UK lineages summary report

This report gives summaries of lineages sampled in Wales for week 2020-05-22. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-04-29. The analysis (eg time since last sample) is therefore undertaken from this date. 2413 sequences from Wales have been included in this analysis. 682 lineages have been recorded, 533 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 688 and the maximum is 980

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

627 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 55 that remain: 9 are pending extinction, ie last seen three weeks ago. 10 lineages have gone quiet, ie haven't been seen this week. 3 lineages have reactivated. 33 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, Apr-29	326	B.3	0
UK158	Mar-20, Apr-29	119	B.1.1.2	0
UK5	Mar-04, Apr-27	109	B.1.1.1	2
UK42	Mar-07, Apr-27	108	B.1.35, B.1	2
UK140	Mar-25, Apr-28	86	B.1.1	1
UK74	Mar-30, Apr-28	84	B.1	1
UK225	Mar-15, Apr-27	82	B.2.6, B.2.2, B.2	2
UK106	Mar-17, Apr-27	65	B.1.44, B.1	2
UK7	Mar-26, Apr-27	64	B.1.p11	2

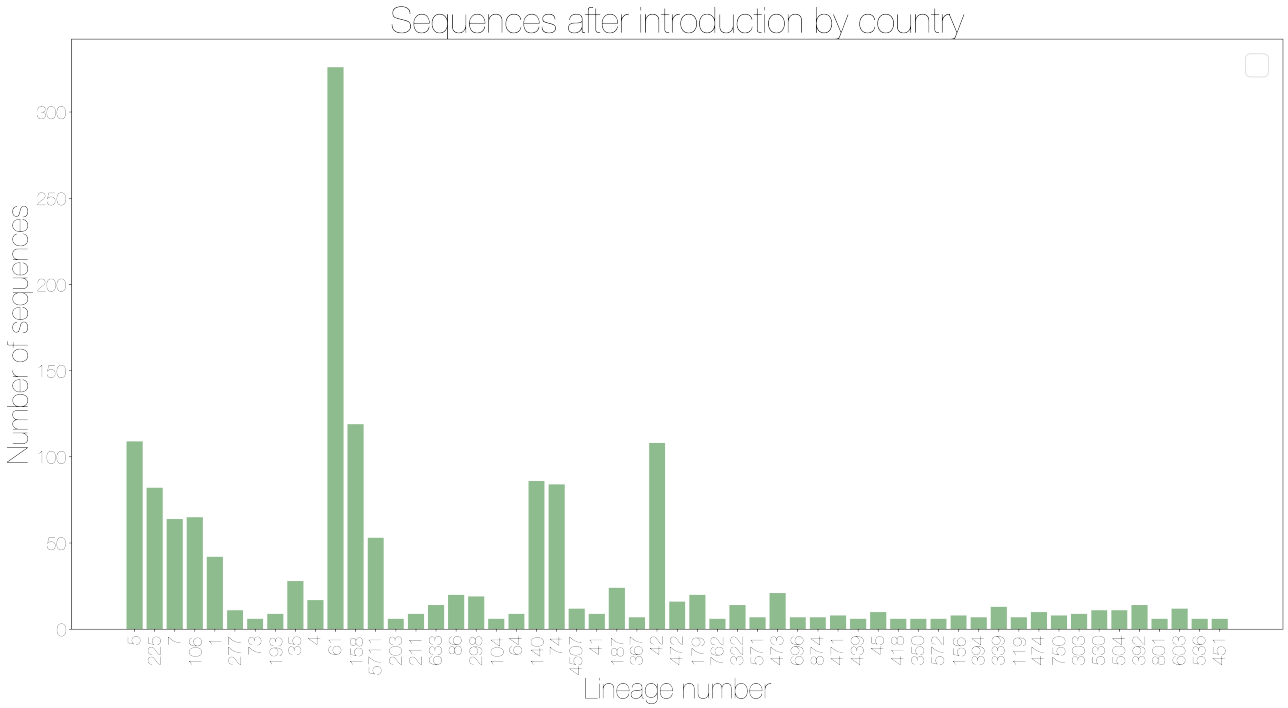
Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)
UK5711	Apr-01, Apr-27	53	B.1.p11	2
UK1	Mar-25, Apr-27	42	B.1	2
UK35	Mar-15, Apr-26	28	B.1.5.6, B.1.5	3
UK187	Apr-01, Apr-27	24	B.1	2
UK473	Apr-02, Apr-24	21	B.1.1	5
UK179	Mar-17, Apr-27	20	B.1.1.p11	2
UK86	Mar-30, Apr-28	20	B.1	1
UK298	Mar-27, Apr-28	19	B.1.1	1
UK4	Mar-11, Apr-24	17	B	5
UK472	Apr-05, Apr-27	16	B.1.1.p11, B.1.1	2
UK633	Apr-06, Apr-28	14	B.1.1.p16, B.1.1.16	1
UK392	Mar-25, Apr-12	14	B.1.67	17
UK322	Mar-30, Apr-26	14	B.1	3
UK339	Mar-14, Apr-14	13	B.3	15
UK4507	Apr-14, Apr-28	12	B.1	1
UK603	Mar-29, Apr-09	12	B.1.1	20
UK504	Mar-30, Apr-13	11	B.1.1	16
UK530	Mar-31, Apr-13	11	B.1.1	16
UK277	Mar-28, Apr-28	11	B.1.1	1
UK45	Mar-01, Apr-20	10	B.1.1	9
UK474	Apr-01, Apr-16	10	B.1.1	13
UK303	Mar-25, Apr-14	9	B.1.1	15
UK193	Apr-01, Apr-14	9	B.1.1	15
UK211	Mar-24, Apr-28	9	B.1.5	1

