

Lineages report for Wales

This report gives summaries of lineages sampled in Wales for week 2020-07-03. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-06-28. The analysis (eg time since last sample) is therefore undertaken from this date. 4129 sequences from Wales have been included in this analysis. 216 lineages have been recorded, 122 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 172 and the maximum is 1319

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

191 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 25 that remain: 17 are pending extinction, ie last seen three weeks ago. 4 lineages have gone quiet, ie haven't been seen this week. 2 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	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK5	Mar-01, Jun-27	1168	B.1.1.16, B.1.1.p16, B.1.1.p11, B.1.1, B.1.1.1, B.1.1.2	1	0.0152
UK61	Mar-08, May-27	419	B.3, B	32	0.0056
UK42	Feb-27, Jun-21	368	B.1.35, B.1, B.1.p11, B.1.71	7	0.0123
UK632	Mar-25, Jun-10	232	B.1.1	18	0.0161
UK3021	Mar-29, Jun-09	225	B.1	19	0.019
UK495	Apr-01, Jun-03	124	B.1.p11	25	0.0214
UK5741	Mar-17, Jun-12	104	B.1.44, B.1	16	0.0201
UK822	Apr-14, Jun-11	102	B.1	17	0.0334
UK5322	Apr-08, Jun-04	86	B.1.1	24	0.0328
UK605	Mar-17, May-24	79	B.1.1, B.1.1.10	35	0.0182

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.

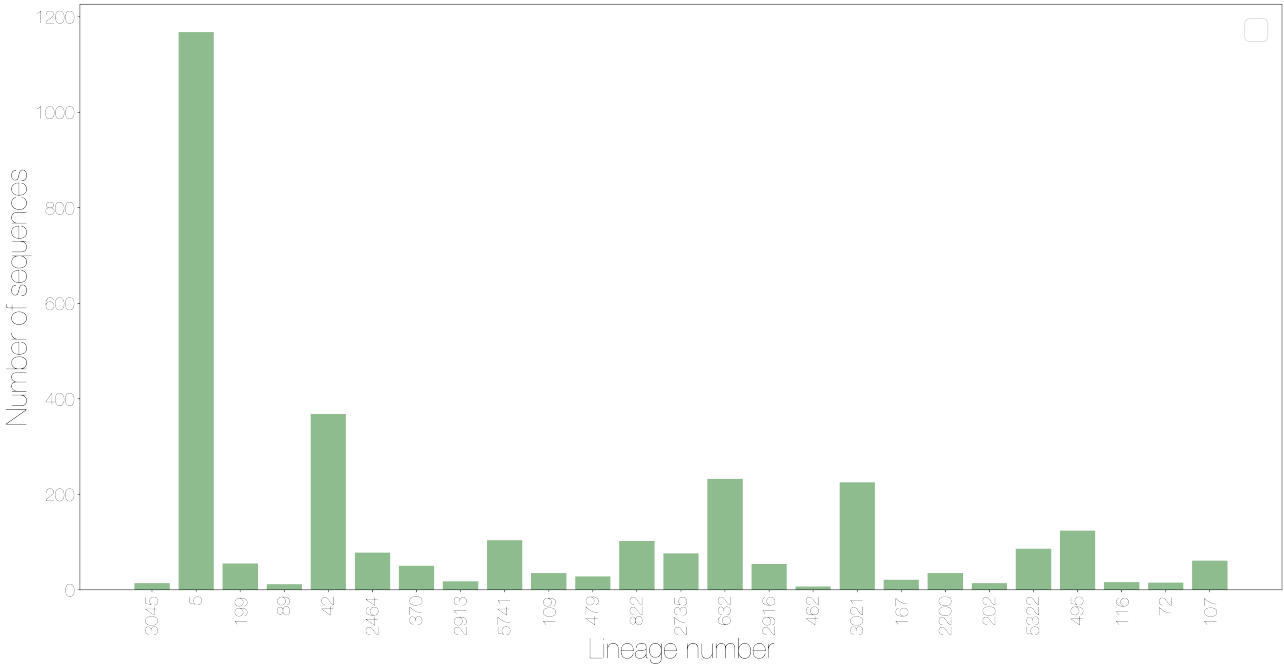


Figure 1: Number of sequences sampled in a lineage by country

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.

