

Lineages report for Wales

This report gives summaries of lineages sampled in Wales 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 from Wales have been included in this analysis. 221 lineages have been recorded, 119 of which only contain one sequence.

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.

185 are lineages which were sampled less than five times in Wales, 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 36 that remain: 8 are pending extinction, ie last seen three weeks ago. 12 lineages have gone quiet, ie haven't been seen this week. 5 lineages have reactivated. 11 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	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK5	Mar-01, Jun-06	821	B.1.1.p11, B.1.1, B.1.1.p16, B.1.1.1, B.1.1.16	0	active today
UK61	Mar-08, May-27	397	B.3	10	0.0183
UK42	Feb-27, Jun-04	371	B.1.71, B.1, B.1.67, B.1.35, B.1.p11	2	0.0466
UK158	Mar-20, Jun-03	301	B.1.1.2, B.1.1	3	0.0751
UK3021	Mar-29, Jun-06	217	B.1	0	active today
UK632	Mar-25, Jun-04	205	B.1.1	2	0.149
UK167	Jan-27, Jun-04	115	B.1	2	0.1688
UK5741	Mar-17, Jun-02	111	B.1.44, B.1	4	0.0784
UK2464	Mar-26, Jun-06	80	B.1.p11	0	active today
UK394	Mar-17, May-24	79	B.1.1, B.1.1.10	13	0.0513

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.

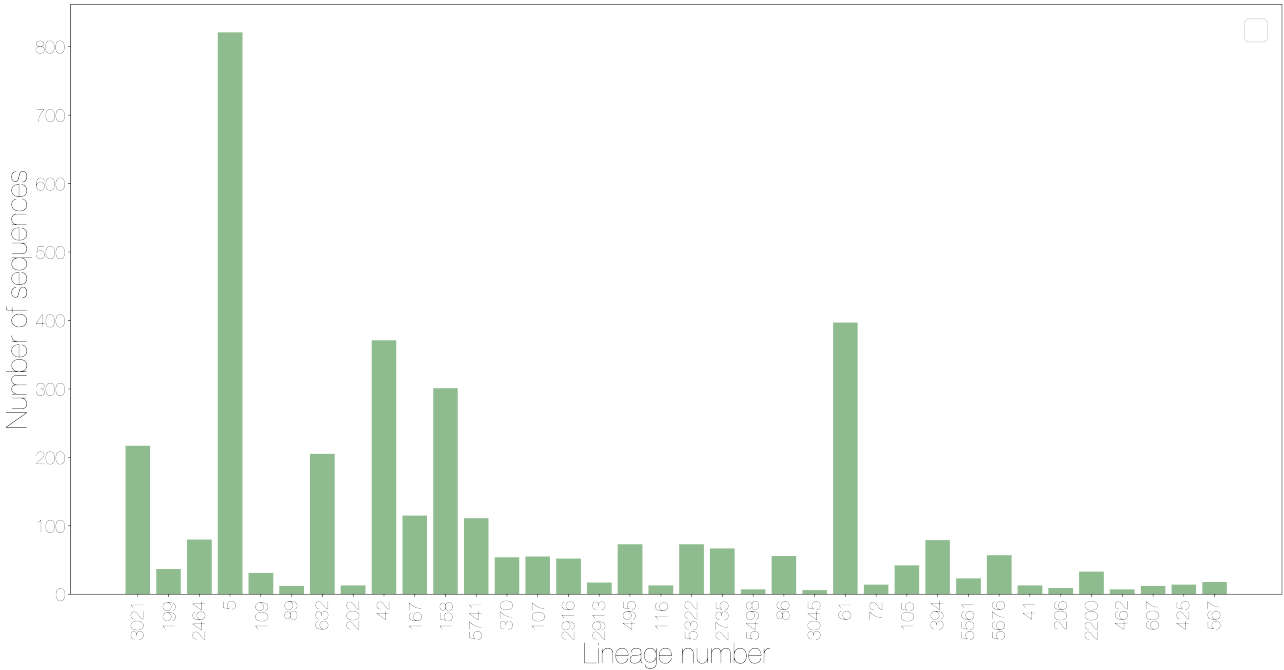


Figure 1: Number of sequences sampled in a lineage by country

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 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.