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)
UK64	Mar-25, Apr-28	9	B.1	1
UK41	Apr-10, Apr-27	9	B.1	2
UK471	Apr-02, Apr-24	8	B.1.1	5
UK156	Mar-28, Apr-14	8	B.1.71	15
UK750	Apr-07, Apr-14	8	B.1	15
UK696	Apr-10, Apr-24	7	B.1.5, B.1	5
UK394	Mar-30, Apr-17	7	B.1.1	12
UK119	Mar-30, Apr-14	7	B.2.5	15
UK367	Mar-25, Apr-27	7	B.1	2
UK571	Apr-06, Apr-25	7	B.1.1	4
UK874	Apr-06, Apr-24	7	B.1	5
UK418	Apr-03, Apr-20	6	B.1.1.10	9
UK572	Apr-07, Apr-19	6	B.1.1	10
UK203	Mar-31, Apr-24	6	B.1.1	5
UK439	Apr-04, Apr-20	6	B.1.1	9
UK762	Apr-11, Apr-26	6	B.1.1	3
UK350	Mar-31, Apr-20	6	B.1.1	9
UK104	Apr-08, Apr-28	6	B.1.1	1
UK536	Mar-27, Apr-09	6	B.1.1	20
UK801	Apr-05, Apr-11	6	B.1	18
UK73	Mar-16, Apr-27	6	B.1.p11	2
UK451	Mar-25, Apr-05	6	B.2.1	24