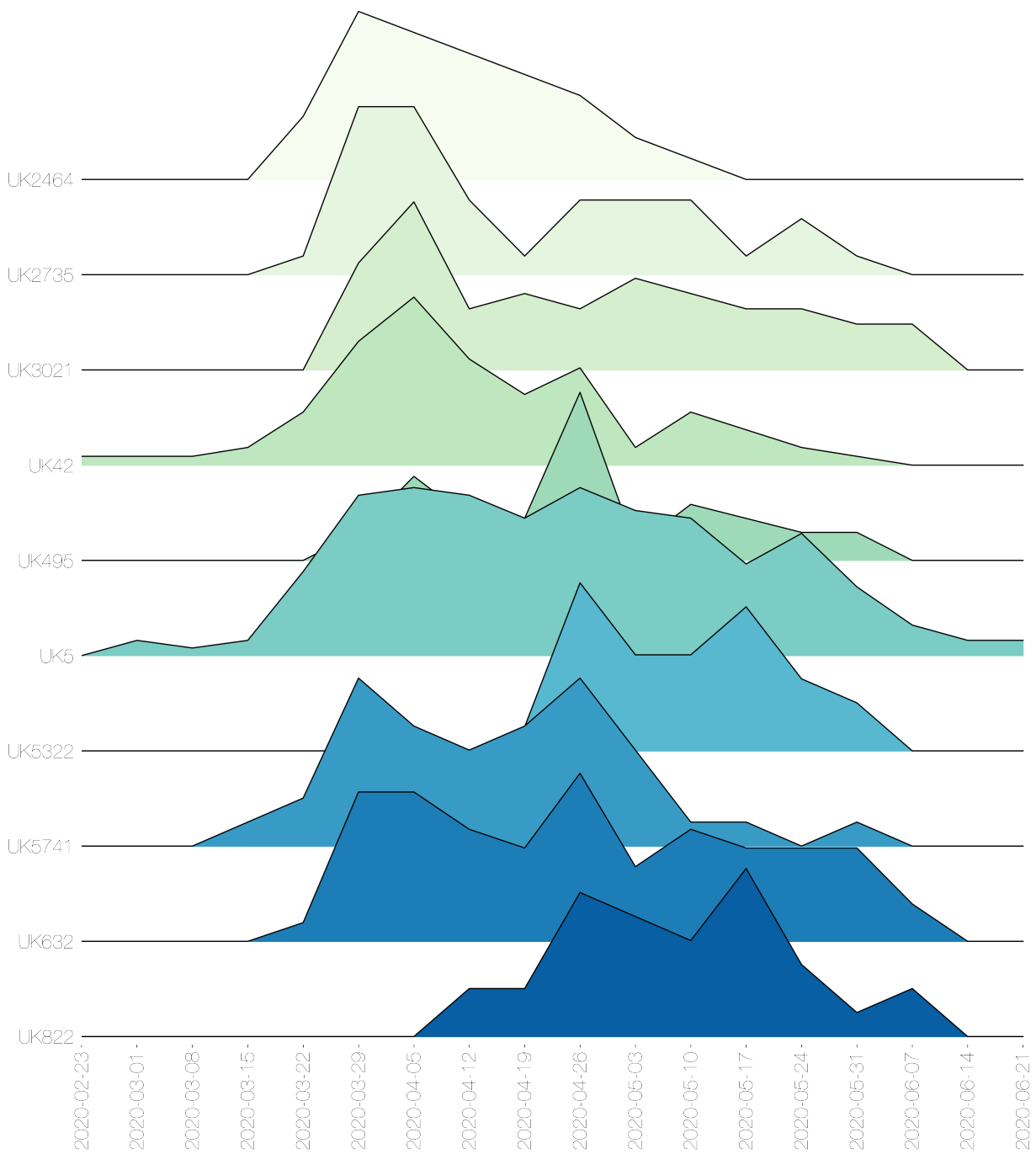


Figure 2: Lineages by number of adm2 regions present by epiweek

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.