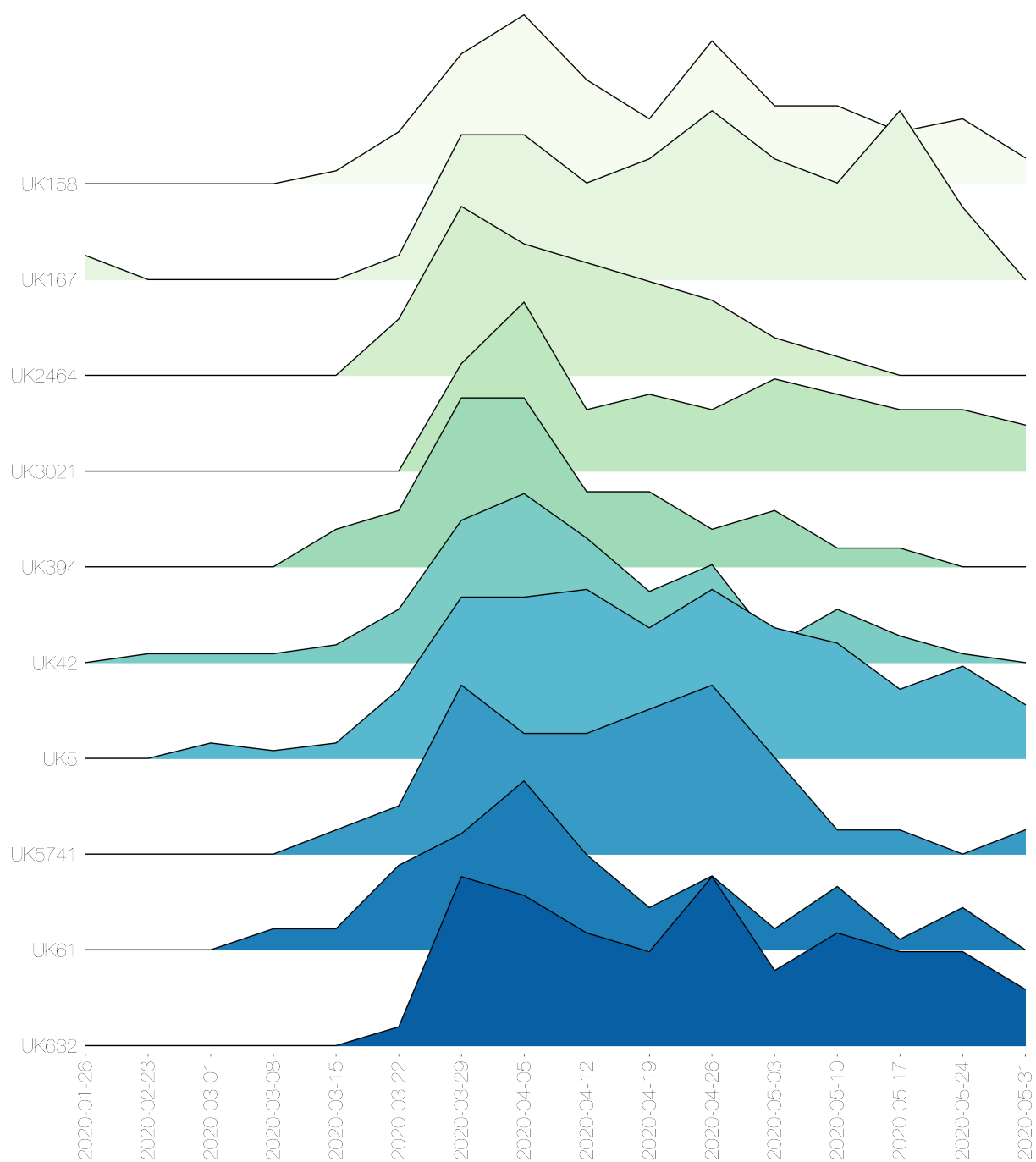


Figure 2: Lineages by number of adm2 regions present by epiweek

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.