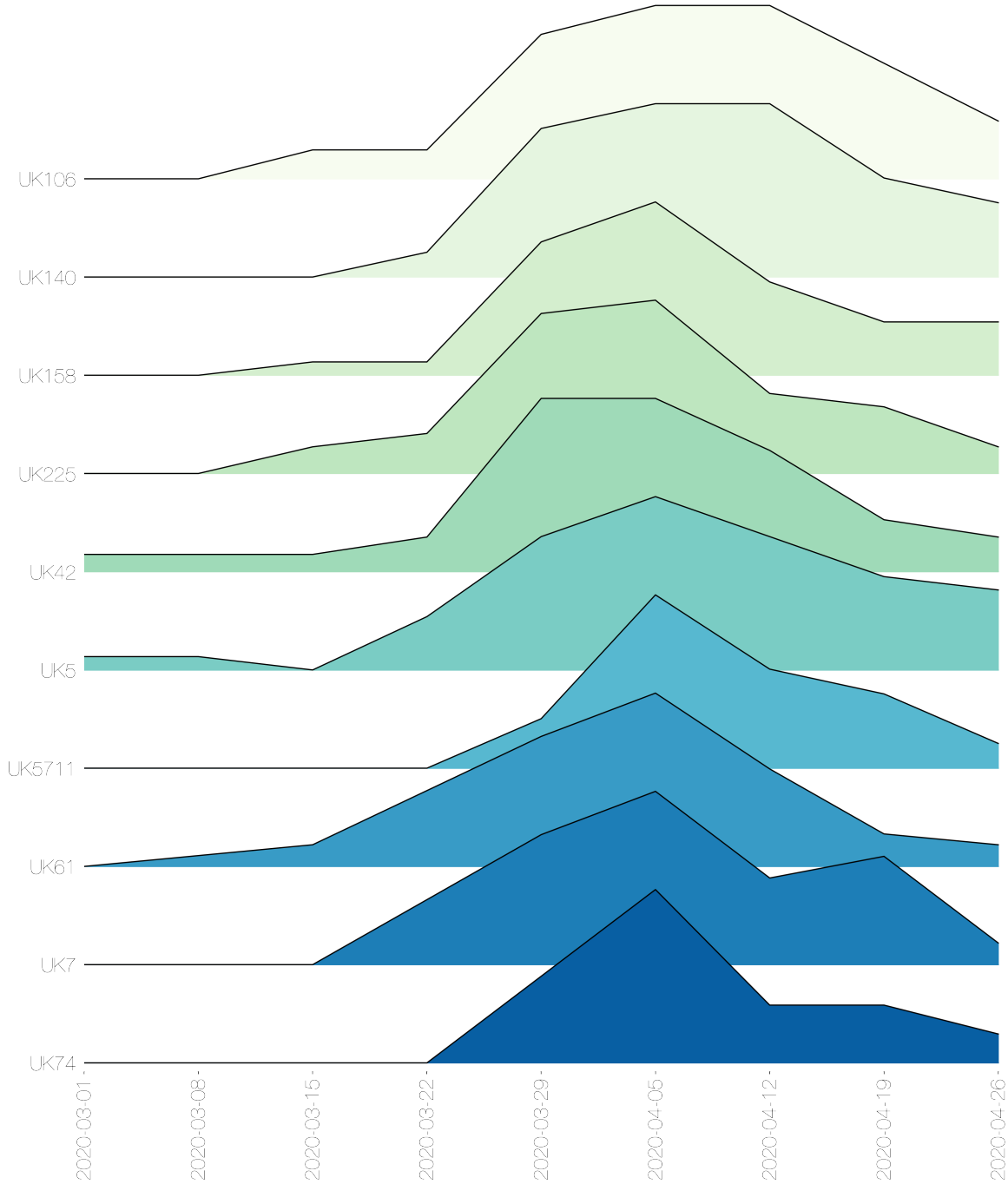
These data is represented in the stacked bar chart below. Note that the number of sequences is likely to be due more to differing sampling efforts in different regions, rather than genuine differences in numbers of

cases.

The raw data for this bar chart are in the table above.



