UK lineages summary report

This report gives summaries of lineages sampled in Wales for week 2020-05-29. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-05-03. The analysis (eg time since last sample) is therefore undertaken from this date. 2724 sequences from Wales have been included in this analysis. 783 lineages have been recorded, 613 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 784 and the maximum is 1121

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

716 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 67 that remain: 14 are pending extinction, ie last seen three weeks ago. 1 has not been seen for more than one month, and so is viewed as extinct, but will continue to be monitored. 10 lineages have gone quiet, ie haven't been seen this week. 11 lineages have reactivated. 31 lineages have been continuously circulating.

The following table contains information about lineages and the number of sequences the dataset, in reverse size order.

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.

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)
UK61	Mar-10, May-01	340	B.3	2
UK158	Mar-20, May-02	142	B.1.1.2	1
UK5	Mar-04, May-03	125	B.1.1, B.1.1.1	0
UK42	Mar-07, Apr-27	112	B.1, B.1.35	6
UK632	Mar-25, May-02	97	B.1.1	1
UK74	Mar-30, May-03	96	B.1	0
UK19	Mar-17, May-02	81	B.1.44, B.1	1
UK2464	Mar-26, May-02	70	B.1.p11	1

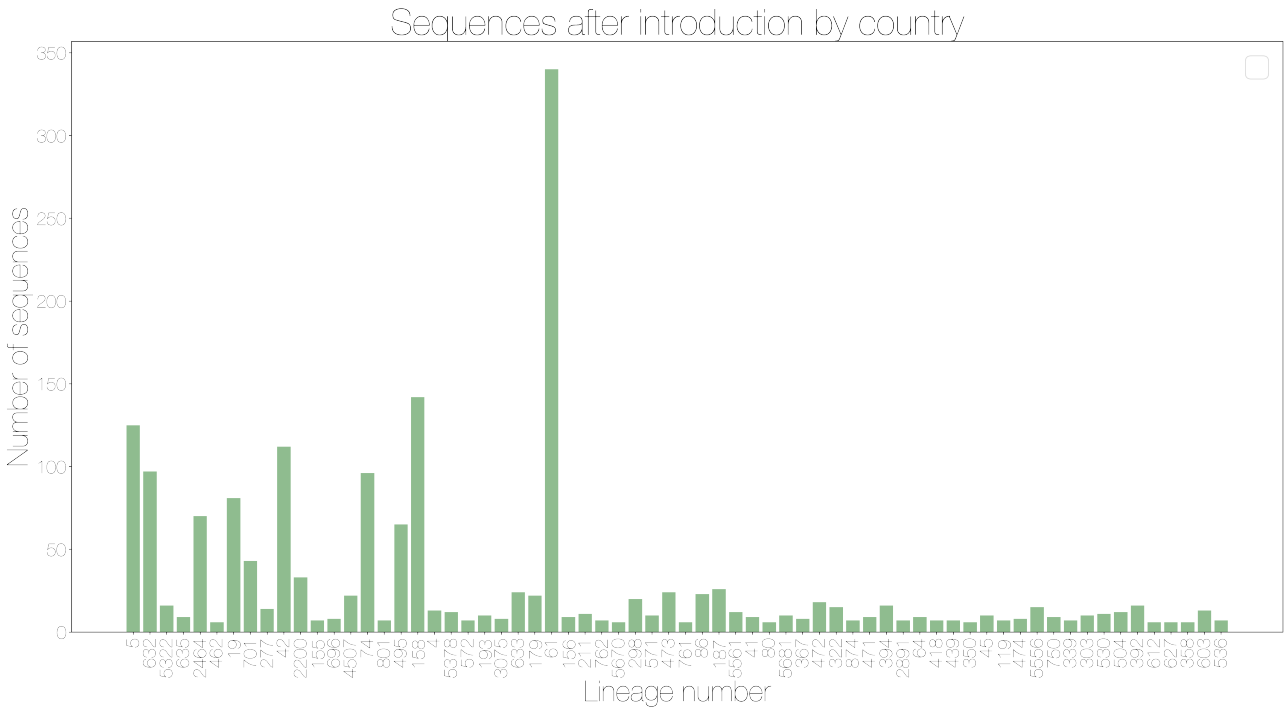
Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)
UK495	Apr-01, May-02	65	B.1.p11	1
UK701	Mar-25, May-01	43	B.1	2
UK2200	Mar-15, Apr-26	33	B.1.5, B.1.5.6	7
UK187	Apr-01, Apr-28	26	B.1	5
UK473	Apr-02, Apr-29	24	B.1.1	4
UK633	Apr-03, May-01	24	B.1.1.16, B.1.1.p16	2
UK86	Mar-30, Apr-28	23	B.1	5
UK4507	Apr-14, May-03	22	B.1	0
UK179	Mar-17, May-01	22	B.1.1.p11	2
UK298	Mar-27, Apr-29	20	B.1.1	4
UK472	Apr-04, Apr-27	18	B.1.1.p11, B.1.1	6
UK392	Mar-25, Apr-12	16	B.1.67	21
UK5322	Apr-08, May-03	16	B.1.1	0
UK394	Mar-24, Apr-24	16	B.1.1.10, B.1.1	9
UK5556	Mar-18, Apr-16	15	B.2.2	17
UK322	Mar-30, Apr-26	15	B.1	7
UK277	Mar-28, Apr-28	14	B.1.1	5
UK603	Mar-29, Apr-09	13	B.1.1	24
UK4	Mar-11, Apr-24	13	B	9
UK504	Mar-30, Apr-13	12	B.1.1	20
UK5378	Apr-02, Apr-25	12	B.1.1	8
UK5561	Mar-30, Apr-27	12	B.2.2	6
UK211	Mar-24, Apr-30	11	B.1.5	3
UK530	Mar-31, Apr-13	11	B.1.1	20