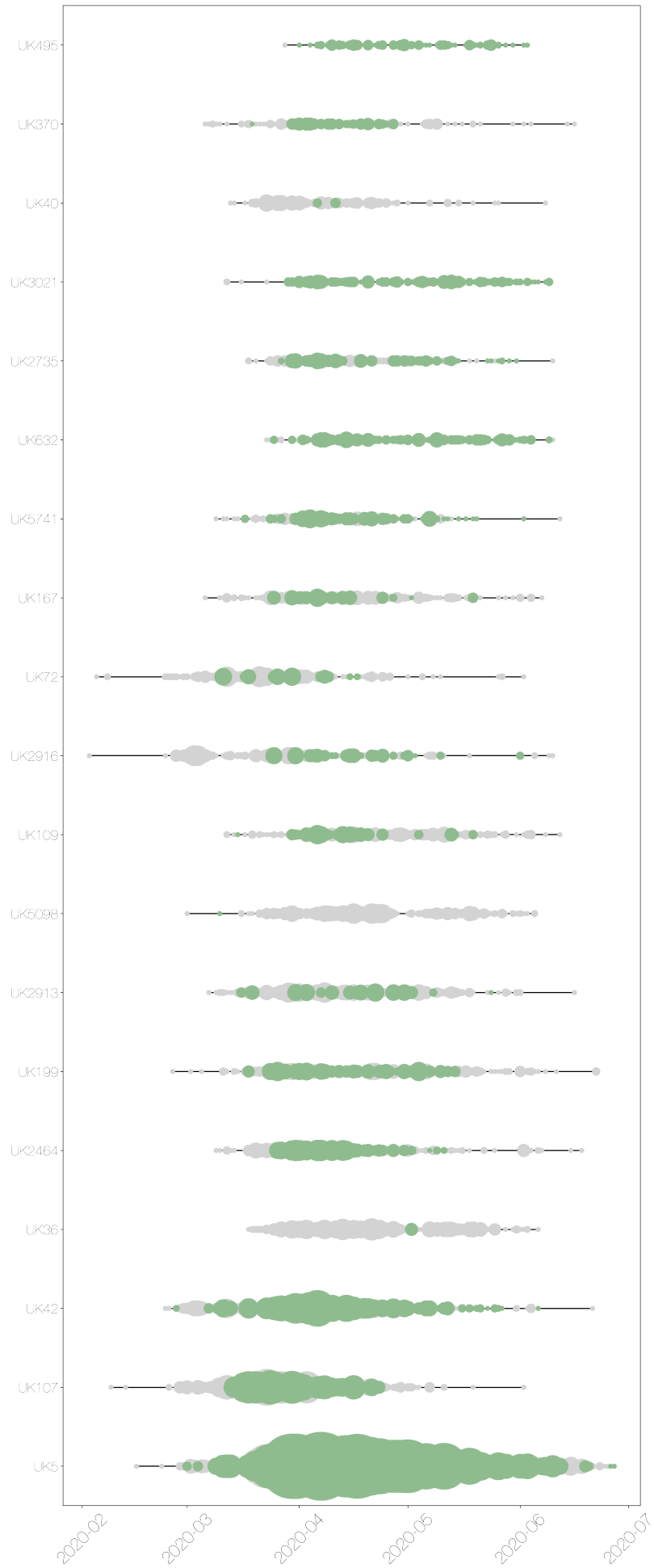


Figure 3: Timeline of lineages, sized by number of sequences from each country.