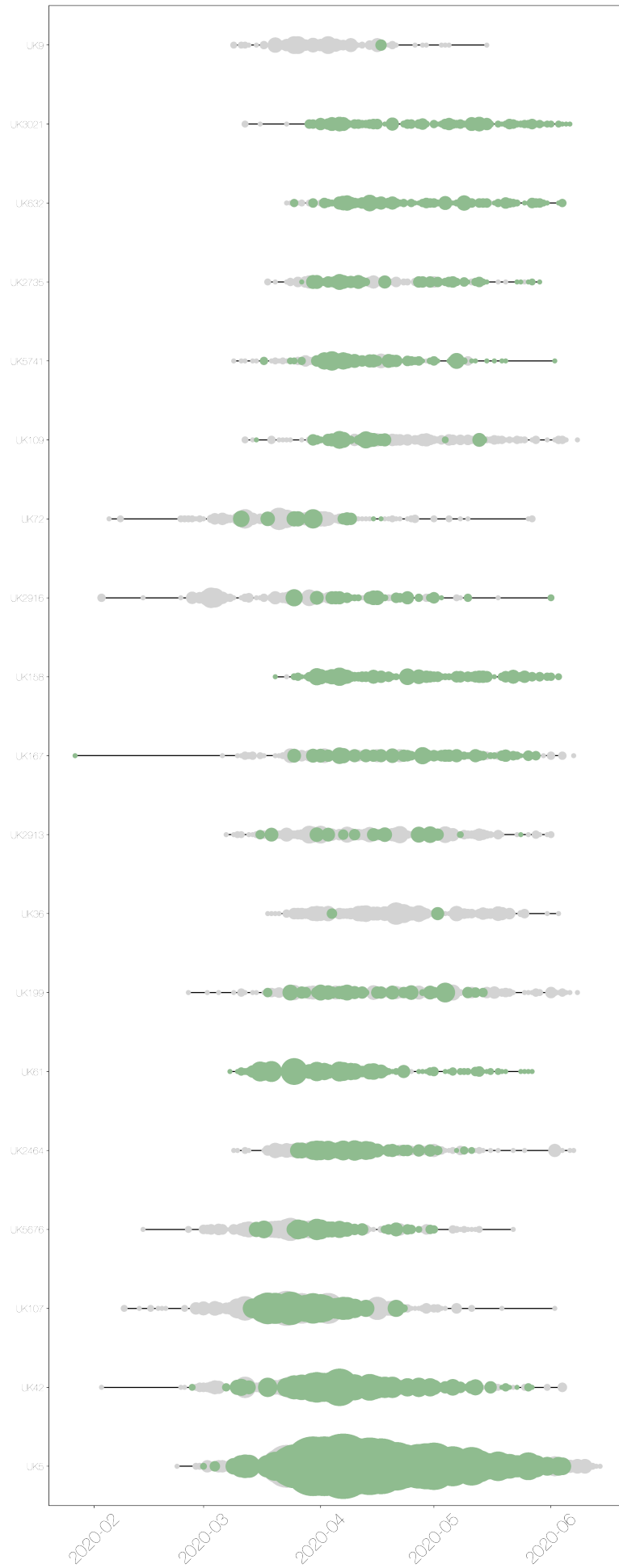


Figure 3: Timeline of lineages, sized by number of sequences from each country.