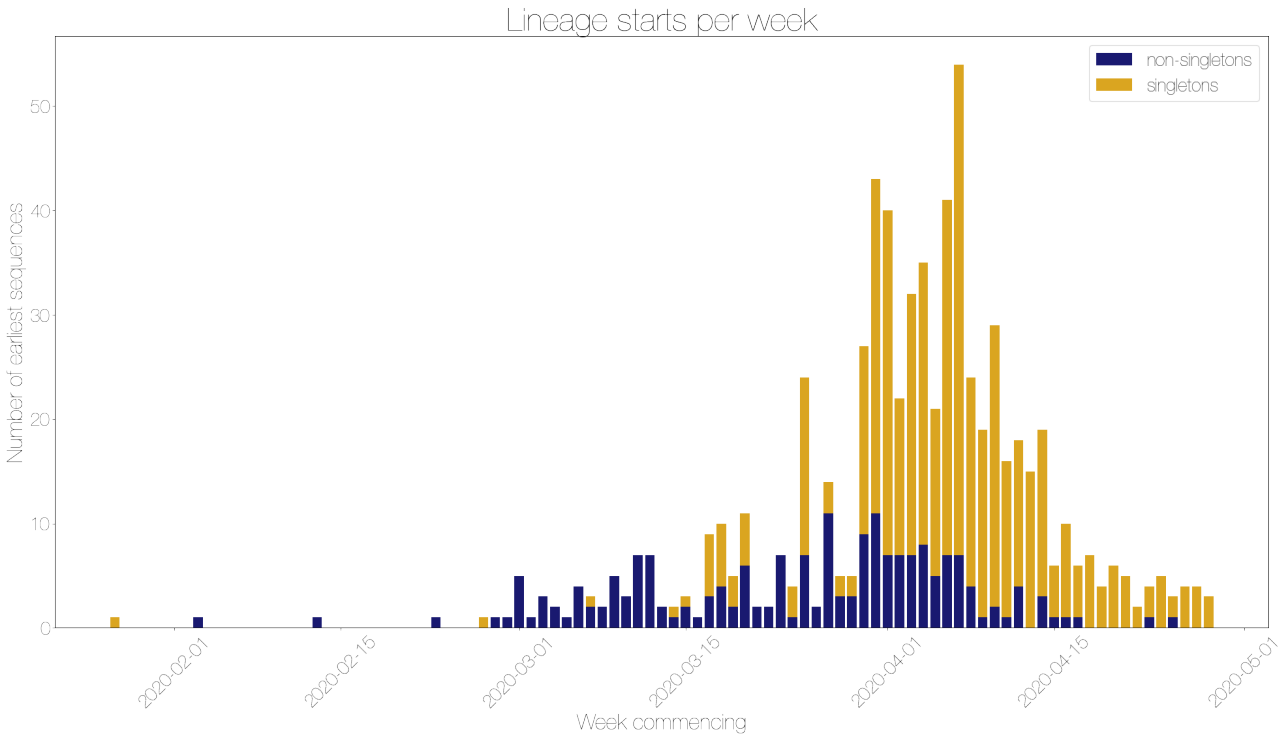
The relative growth and decline of the ten most sampled lineages in terms of number of counties they are present in is shown below. The raw data for the plot is shown below it, with each column representing a lineage, and the number of admin2 regions it is present in in each week.



Week commencing	UK61	UK158	UK5	UK42	UK140	UK74	UK225	UK106	UK7	UK5711
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	2	1	0	0
2020-03-22	7	1	4	2	1	0	3	1	3	0
2020-03-29	12	10	10	10	6	6	12	5	6	2
2020-04-05	16	13	13	10	7	12	13	6	8	7
2020-04-12	9	7	10	7	7	4	6	6	4	4
2020-04-19	3	4	7	3	4	4	5	4	5	3
2020-04-26	2	4	6	2	3	2	2	2	1	1

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

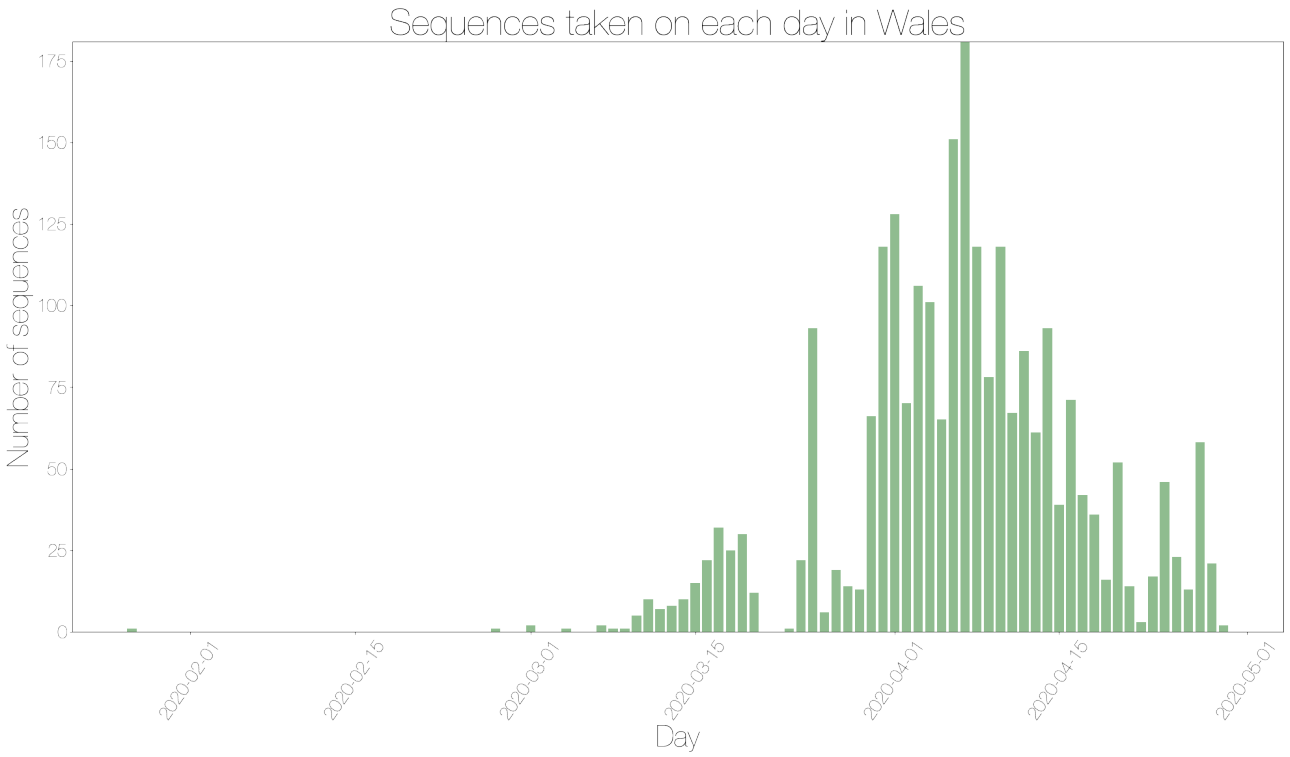
NB the lineage may have started anywhere in the UK, but has been recorded at least once in Wales



