UK lineages summary report

This report gives summaries of UK specific lineages sequenced by PHWC 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 in the UK from the sequencing centre PHWC 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 5185 and the maximum is 8310

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

627 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 55 that remain: 13 are pending extinction, ie last seen three weeks ago. 7 lineages have gone quiet, ie haven't been seen this week. 7 lineages have reactivated. 28 lineages have been continuously circulating.

The following table contains information about lineages and the number of sequences in each country in the UK for each lineage, 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"

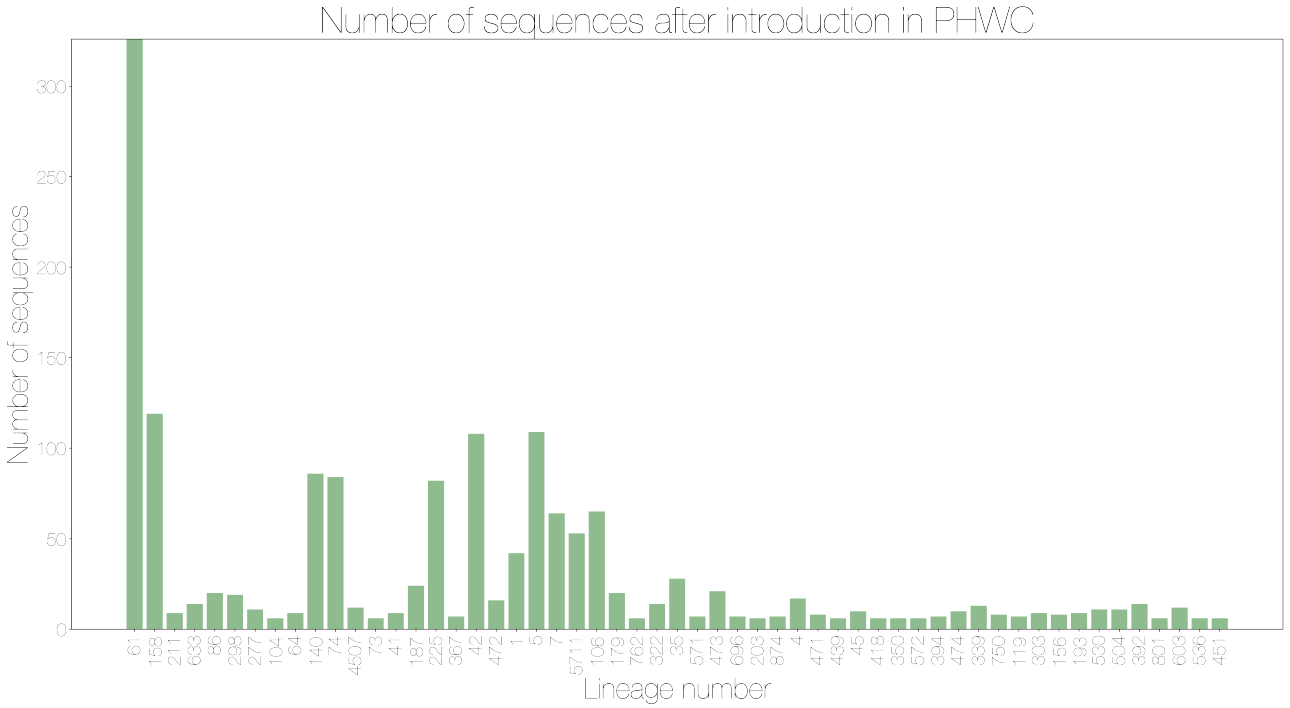
Lineage name	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)
UK61	326 (100.0%)	Mar-10, Apr-29	326	B.3	0
UK158	119 (100.0%)	Mar-20, Apr-29	119	B.1.1.2	0
UK5	109 (100.0%)	Mar-04, Apr-27	109	B.1.1.1	2
UK42	108 (100.0%)	Mar-07, Apr-27	108	B.1.35, B.1	2
UK140	86 (100.0%)	Mar-25, Apr-28	86	B.1.1	1
UK74	84 (100.0%)	Mar-30, Apr-28	84	B.1	1
UK225	82 (100.0%)	Mar-15, Apr-27	82	B.2.6, B.2, B.2.2	2
UK106	65 (100.0%)	Mar-17, Apr-27	65	B.1.44, B.1	2