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)
UK571	Apr-06, Apr-29	10	B.1.1	4
UK45	Mar-01, Apr-20	10	B.1.1	13
UK5681	Apr-03, Apr-27	10	B.2	6
UK193	Apr-01, May-01	10	B.1.1	2
UK303	Mar-25, Apr-14	10	B.1.1	19
UK64	Mar-25, Apr-20	9	B.1	13
UK471	Apr-02, Apr-24	9	B.1.1	9
UK750	Apr-07, Apr-15	9	B.1	18
UK41	Apr-10, Apr-27	9	B.1	6
UK635	Apr-07, May-02	9	B.1.1	1
UK156	Mar-28, Apr-30	9	B.1.71	3
UK367	Mar-25, Apr-27	8	B.1	6
UK474	Apr-01, Apr-16	8	B.1.1	17
UK3075	Apr-17, May-01	8	B.1.1	2
UK696	Apr-10, May-01	8	B.1.5, B.1	2
UK572	Apr-07, May-01	7	B.1.1	2
UK339	Mar-25, Apr-14	7	B.3	19
UK2891	Mar-31, Apr-24	7	B.1.1	9
UK874	Apr-06, Apr-24	7	B.1	9
UK801	Apr-05, May-02	7	B.1	1
UK536	Mar-27, Apr-09	7	B.1.1	24
UK119	Mar-30, Apr-14	7	B.2.5	19
UK762	Apr-11, Apr-30	7	B.1.1	3
UK155	Mar-25, May-03	7	B.1	0