Day	Number of singleton starts	Number of non-singleton starts	Total
2020-01-27	1	0	1
2020-02-03	0	1	1
2020-02-13	0	1	1
2020-02-23	0	1	1
2020-02-27	1	0	1
2020-02-28	0	1	1
2020-02-29	0	1	1
2020-03-01	0	5	5
2020-03-02	0	1	1
2020-03-03	0	3	3
2020-03-04	0	2	2
2020-03-05	0	1	1
2020-03-06	0	4	4
2020-03-07	1	2	3
2020-03-08	0	2	2
2020-03-09	0	5	5
2020-03-10	0	3	3
2020-03-11	0	7	7
2020-03-12	0	7	7
2020-03-13	0	2	2
2020-03-14	1	1	2
2020-03-15	1	2	3
2020-03-16	0	1	1
2020-03-17	6	3	9
2020-03-18	6	4	10
2020-03-19	3	2	5

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-03-20	5	6	11
2020-03-21	0	2	2
2020-03-22	0	2	2
2020-03-23	0	7	7
2020-03-24	3	1	4
2020-03-25	17	7	24
2020-03-26	0	2	2
2020-03-27	3	11	14
2020-03-28	2	3	5
2020-03-29	2	3	5
2020-03-30	18	9	27
2020-03-31	32	11	43
2020-04-01	33	7	40
2020-04-02	15	7	22
2020-04-03	25	7	32
2020-04-04	27	8	35
2020-04-05	16	5	21
2020-04-06	34	7	41
2020-04-07	47	7	54
2020-04-08	20	4	24
2020-04-09	18	1	19
2020-04-10	27	2	29
2020-04-11	15	1	16
2020-04-12	14	4	18
2020-04-13	15	0	15
2020-04-14	16	3	19
2020-04-15	5	1	6
2020-04-16	9	1	10
2020-04-17	5	1	6
2020-04-18	7	0	7
2020-04-19	4	0	4
2020-04-20	6	0	6
2020-04-21	5	0	5
2020-04-22	2	0	2
2020-04-23	3	1	4
2020-04-24	5	0	5
2020-04-25	2	1	3
2020-04-26	4	0	4
2020-04-27	4	0	4
2020-04-28	3	0	3