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

NB the lineage may have started anywhere in the UK, but has been recorded at least once in Wales

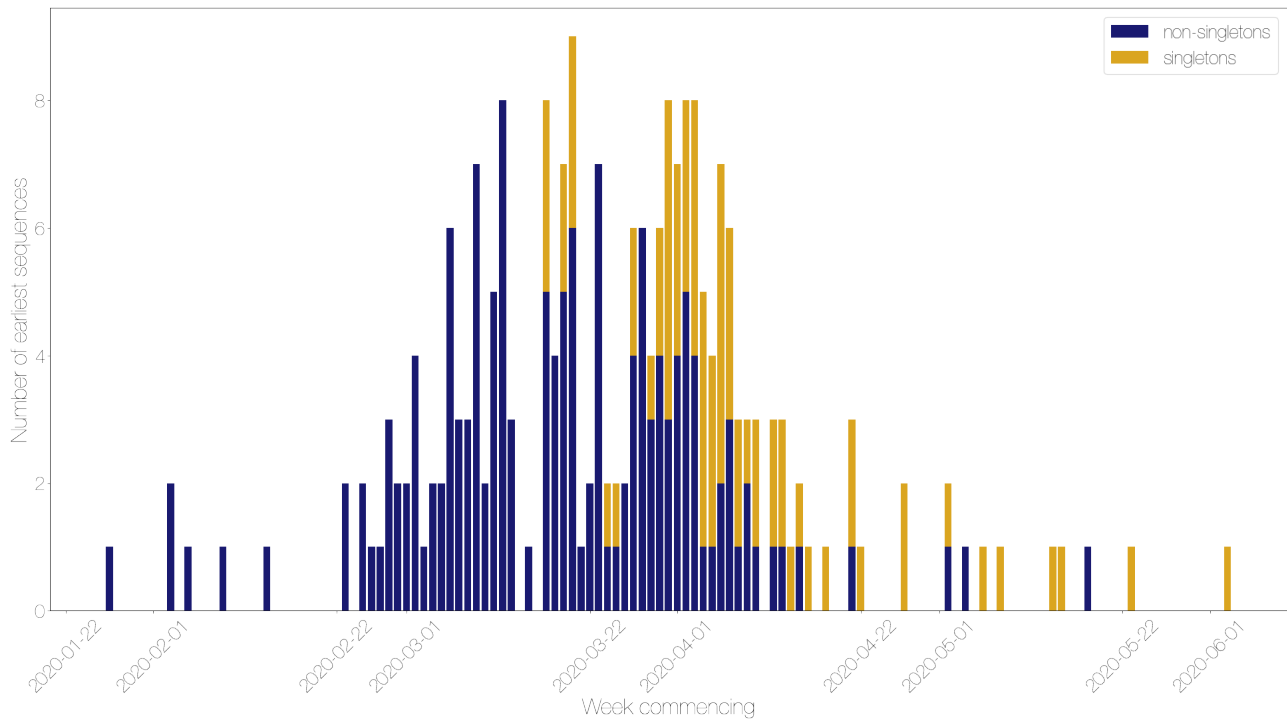


Figure 4: Lineage starts per week, split by singletons and non-singletons

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.