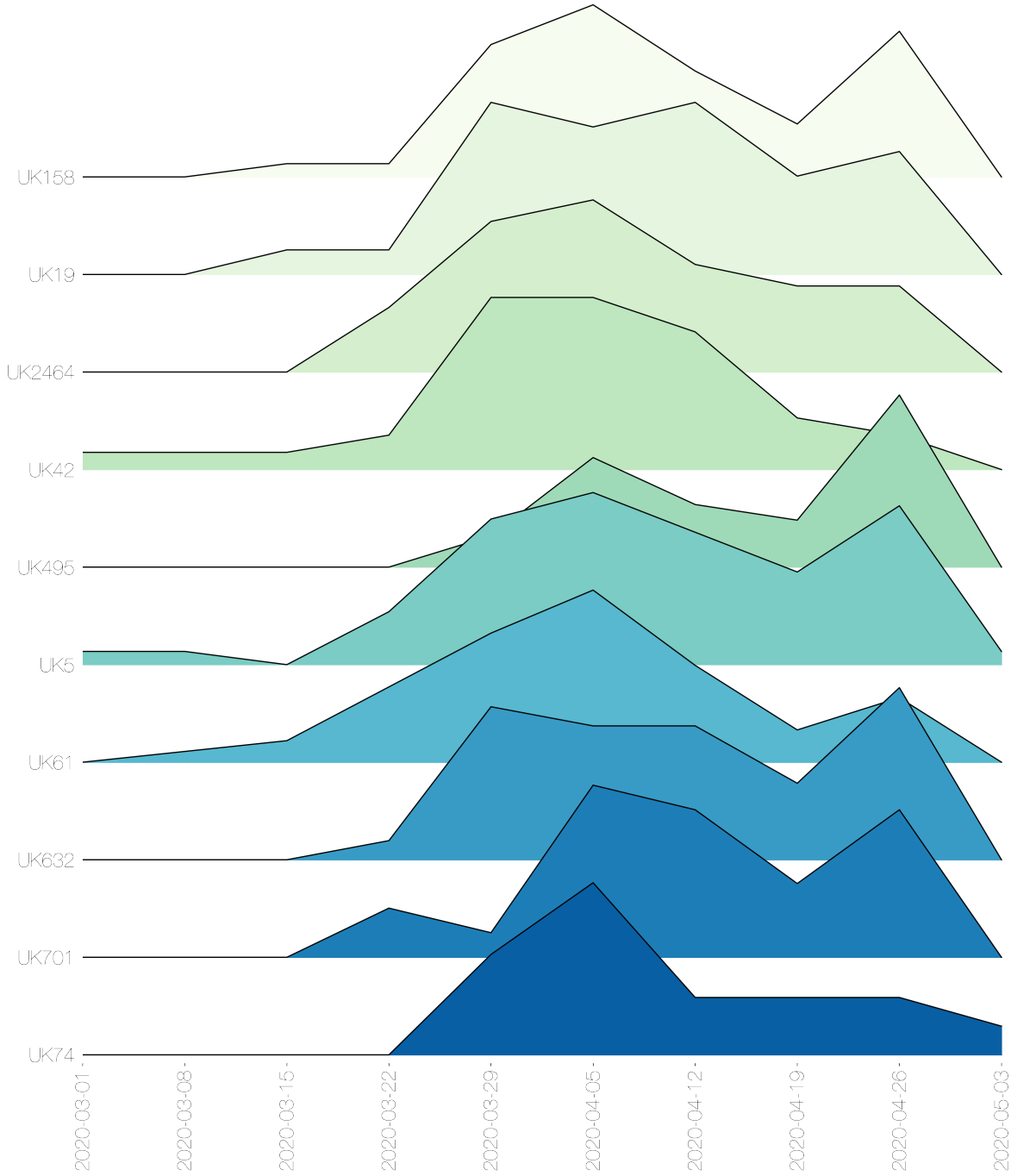
Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)
UK439	Apr-02, Apr-20	7	B.1.1	13
UK418	Apr-03, Apr-20	7	B.1.1.10	13
UK761	Apr-12, Apr-28	6	B.1.1	5
UK358	Mar-31, Apr-09	6	B.2.1	24
UK5670	Apr-01, Apr-30	6	B.2	3
UK80	Mar-31, Apr-27	6	B.1.1.p15	6
UK612	Mar-31, Apr-11	6	B.2.1	22
UK462	Apr-01, Apr-16	6	B.1	17
UK350	Mar-31, Apr-20	6	B.1.1	13
UK627	Mar-31, Apr-10	6	B.1	23
UK451	Mar-25, Apr-05	6	B.2.1	28

These data is represented in the stacked bar chart below. 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.