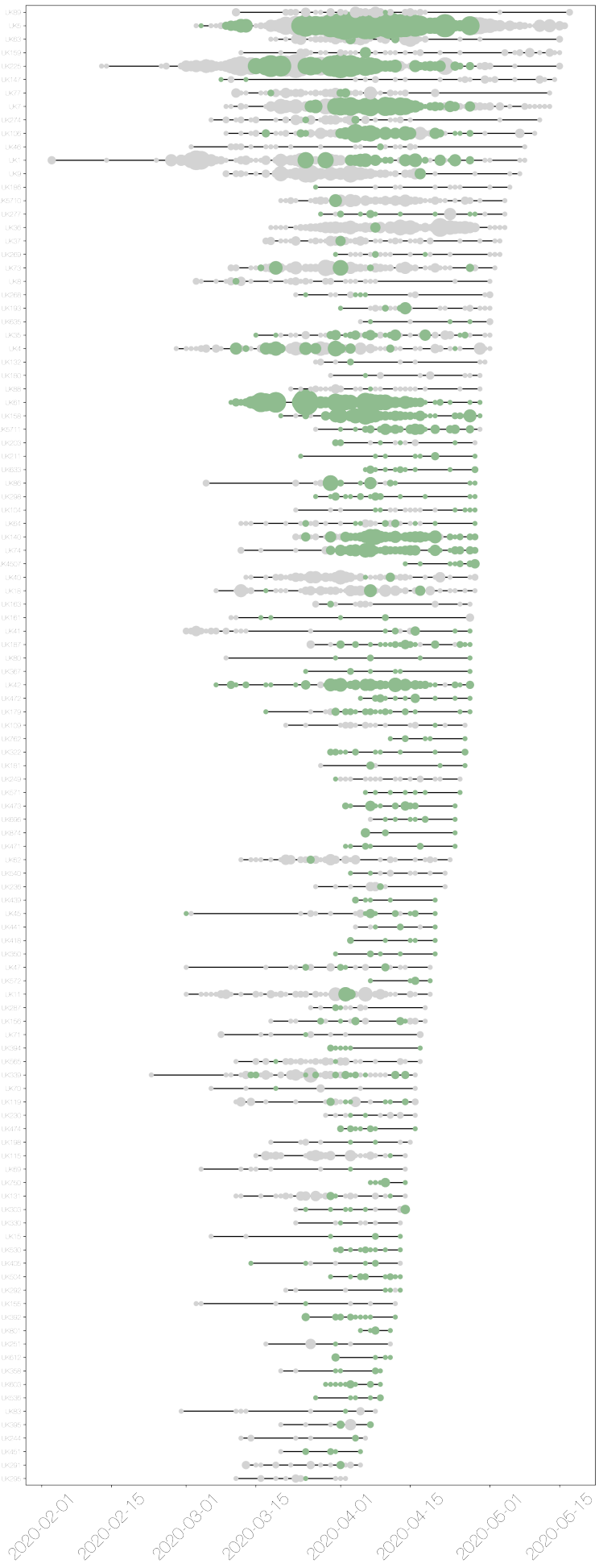
For comparison, here is a plot of the day that every sequence was taken, coloured by country. Note that sequences without dates were not included.