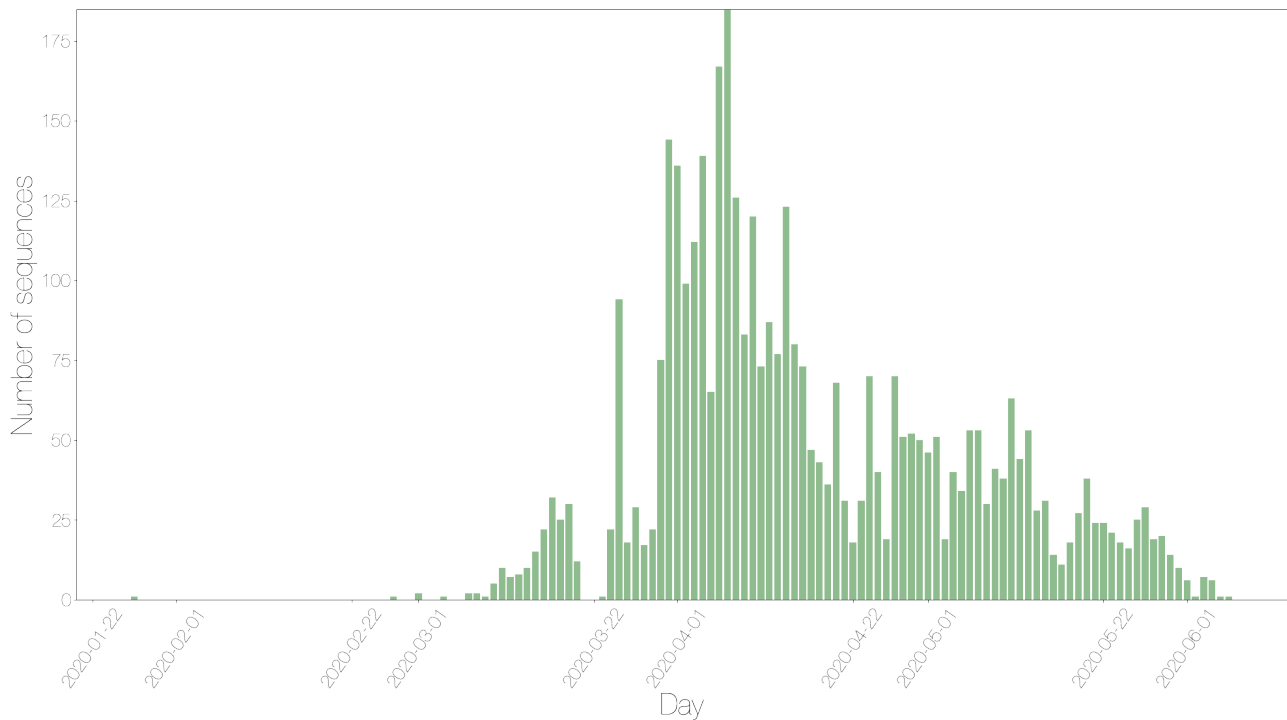


Figure 5: Sequences taken on each day by country

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.