The relative growth and decline of the ten most sampled lineages in terms of number of counties they are present in is shown below. The raw data for the plot is shown below it, with each column representing a lineage, and the number of admin2 regions it is present in in each week.

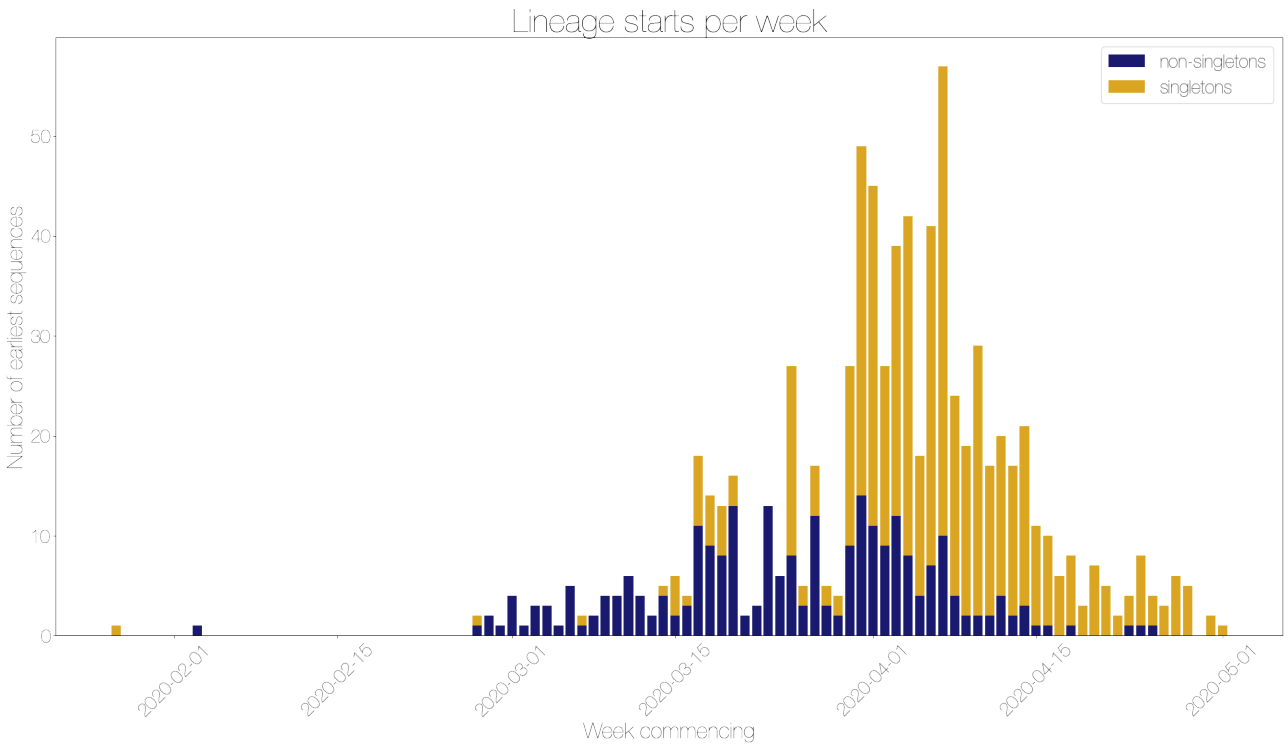


Week commencing	UK61	UK158	UK5	UK42	UK632	UK74	UK19	UK2464	UK495	UK701
2020-03-01	0	0	1	1	0	0	0	0	0	0
2020-03-08	1	0	1	1	0	0	0	0	0	0
2020-03-15	2	1	0	1	0	0	1	0	0	0
2020-03-22	7	1	4	2	1	0	1	3	0	2
2020-03-29	12	10	11	10	8	7	7	7	2	1
2020-04-05	16	13	13	10	7	12	6	8	7	7
2020-04-12	9	8	10	8	7	4	7	5	4	6
2020-04-19	3	4	7	3	4	4	4	4	3	3