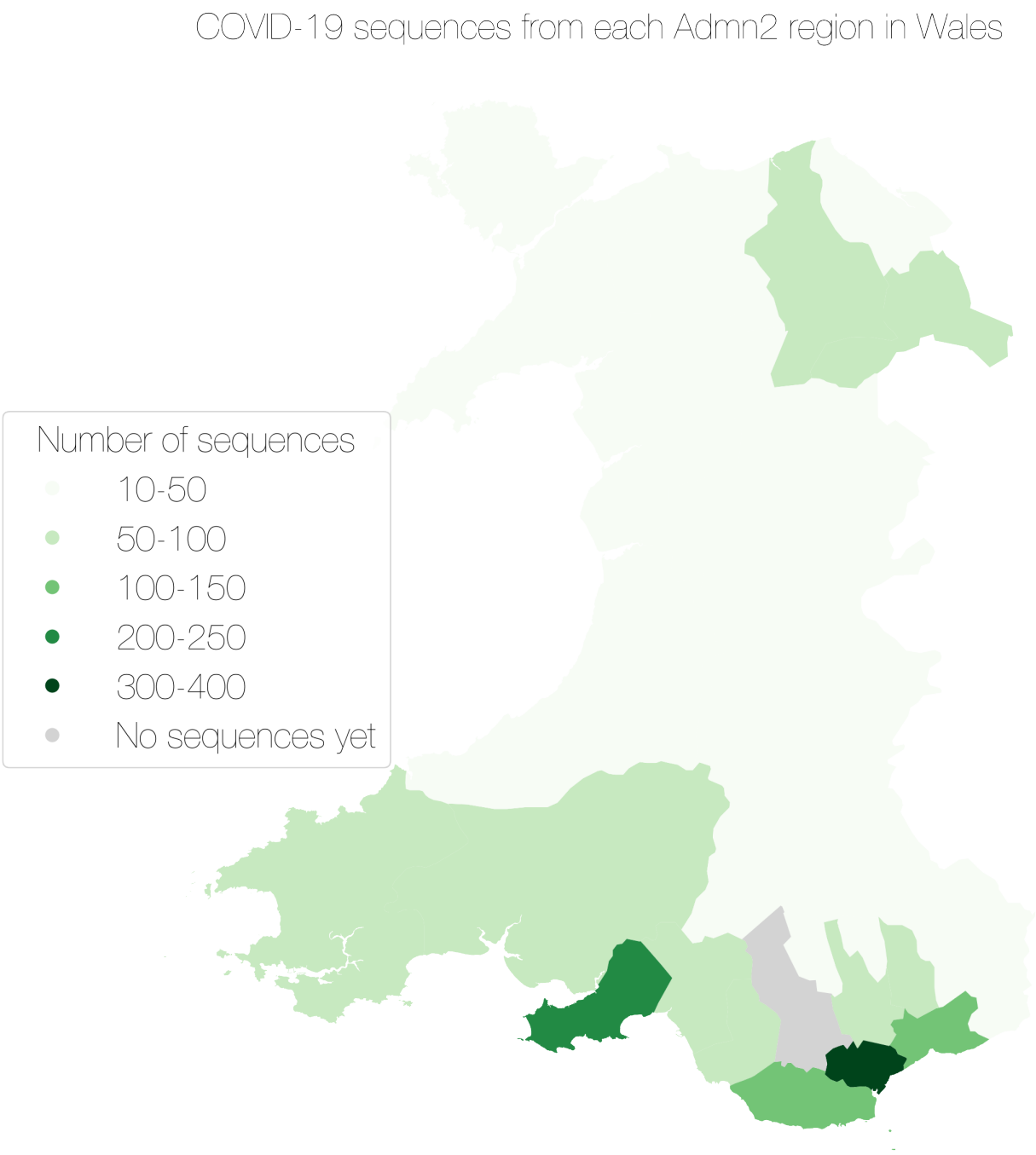
Day	Wales
2020-01-27	1
2020-02-27	1
2020-03-01	2
2020-03-04	1
2020-03-07	2
2020-03-08	1
2020-03-09	1
2020-03-10	5
2020-03-11	10
2020-03-12	7
2020-03-13	8
2020-03-14	10
2020-03-15	15
2020-03-16	22
2020-03-17	32
2020-03-18	25
2020-03-19	30
2020-03-20	12
2020-03-23	1
2020-03-24	22
2020-03-25	93
2020-03-26	6
2020-03-27	19
2020-03-28	14
2020-03-29	13
2020-03-30	66
2020-03-31	118

Day	Wales
2020-04-01	128
2020-04-02	70
2020-04-03	106
2020-04-04	101
2020-04-05	65
2020-04-06	151
2020-04-07	181
2020-04-08	118
2020-04-09	78
2020-04-10	118
2020-04-11	67
2020-04-12	86
2020-04-13	61
2020-04-14	93
2020-04-15	39
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	46
2020-04-25	23
2020-04-26	13
2020-04-27	58
2020-04-28	21
2020-04-29	2

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	18	10-50
BLAENAU GWENT	Wales	42	10-50
BRIDGEND	Wales	83	50-100
CAERPHILLY	Wales	97	50-100
CARDIFF	Wales	310	300-400
CARMARTHENSHIRE	Wales	75	50-100

Admin2	Country	Number of sequences	Sequence group
CEREDIGION	Wales	16	10-50
CONWY	Wales	37	10-50
DENBIGHSHIRE	Wales	64	50-100
FLINTSHIRE	Wales	46	10-50
GWYNEDD	Wales	39	10-50
MERTHYR TYDFIL	Wales	41	10-50
MONMOUTHSHIRE	Wales	46	10-50
NEATH PORT TALBOT	Wales	85	50-100
NEWPORT	Wales	112	100-150
PEMBROKESHIRE	Wales	56	50-100
POWYS	Wales	35	10-50
RHONDDA, CYNON, TAFF	Wales	0	0
SWANSEA	Wales	200	200-250
TORFAEN	Wales	71	50-100
VALE OF GLAMORGAN	Wales	118	100-150
WREXHAM	Wales	64	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.