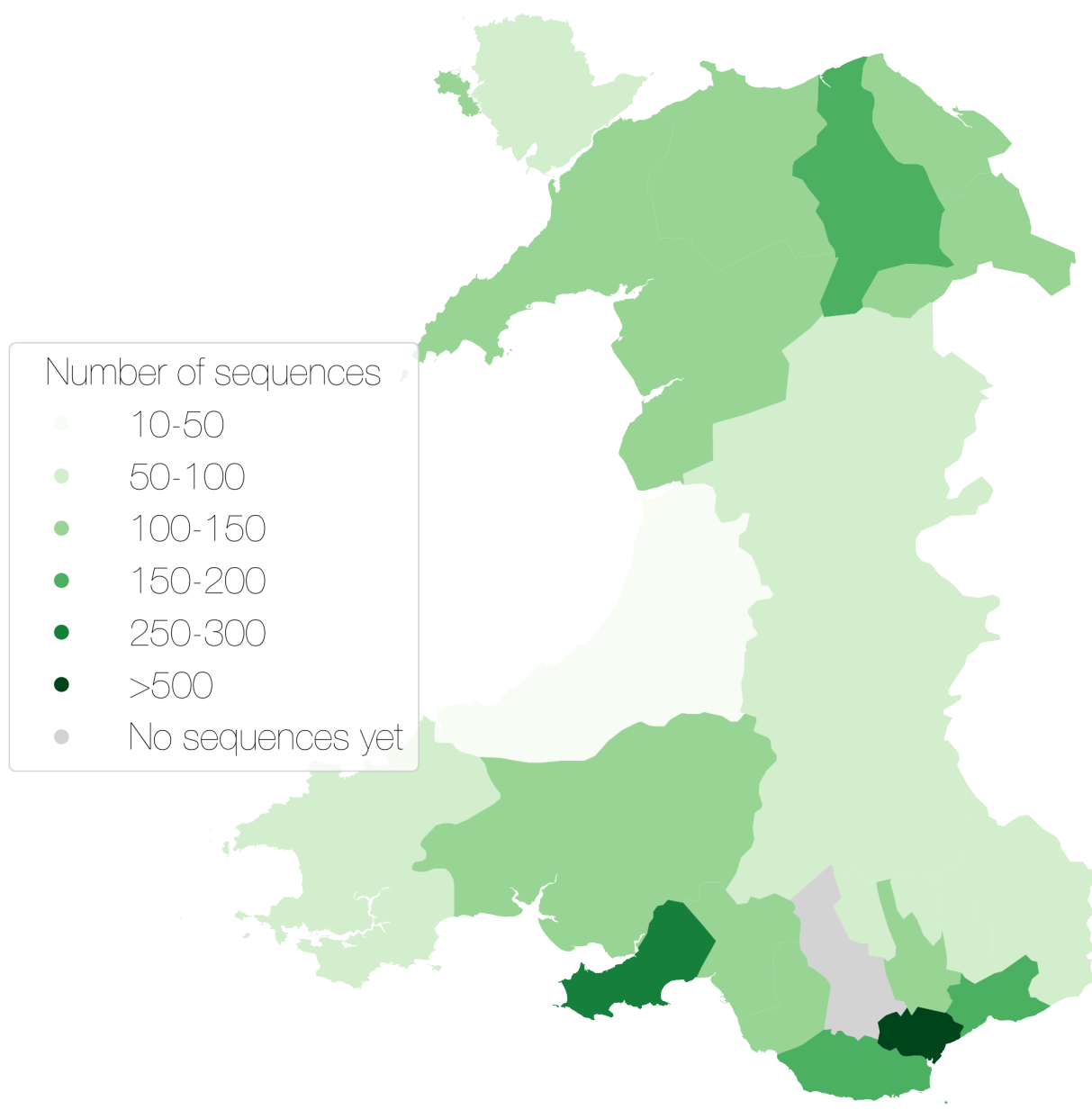


Figure 6: Map showing the number of sequences sampled by adm2 region

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	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK5	Mar-01, Jun-06	821	B.1.1.p11, B.1.1, B.1.1.p16, B.1.1.1, B.1.1.16	0	active today
UK61	Mar-08, May-27	397	B.3	10	0.0183
UK42	Feb-27, Jun-04	371	B.1.71, B.1, B.1.67, B.1.35, B.1.p11	2	0.0466
UK158	Mar-20, Jun-03	301	B.1.1.2, B.1.1	3	0.0751
UK3021	Mar-29, Jun-06	217	B.1	0	active today
UK632	Mar-25, Jun-04	205	B.1.1	2	0.149
UK167	Jan-27, Jun-04	115	B.1	2	0.1688
UK5741	Mar-17, Jun-02	111	B.1.44, B.1	4	0.0784
UK2464	Mar-26, Jun-06	80	B.1.p11	0	active today
UK394	Mar-17, May-24	79	B.1.1, B.1.1.10	13	0.0513
UK5322	Apr-08, May-30	73	B.1.1	7	0.1262
UK495	Apr-01, May-30	73	B.1.p11	7	0.1155
UK2735	Mar-27, May-29	67	B.1.1	8	0.0366
UK5676	Mar-15, May-22	57	B.3, B.2	15	0.0124
UK86	Mar-30, May-28	56	B.1	9	0.1279
UK107	Mar-14, Jun-02	55	B.2.1	4	0.0221
UK370	Mar-19, Jun-02	54	B.1.1.10	4	0.1264
UK2916	Mar-25, Jun-01	52	B.1	5	0.0735
UK105	Apr-04, May-26	42	B.1.p11	11	0.1299
UK199	Mar-18, Jun-06	37	B.1.5, B.1	0	active today
UK2200	Mar-15, May-20	33	B.1.5.6, B.1.5	17	0.0513
UK109	Mar-15, Jun-05	31	B.1.5	1	0.2848
UK187	Mar-30, Apr-30	29	B.1	37	0.0223
UK5561	Mar-18, May-24	23	B.2.2	13	0.0466
UK179	Mar-17, May-07	23	B.1.1, B.1.1.p11	30	0.0447
UK567	Mar-30, May-15	18	B.2.2	22	0.0979