Week commencing	UK61	UK158	UK5	UK42	UK632	UK74	UK19	UK2464	UK495	UK701
2020-04-26	6	11	12	2	9	4	5	4	11	6
2020-05-03	0	0	1	0	0	2	0	0	0	0

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

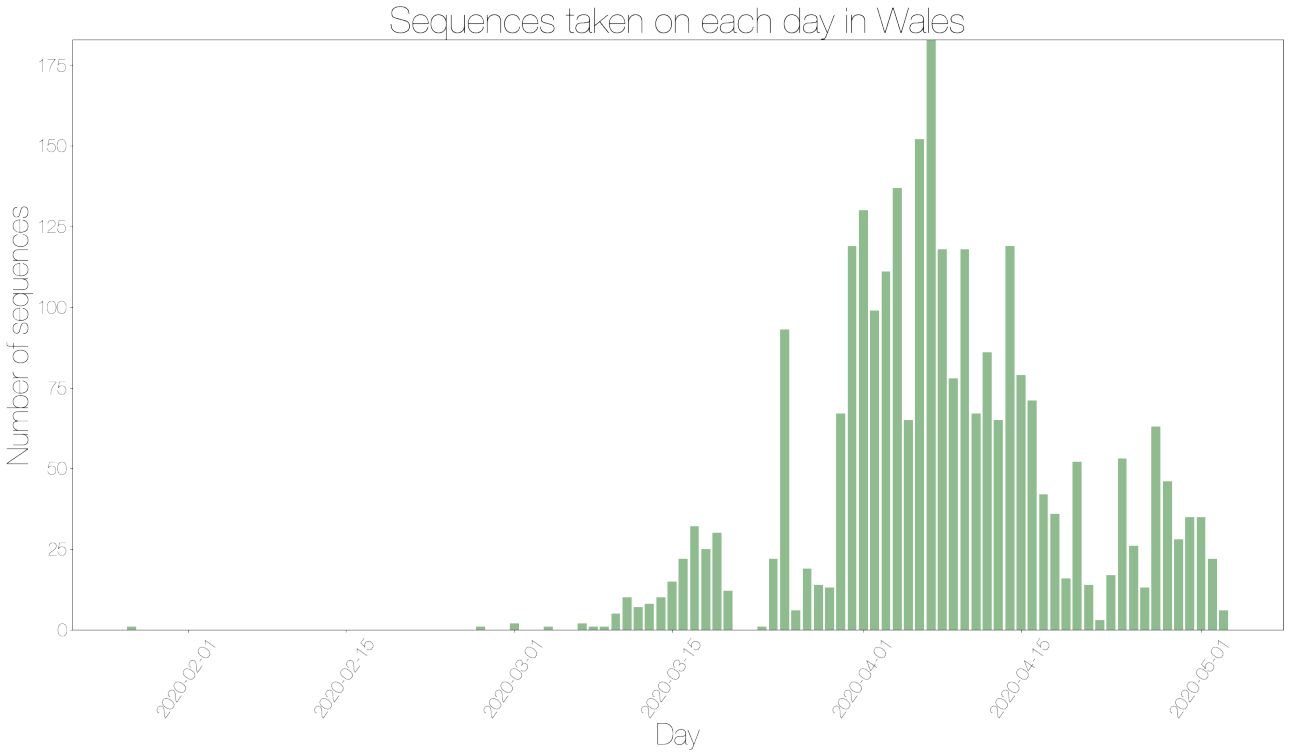
NB the lineage may have started anywhere in the UK, but has been recorded at least once in Wales