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

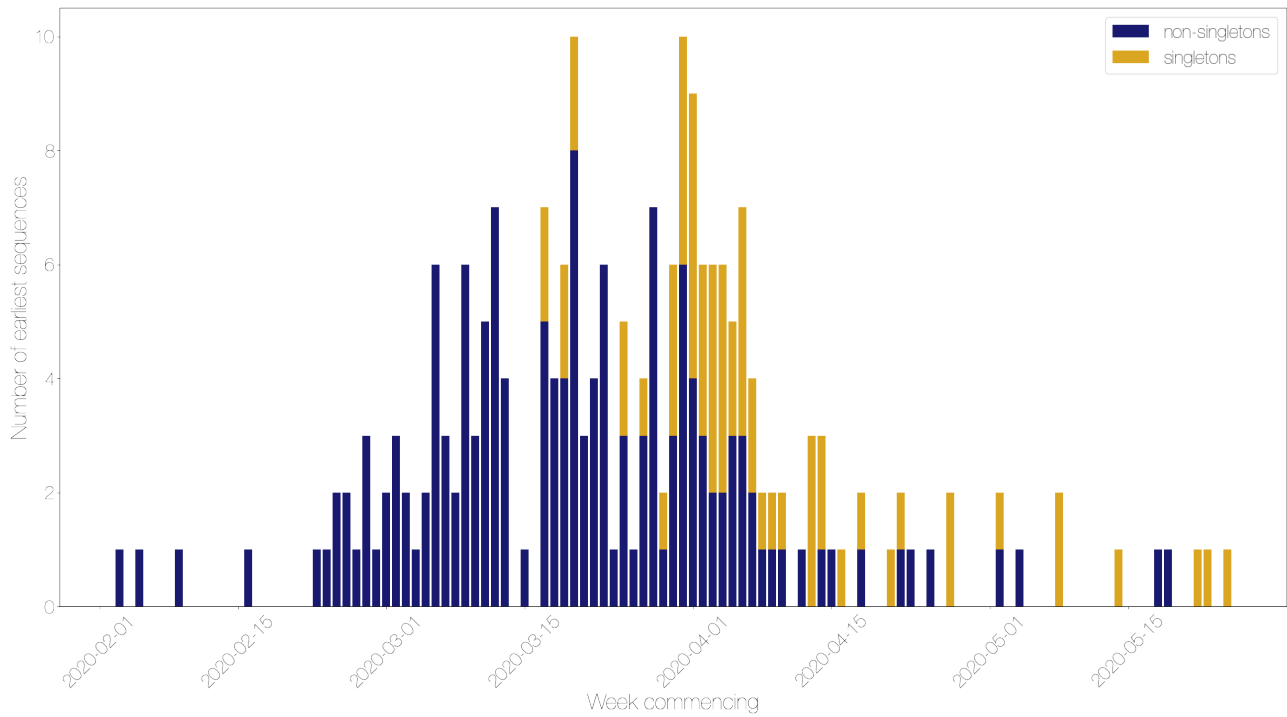


Figure 4: Lineage starts per week, split by singletons and non-singletons

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.