Lineage name	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)
UK7	64 (100.0%)	Mar-26, Apr-27	64	B.1.p11	2
UK5711	53 (100.0%)	Apr-01, Apr-27	53	B.1.p11	2
UK1	42 (100.0%)	Mar-25, Apr-27	42	B.1	2
UK35	28 (100.0%)	Mar-15, Apr-26	28	B.1.5.6, B.1.5	3
UK187	24 (100.0%)	Apr-01, Apr-27	24	B.1	2
UK473	21 (100.0%)	Apr-02, Apr-24	21	B.1.1	5
UK86	20 (100.0%)	Mar-30, Apr-28	20	B.1	1
UK179	20 (100.0%)	Mar-17, Apr-27	20	B.1.1.p11	2
UK298	19 (100.0%)	Mar-27, Apr-28	19	B.1.1	1
UK4	17 (100.0%)	Mar-11, Apr-24	17	B	5
UK472	16 (100.0%)	Apr-05, Apr-27	16	B.1.1, B.1.1.p11	2
UK392	14 (100.0%)	Mar-25, Apr-12	14	B.1.67	17
UK633	14 (100.0%)	Apr-06, Apr-28	14	B.1.1.16, B.1.1.p16	1
UK322	14 (100.0%)	Mar-30, Apr-26	14	B.1	3
UK339	13 (100.0%)	Mar-14, Apr-14	13	B.3	15
UK4507	12 (100.0%)	Apr-14, Apr-28	12	B.1	1
UK603	12 (100.0%)	Mar-29, Apr-09	12	B.1.1	20
UK530	11 (100.0%)	Mar-31, Apr-13	11	B.1.1	16
UK277	11 (100.0%)	Mar-28, Apr-28	11	B.1.1	1
UK504	11 (100.0%)	Mar-30, Apr-13	11	B.1.1	16
UK474	10 (100.0%)	Apr-01, Apr-16	10	B.1.1	13
UK45	10 (100.0%)	Mar-01, Apr-20	10	B.1.1	9
UK303	9 (100.0%)	Mar-25, Apr-14	9	B.1.1	15
UK41	9 (100.0%)	Apr-10, Apr-27	9	B.1	2