Day	Number of singleton starts	Number of non-singleton starts	Total
2020-01-27	1	0	1
2020-02-03	0	1	1
2020-02-27	1	1	2
2020-02-28	0	2	2
2020-02-29	0	1	1
2020-03-01	0	4	4
2020-03-02	0	1	1
2020-03-03	0	3	3
2020-03-04	0	3	3
2020-03-05	0	1	1
2020-03-06	0	5	5
2020-03-07	1	1	2
2020-03-08	0	2	2
2020-03-09	0	4	4
2020-03-10	0	4	4
2020-03-11	0	6	6
2020-03-12	0	4	4
2020-03-13	0	2	2
2020-03-14	1	4	5

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-03-15	4	2	6
2020-03-16	1	3	4
2020-03-17	7	11	18
2020-03-18	5	9	14
2020-03-19	5	8	13
2020-03-20	3	13	16
2020-03-21	0	2	2
2020-03-22	0	3	3
2020-03-23	0	13	13
2020-03-24	0	6	6
2020-03-25	19	8	27
2020-03-26	2	3	5
2020-03-27	5	12	17
2020-03-28	2	3	5
2020-03-29	2	2	4
2020-03-30	18	9	27
2020-03-31	35	14	49
2020-04-01	34	11	45
2020-04-02	18	9	27
2020-04-03	27	12	39
2020-04-04	34	8	42
2020-04-05	14	4	18
2020-04-06	34	7	41
2020-04-07	47	10	57
2020-04-08	20	4	24
2020-04-09	17	2	19
2020-04-10	27	2	29
2020-04-11	15	2	17
2020-04-12	16	4	20
2020-04-13	15	2	17
2020-04-14	18	3	21
2020-04-15	10	1	11
2020-04-16	9	1	10
2020-04-17	6	0	6
2020-04-18	7	1	8
2020-04-19	3	0	3
2020-04-20	7	0	7
2020-04-21	5	0	5
2020-04-22	2	0	2
2020-04-23	3	1	4
2020-04-24	7	1	8
2020-04-25	3	1	4
2020-04-26	3	0	3
2020-04-27	6	0	6
2020-04-28	5	0	5
2020-04-30	2	0	2
2020-05-01	1	0	1

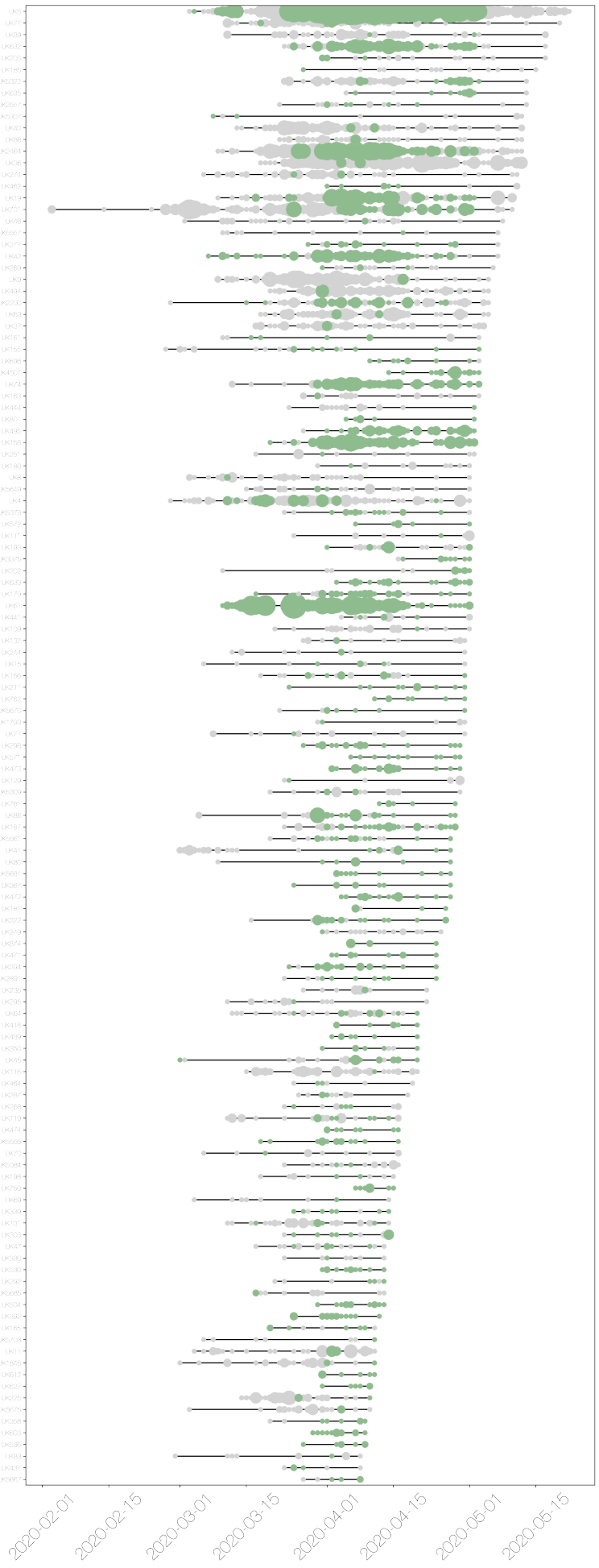
For comparison, here is a plot of the day that every sequence was taken, coloured by country. Note that sequences without dates were not included.