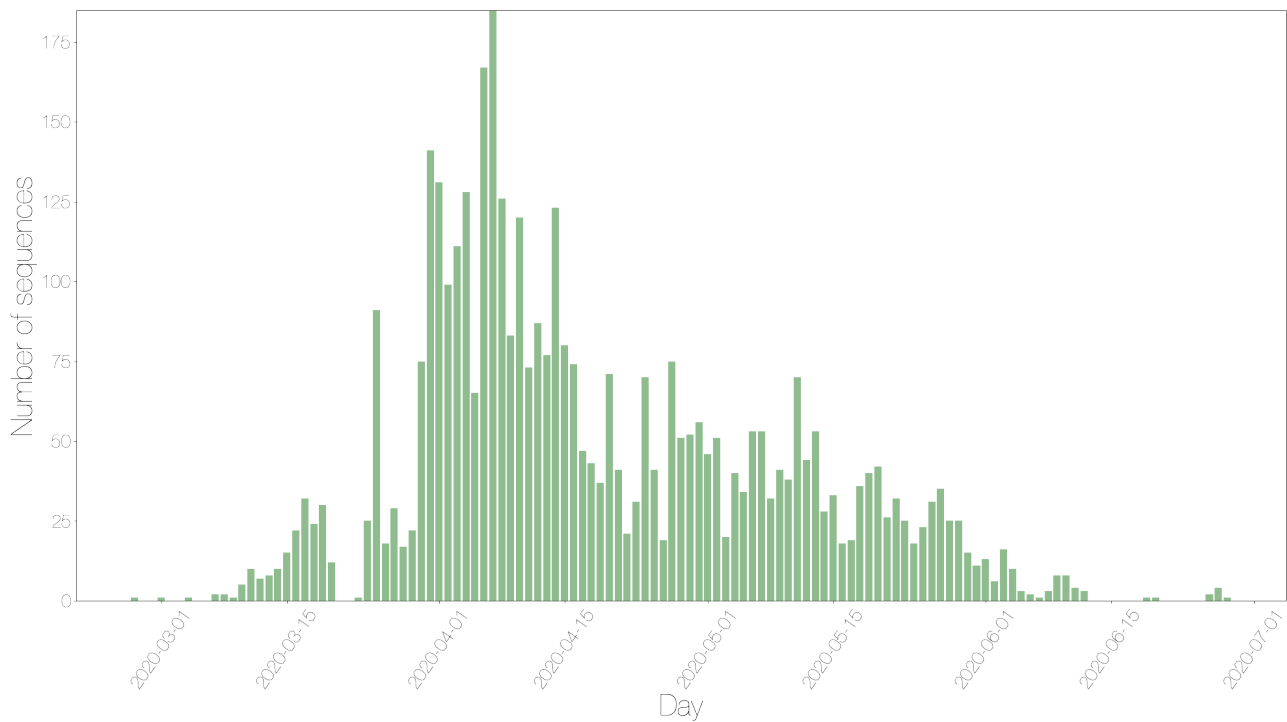


Figure 5: Sequences taken on each day by country

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.