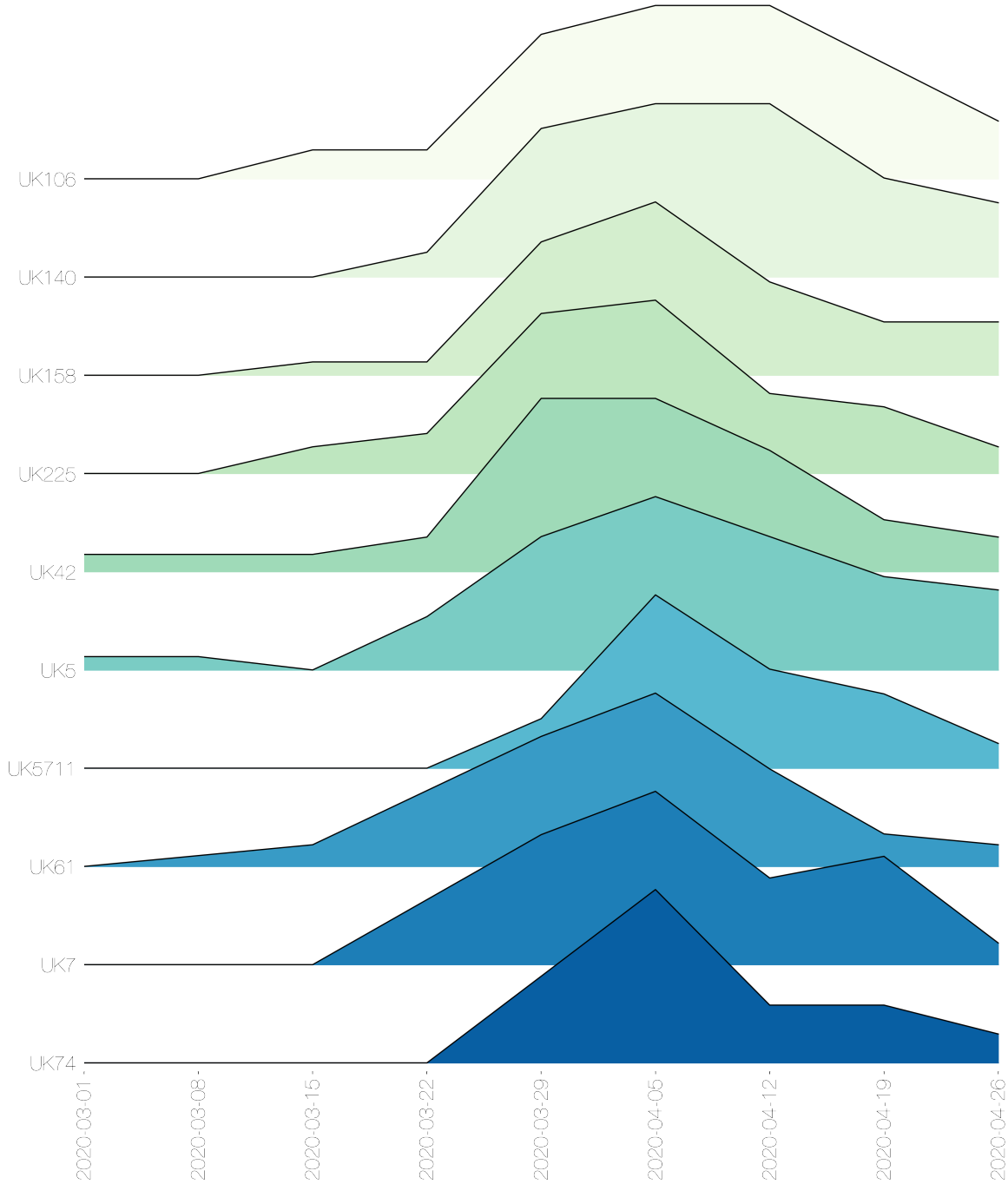
Lineage name	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)
UK211	9 (100.0%)	Mar-24, Apr-28	9	B.1.5	1
UK64	9 (100.0%)	Mar-25, Apr-28	9	B.1	1
UK193	9 (100.0%)	Apr-01, Apr-14	9	B.1.1	15
UK750	8 (100.0%)	Apr-07, Apr-14	8	B.1	15
UK471	8 (100.0%)	Apr-02, Apr-24	8	B.1.1	5
UK156	8 (100.0%)	Mar-28, Apr-14	8	B.1.71	15
UK119	7 (100.0%)	Mar-30, Apr-14	7	B.2.5	15
UK367	7 (100.0%)	Mar-25, Apr-27	7	B.1	2
UK394	7 (100.0%)	Mar-30, Apr-17	7	B.1.1	12
UK571	7 (100.0%)	Apr-06, Apr-25	7	B.1.1	4
UK696	7 (100.0%)	Apr-10, Apr-24	7	B.1.5, B.1	5
UK874	7 (100.0%)	Apr-06, Apr-24	7	B.1	5
UK104	6 (100.0%)	Apr-08, Apr-28	6	B.1.1	1
UK73	6 (100.0%)	Mar-16, Apr-27	6	B.1.p11	2
UK451	6 (100.0%)	Mar-25, Apr-05	6	B.2.1	24
UK350	6 (100.0%)	Mar-31, Apr-20	6	B.1.1	9
UK203	6 (100.0%)	Mar-31, Apr-24	6	B.1.1	5
UK536	6 (100.0%)	Mar-27, Apr-09	6	B.1.1	20
UK418	6 (100.0%)	Apr-03, Apr-20	6	B.1.1.10	9
UK439	6 (100.0%)	Apr-04, Apr-20	6	B.1.1	9
UK762	6 (100.0%)	Apr-11, Apr-26	6	B.1.1	3
UK801	6 (100.0%)	Apr-05, Apr-11	6	B.1	18
UK572	6 (100.0%)	Apr-07, Apr-19	6	B.1.1	10