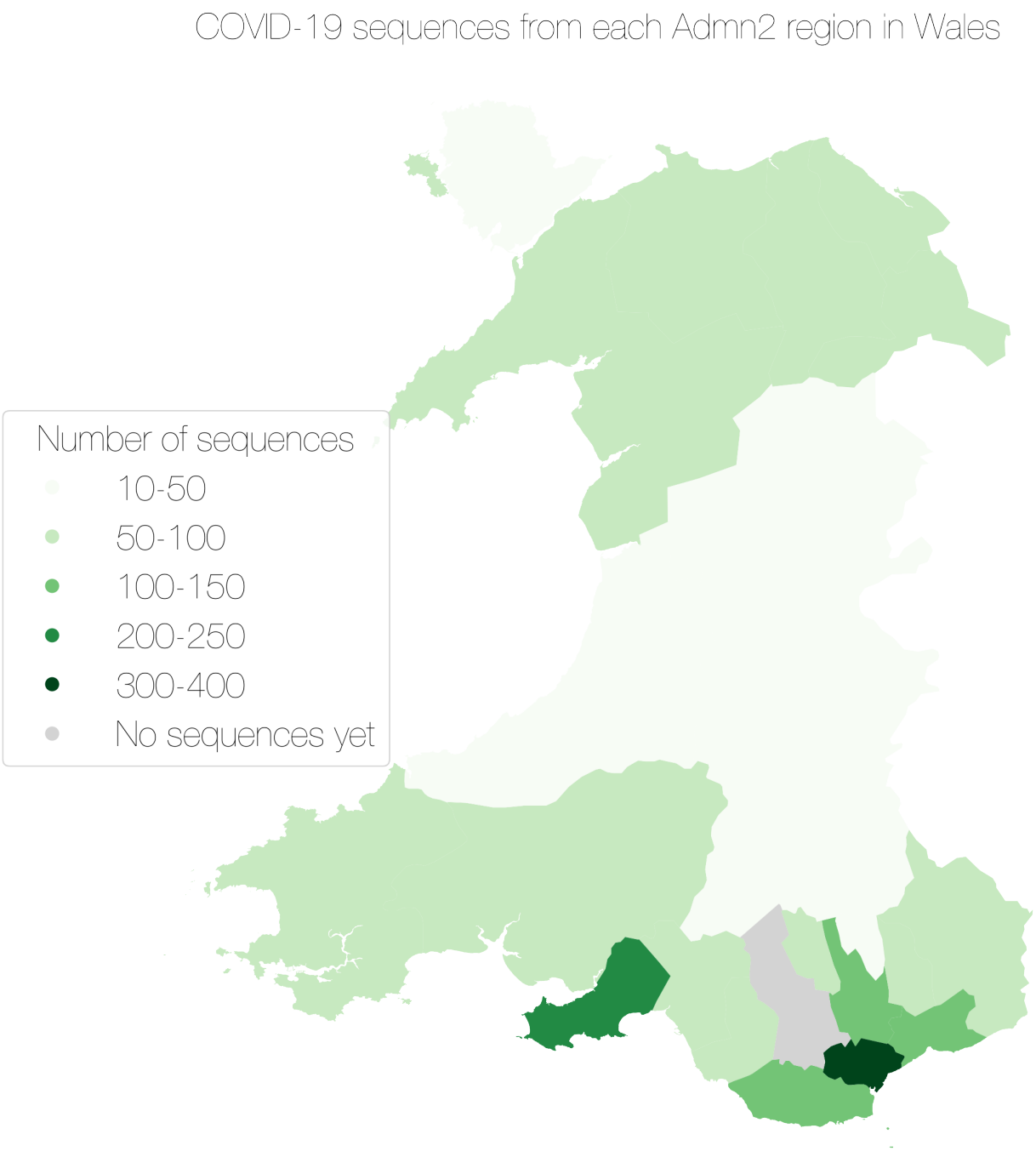
Day	Wales
2020-01-27	1
2020-02-27	1
2020-03-01	2
2020-03-04	1
2020-03-07	2
2020-03-08	1
2020-03-09	1
2020-03-10	5
2020-03-11	10
2020-03-12	7
2020-03-13	8
2020-03-14	10
2020-03-15	15
2020-03-16	22
2020-03-17	32
2020-03-18	25
2020-03-19	30
2020-03-20	12
2020-03-23	1
2020-03-24	22
2020-03-25	93
2020-03-26	6
2020-03-27	19
2020-03-28	14
2020-03-29	13