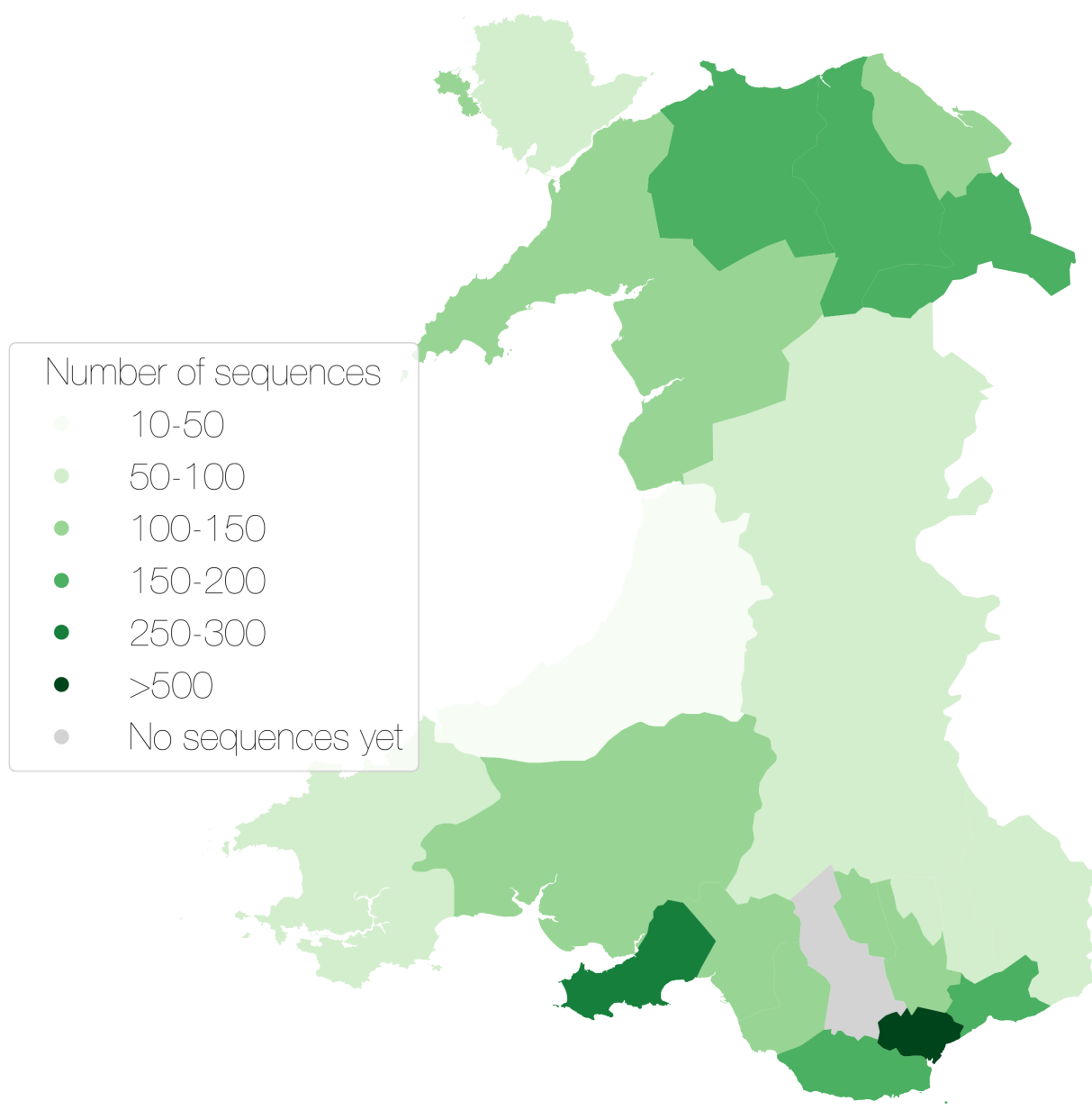


Figure 6: Map showing the number of sequences sampled by adm2 region

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-27	1168	B.1.1.16, B.1.1.p16, B.1.1.p11, B.1.1, B.1.1.1, B.1.1.2	1	0.0152
UK61	Mar-08, May-27	419	B.3, B	32	0.0056
UK42	Feb-27, Jun-21	368	B.1.35, B.1, B.1.p11, B.1.71	7	0.0123
UK632	Mar-25, Jun-10	232	B.1.1	18	0.0161
UK3021	Mar-29, Jun-09	225	B.1	19	0.019
UK495	Apr-01, Jun-03	124	B.1.p11	25	0.0214
UK5741	Mar-17, Jun-12	104	B.1.44, B.1	16	0.0201
UK822	Apr-14, Jun-11	102	B.1	17	0.0334
UK5322	Apr-08, Jun-04	86	B.1.1	24	0.0328
UK605	Mar-17, May-24	79	B.1.1, B.1.1.10	35	0.0182
UK2464	Mar-26, Jun-18	78	B.1.p11	10	0.0182
UK2735	Mar-27, Jun-10	76	B.1.1	18	0.0172
UK86	Mar-30, May-30	61	B.1	29	0.0424
UK107	Mar-14, Jun-02	61	B.2.1, B	26	0.0031
UK199	Mar-18, Jun-22	55	B.1, B.1.5	6	0.0358
UK2916	Mar-25, Jun-10	54	B.1	18	0.0189
UK5676	Mar-15, May-27	54	B.2	32	0.005
UK370	Mar-19, Jun-16	50	B.1.1.10	12	0.0447
UK2200	Mar-15, Jun-06	35	B.1.5.6, B.1.5	22	0.0421
UK109	Mar-15, Jun-12	35	B.1.5	16	0.0145
UK187	Mar-30, Apr-30	29	B.1	59	0.0133
UK479	Apr-07, Jun-12	28	B.1.1	16	0.1288
UK5561	Mar-18, May-24	23	B.2.2	35	0.0128
UK600	Apr-01, May-26	22	B.1.1	33	0.0667
UK167	Mar-25, Jun-07	21	B.1	21	0.0139
UK179	Mar-17, May-07	20	B.1.1.p11	52	0.0272