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK2913	Mar-16, Jun-01	17	B.1, B.1.p11	5	0.0416
UK198	Mar-26, May-07	15	B.1.5	30	0.0288
UK339	Mar-14, Apr-16	15	B.3	51	0.0132
UK72	Mar-11, May-27	14	B	10	0.0348
UK425	Mar-28, May-15	14	B.1.1	22	0.1283
UK202	Apr-24, Jun-04	13	B.1.1	2	1.9545
UK116	May-08, May-30	13	B.1	7	0.4786
UK41	Apr-10, May-21	13	B.1	16	0.122
UK89	Apr-10, Jun-05	12	B.1.1.9	1	1.434
UK317	Mar-19, Apr-20	12	B.3	47	0.0337
UK64	Mar-25, May-05	12	B.1	32	0.0433
UK607	Mar-11, May-18	12	B	19	0.0633
UK15	Mar-17, May-06	11	B.1.1	31	0.0133
UK801	Apr-05, May-05	10	B.1	32	0.1042
UK206	Apr-02, May-20	9	B.1	17	0.3529
UK696	Apr-10, May-01	9	B.1.5, B.1	36	0.0729
UK275	Mar-31, Apr-27	8	B.1.13	40	0.0266
UK633	Apr-03, Apr-28	8	B.1.1.p16, B.1.1.16	39	0.0916
UK491	Mar-31, Apr-14	8	B, B.2.1	53	0.0345
UK462	Apr-01, May-18	7	B.1	19	0.1076
UK119	Mar-30, Apr-24	7	B.2.5	43	0.0301
UK5498	Apr-01, May-28	7	B.2	9	0.1111
UK506	Apr-02, Apr-20	7	B.1.1	47	0.0479
UK3045	May-15, May-27	6	B.1.1, B.1.1.p11	10	0.7
UK451	Mar-25, Apr-05	6	B.2.1	62	0.043

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

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-04-22	1	0	1
2020-04-27	2	0	2
2020-05-02	1	1	2
2020-05-04	0	1	1
2020-05-06	1	0	1
2020-05-08	1	0	1
2020-05-14	1	0	1
2020-05-15	1	0	1
2020-05-18	0	1	1
2020-05-23	1	0	1
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	563	>500
CARMARTHENSHIRE	Wales	144	100-150
CEREDIGION	Wales	16	10-50
CONWY	Wales	143	100-150
DENBIGHSHIRE	Wales	168	150-200
FLINTSHIRE	Wales	120	100-150
GWYNEDD	Wales	115	100-150
MERTHYR TYDFIL	Wales	91	50-100
MONMOUTHSHIRE	Wales	77	50-100
NEATH PORT TALBOT	Wales	114	100-150
NEWPORT	Wales	164	150-200
PEMBROKESHIRE	Wales	70	50-100
POWYS	Wales	77	50-100
RHONDDA, CYNON, TAFF	Wales	0	0
SWANSEA	Wales	269	250-300
TORFAEN	Wales	91	50-100
VALE OF GLAMORGAN	Wales	187	150-200
WREXHAM	Wales	149	100-150