Day	Wales
2020-03-30	67
2020-03-31	119
2020-04-01	130
2020-04-02	99
2020-04-03	111
2020-04-04	137
2020-04-05	65
2020-04-06	152
2020-04-07	183
2020-04-08	118
2020-04-09	78
2020-04-10	118
2020-04-11	67
2020-04-12	86
2020-04-13	65
2020-04-14	119
2020-04-15	79
2020-04-16	71
2020-04-17	42
2020-04-18	36
2020-04-19	16
2020-04-20	52
2020-04-21	14
2020-04-22	3
2020-04-23	17
2020-04-24	53
2020-04-25	26
2020-04-26	13
2020-04-27	63
2020-04-28	46
2020-04-29	28
2020-04-30	35
2020-05-01	35
2020-05-02	22
2020-05-03	6

These lineages are shown on the timeline below. 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.


2020-02-01
2020-02-15
2020-03-01
2020-03-15
2020-04-01
2020-04-15
2020-05-01
2020-05-15

The map below 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.



Admin2	Country	Number of sequences	Sequence group
ANGLESEY	Wales	23	10-50
BLAENAU GWENT	Wales	46	10-50
BRIDGEND	Wales	96	50-100
CAERPHILLY	Wales	108	100-150
CARDIFF	Wales	368	300-400
CARMARTHENSHIRE	Wales	80	50-100

Admin2	Country	Number of sequences	Sequence group
CEREDIGION	Wales	16	10-50
CONWY	Wales	57	50-100
DENBIGHSHIRE	Wales	86	50-100
FLINTSHIRE	Wales	55	50-100
GWYNEDD	Wales	51	50-100
MERTHYR TYDFIL	Wales	52	50-100
MONMOUTHSHIRE	Wales	53	50-100
NEATH PORT TALBOT	Wales	94	50-100
NEWPORT	Wales	121	100-150
PEMBROKESHIRE	Wales	62	50-100
POWYS	Wales	46	10-50
RHONDDA, CYNON, TAFF	Wales	0	0
SWANSEA	Wales	223	200-250
TORFAEN	Wales	76	50-100
VALE OF GLAMORGAN	Wales	137	100-150
WREXHAM	Wales	73	50-100

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.