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK567	Mar-30, May-15	20	B.2.2	44	0.0439
UK206	Apr-02, May-20	19	B.1	39	0.0684
UK2913	Mar-16, Jun-16	18	B.1, B.1.p11	12	0.0159
UK116	May-08, Jun-02	16	B.1	26	0.1122
UK695	Mar-25, Apr-12	16	B.1.67	77	0.0156
UK72	Mar-11, Jun-02	15	B	26	0.0133
UK202	Apr-24, Jun-04	14	B.1.1	24	0.1493
UK425	Mar-28, May-15	14	B.1.1	44	0.0642
UK3045	May-15, Jun-28	14	B.1.1, B.1.1.p11	0	active today
UK64	Mar-25, May-05	12	B.1	54	0.0238
UK89	Apr-10, Jun-22	12	B.1.1.9	6	0.2246
UK317	Mar-19, Apr-20	12	B.3	69	0.022
UK607	Mar-11, May-18	12	B	41	0.0272
UK15	Mar-17, May-06	11	B.1.1	53	0.0069
UK327	Apr-05, May-05	10	B.1	54	0.0617
UK275	Mar-31, Apr-27	8	B.1.13	62	0.0152
UK633	Apr-03, Apr-28	8	B.1.1.p16, B.1.1.16	61	0.0585
UK696	Apr-10, May-01	8	B.1.5, B.1	58	0.0517
UK5498	Apr-01, May-28	7	B.2	31	0.0297
UK119	Mar-30, Apr-24	7	B.2.5	65	0.0188
UK462	Apr-01, Jun-09	7	B.1	19	0.1171
UK451	Mar-25, Apr-05	6	B.2.1	84	0.0317

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	5

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

Week commencing	UK5	UK42	UK632	UK3021	UK495	UK5741	UK822	UK5322	UK2464	UK2735
2020-02-23	0	1	0	0	0	0	0	0	0	0
2020-03-01	2	1	0	0	0	0	0	0	0	0
2020-03-08	1	1	0	0	0	0	0	0	0	0
2020-03-15	2	2	0	0	0	1	0	0	0	0
2020-03-22	11	6	1	0	0	2	0	0	3	1
2020-03-29	21	14	8	7	2	7	0	0	8	9
2020-04-05	22	19	8	11	6	5	0	1	7	9
2020-04-12	21	12	6	4	3	4	2	0	6	4
2020-04-19	18	8	5	5	3	5	2	1	5	1
2020-04-26	22	11	9	4	12	7	6	7	4	4
2020-05-03	19	2	4	6	1	4	5	4	2	4
2020-05-10	18	6	6	5	4	1	4	4	1	4
2020-05-17	12	4	5	4	3	1	7	6	0	1
2020-05-24	16	2	5	4	2	0	3	3	0	3
2020-05-31	9	1	5	3	2	1	1	2	0	1
2020-06-07	4	0	2	3	0	0	2	0	0	0
2020-06-14	2	0	0	0	0	0	0	0	0	0
2020-06-21	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	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	1	1
2020-02-25	0	2	2
2020-02-26	0	2	2
2020-02-27	0	1	1
2020-02-28	0	3	3
2020-02-29	0	1	1
2020-03-01	0	2	2
2020-03-02	0	3	3
2020-03-03	0	2	2
2020-03-04	0	1	1
2020-03-05	0	2	2
2020-03-06	0	6	6
2020-03-07	0	3	3
2020-03-08	0	2	2
2020-03-09	0	6	6
2020-03-10	0	3	3
2020-03-11	0	5	5
2020-03-12	0	7	7
2020-03-13	0	4	4
2020-03-15	0	1	1
2020-03-17	2	5	7
2020-03-18	0	4	4
2020-03-19	2	4	6
2020-03-20	2	8	10
2020-03-21	0	3	3
2020-03-22	0	4	4
2020-03-23	0	6	6
2020-03-24	0	1	1
2020-03-25	2	3	5
2020-03-26	0	1	1
2020-03-27	1	3	4
2020-03-28	0	7	7
2020-03-29	1	1	2
2020-03-30	3	3	6
2020-03-31	4	6	10
2020-04-01	5	4	9
2020-04-02	3	3	6
2020-04-03	4	2	6
2020-04-04	4	2	6
2020-04-05	2	3	5
2020-04-06	4	3	7
2020-04-07	2	2	4
2020-04-08	1	1	2
2020-04-09	1	1	2
2020-04-10	1	1	2
2020-04-12	0	1	1
2020-04-13	3	0	3
2020-04-14	2	1	3
2020-04-15	0	1	1
2020-04-16	1	0	1
2020-04-18	1	1	2
2020-04-21	1	0	1

