8	B.1.1.p16, B.1.1.16	39	0.0916
UK275	8 (100.0%)	Mar-31, Apr-18	8	B.1.13	49	0.0525
UK491	8 (100.0%)	Mar-31, Apr-12	8	B, B.2.1	55	0.0312
UK119	7 (100.0%)	Mar-30, Apr-14	7	B.2.5	53	0.0472
UK506	7 (100.0%)	Apr-02, Apr-20	7	B.1.1	47	0.0638
UK5498	7 (100.0%)	Apr-01, Apr-14	7	B.2	53	0.0409
UK462	7 (100.0%)	Apr-01, Apr-20	7	B.1	47	0.0674
UK3045	6 (100.0%)	May-15, May-27	6	B.1.1.p11, B.1.1	10	0.24
UK451	6 (100.0%)	Mar-25, Apr-05	6	B.2.1	62	0.0355

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	PHWC	13

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

Week commencing	UK5	UK61	UK42	UK158	UK3021	UK632	UK167	UK5741	UK2464	UK394
2020-01-26	0	0	0	0	0	0	1	0	0	0
2020-02-23	0	0	1	0	0	0	0	0	0	0
2020-03-01	2	0	1	0	0	0	0	0	0	0
2020-03-08	1	2	1	0	0	0	0	0	0	0
2020-03-15	2	2	2	1	0	0	0	1	0	2
2020-03-22	9	8	6	4	0	1	1	2	3	3
2020-03-29	21	11	16	10	7	9	6	7	9	9
2020-04-05	21	16	19	13	11	8	6	5	7	9
2020-04-12	22	9	14	8	4	6	4	5	6	4
2020-04-19	17	4	8	5	5	5	5	6	5	4
2020-04-26	22	7	11	11	4	9	7	7	4	2
2020-05-03	17	2	2	6	6	4	5	4	2	3
2020-05-10	15	6	6	6	5	6	4	1	1	1
2020-05-17	9	1	3	4	4	5	7	1	0	1
2020-05-24	12	4	1	5	4	5	3	0	0	0
2020-05-31	7	0	0	2	3	3	0	1	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
2020-01-27	0	1	1
2020-02-27	0	1	1
2020-03-01	0	2	2
2020-03-07	0	1	1
2020-03-08	0	2	2
2020-03-11	1	2	3
2020-03-12	1	0	1
2020-03-14	0	3	3
2020-03-15	0	3	3
2020-03-16	0	2	2
2020-03-17	3	6	9
2020-03-18	1	2	3
2020-03-19	3	3	6
2020-03-20	3	1	4
2020-03-24	1	0	1
2020-03-25	3	9	12
2020-03-26	0	3	3
2020-03-27	3	3	6
2020-03-28	0	1	1
2020-03-29	1	3	4
2020-03-30	4	8	12
2020-03-31	7	7	14
2020-04-01	7	8	15
2020-04-02	8	4	12
2020-04-03	8	2	10
2020-04-04	7	4	11
2020-04-05	3	1	4
2020-04-06	7	2	9
2020-04-07	6	3	9
2020-04-08	3	2	5
2020-04-09	4	2	6
2020-04-10	3	3	6
2020-04-11	1	0	1
2020-04-12	4	1	5
2020-04-13	2	0	2
2020-04-14	1	0	1
2020-04-15	1	0	1
2020-04-16	1	0	1
2020-04-17	1	0	1
2020-04-18	1	0	1
2020-04-20	2	1	3
2020-04-21	3	0	3
2020-04-22	1	0	1
2020-04-24	0	1	1
2020-04-27	3	0	3
2020-04-30	1	0	1
2020-05-01	1	0	1
2020-05-02	1	1	2
2020-05-04	0	1	1
2020-05-05	1	0	1
2020-05-06	1	0	1
2020-05-08	1	1	2
2020-05-14	1	0	1
2020-05-15	1	1	2
2020-05-18	0	1	1
2020-05-19	1	0	1
2020-05-23	1	0	1

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-06-03	1	0	1

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

Day	Wales
2020-01-27	1
2020-02-27	1
2020-03-01	2
2020-03-04	1
2020-03-07	2
2020-03-08	2
2020-03-09	1
2020-03-10	5
2020-03-11	10
2020-03-12	7
2020-03-13	8
2020-03-14	10
2020-03-15	15
2020-03-16	22
2020-03-17	32
2020-03-18	25
2020-03-19	30
2020-03-20	12
2020-03-23	1
2020-03-24	22
2020-03-25	94
2020-03-26	18
2020-03-27	29
2020-03-28	17
2020-03-29	22
2020-03-30	75
2020-03-31	144
2020-04-01	136
2020-04-02	99
2020-04-03	112
2020-04-04	139
2020-04-05	65
2020-04-06	167
2020-04-07	185
2020-04-08	126
2020-04-09	83
2020-04-10	120
2020-04-11	73
2020-04-12	87
2020-04-13	77
2020-04-14	123
2020-04-15	80
2020-04-16	73
2020-04-17	47
2020-04-18	43
2020-04-19	36
2020-04-20	68
2020-04-21	31
2020-04-22	18
2020-04-23	31
2020-04-24	70
2020-04-25	40
2020-04-26	19
2020-04-27	70
2020-04-28	51
2020-04-29	52
2020-04-30	50

Day	Wales
2020-05-01	46
2020-05-02	51
2020-05-03	19
2020-05-04	40
2020-05-05	34
2020-05-06	53
2020-05-07	53
2020-05-08	30
2020-05-09	41
2020-05-10	38
2020-05-11	63
2020-05-12	44
2020-05-13	53
2020-05-14	28
2020-05-15	31
2020-05-16	14
2020-05-17	11
2020-05-18	18
2020-05-19	27
2020-05-20	38
2020-05-21	24
2020-05-22	24
2020-05-23	21
2020-05-24	18
2020-05-25	16
2020-05-26	25
2020-05-27	29
2020-05-28	19
2020-05-29	20
2020-05-30	14
2020-05-31	10
2020-06-01	6
2020-06-02	1
2020-06-03	7
2020-06-04	6
2020-06-05	1
2020-06-06	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
ANGLESEY	Wales	71	50-100
BLAENAU GWENT	Wales	60	50-100
BRIDGEND	Wales	115	100-150
CAERPHILLY	Wales	142	100-150
CARDIFF	Wales	562	>500
CARMARTHENSHIRE	Wales	143	100-150
CEREDIGION	Wales	16	10-50
CONWY	Wales	143	100-150
DENBIGHSHIRE	Wales	168	150-200
FLINTSHIRE	Wales	120	100-150
GWYNEDD	Wales	113	100-150
MERTHYR TYDFIL	Wales	91	50-100
MONMOUTHSHIRE	Wales	74	50-100
NEATH PORT TALBOT	Wales	114	100-150
NEWPORT	Wales	164	150-200
PEMBROKESHIRE	Wales	70	50-100
POWYS	Wales	77	50-100
SWANSEA	Wales	269	250-300
TORFAEN	Wales	91	50-100
VALE OF GLAMORGAN	Wales	187	150-200
WREXHAM	Wales	149	100-150