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-04-22	1	1	2
2020-04-23	0	1	1
2020-04-25	0	1	1
2020-04-27	2	0	2
2020-05-02	1	1	2
2020-05-04	0	1	1
2020-05-08	2	0	2
2020-05-14	1	0	1
2020-05-18	0	1	1
2020-05-19	0	1	1
2020-05-22	1	0	1
2020-05-23	1	0	1
2020-05-25	1	0	1

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

Day	Wales
2020-02-27	1
2020-03-01	1
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	24
2020-03-19	30
2020-03-20	12
2020-03-23	1
2020-03-24	25
2020-03-25	91
2020-03-26	18
2020-03-27	29
2020-03-28	17
2020-03-29	22
2020-03-30	75
2020-03-31	141
2020-04-01	131
2020-04-02	99
2020-04-03	111
2020-04-04	128
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	74
2020-04-17	47
2020-04-18	43
2020-04-19	37
2020-04-20	71
2020-04-21	41
2020-04-22	21
2020-04-23	31
2020-04-24	70
2020-04-25	41
2020-04-26	19
2020-04-27	75
2020-04-28	51
2020-04-29	52
2020-04-30	56
2020-05-01	46

Day	Wales
2020-05-02	51
2020-05-03	20
2020-05-04	40
2020-05-05	34
2020-05-06	53
2020-05-07	53
2020-05-08	32
2020-05-09	41
2020-05-10	38
2020-05-11	70
2020-05-12	44
2020-05-13	53
2020-05-14	28
2020-05-15	33
2020-05-16	18
2020-05-17	19
2020-05-18	36
2020-05-19	40
2020-05-20	42
2020-05-21	26
2020-05-22	32
2020-05-23	25
2020-05-24	18
2020-05-25	23
2020-05-26	31
2020-05-27	35
2020-05-28	25
2020-05-29	25
2020-05-30	15
2020-05-31	11
2020-06-01	13
2020-06-02	6
2020-06-03	16
2020-06-04	10
2020-06-05	3
2020-06-06	2
2020-06-07	1
2020-06-08	3
2020-06-09	8
2020-06-10	8
2020-06-11	4
2020-06-12	3
2020-06-19	1
2020-06-20	1
2020-06-26	2
2020-06-27	4
2020-06-28	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	80	50-100
BLAENAU GWENT	Wales	59	50-100
BRIDGEND	Wales	114	100-150
CAERPHILLY	Wales	142	100-150
CARDIFF	Wales	586	>500
CARMARTHENSHIRE	Wales	148	100-150
CEREDIGION	Wales	16	10-50
CONWY	Wales	162	150-200
DENBIGHSHIRE	Wales	194	150-200
FLINTSHIRE	Wales	131	100-150
GWYNEDD	Wales	125	100-150
MERTHYR TYDFIL	Wales	103	100-150
MONMOUTHSHIRE	Wales	88	50-100
NEATH PORT TALBOT	Wales	119	100-150
NEWPORT	Wales	165	150-200
PEMBROKESHIRE	Wales	73	50-100
POWYS	Wales	77	50-100
RHONDDA, CYNON, TAFF	Wales	0	0
SWANSEA	Wales	276	250-300
TORFAEN	Wales	91	50-100
VALE OF GLAMORGAN	Wales	191	150-200
WREXHAM	Wales	166	150-200