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.

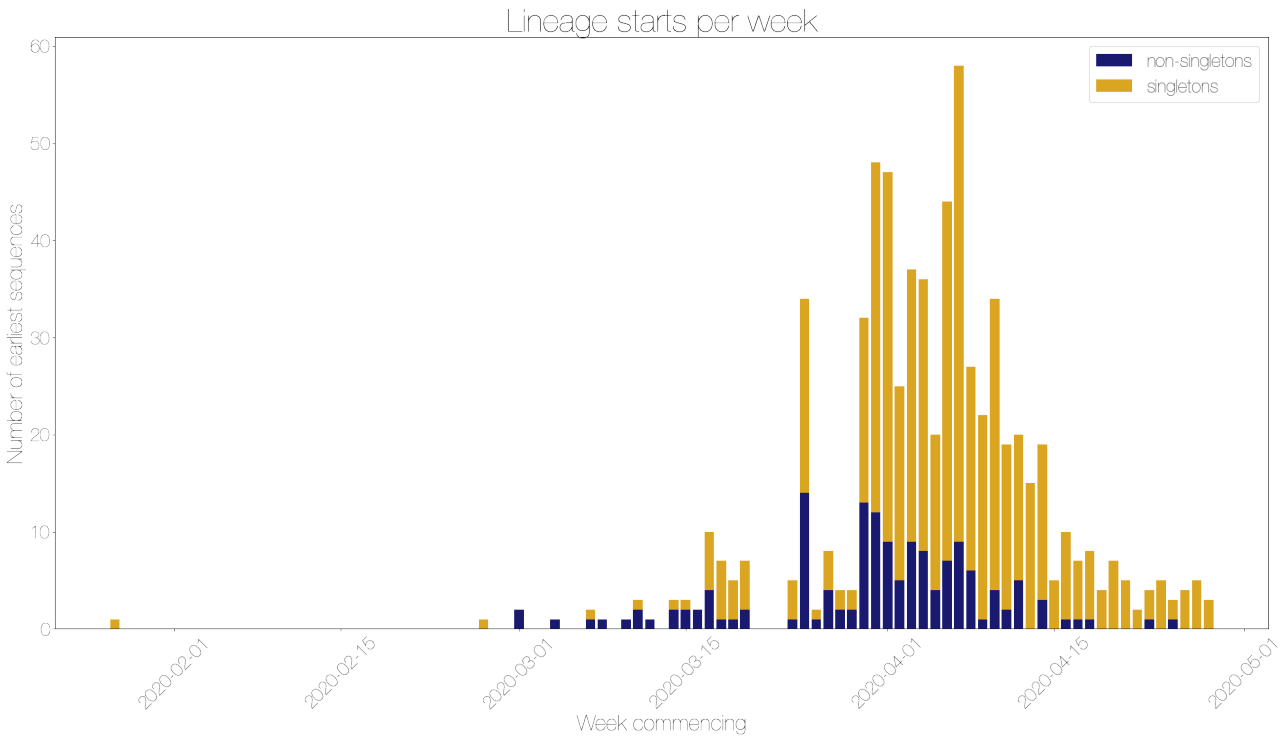


The relative growth and decline of the ten most sampled lineages in terms of number of counties they are present in is shown below. The raw data for the plot is shown below it, with each column representing a lineage, and the number of admin2 regions it is present in in each week.



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

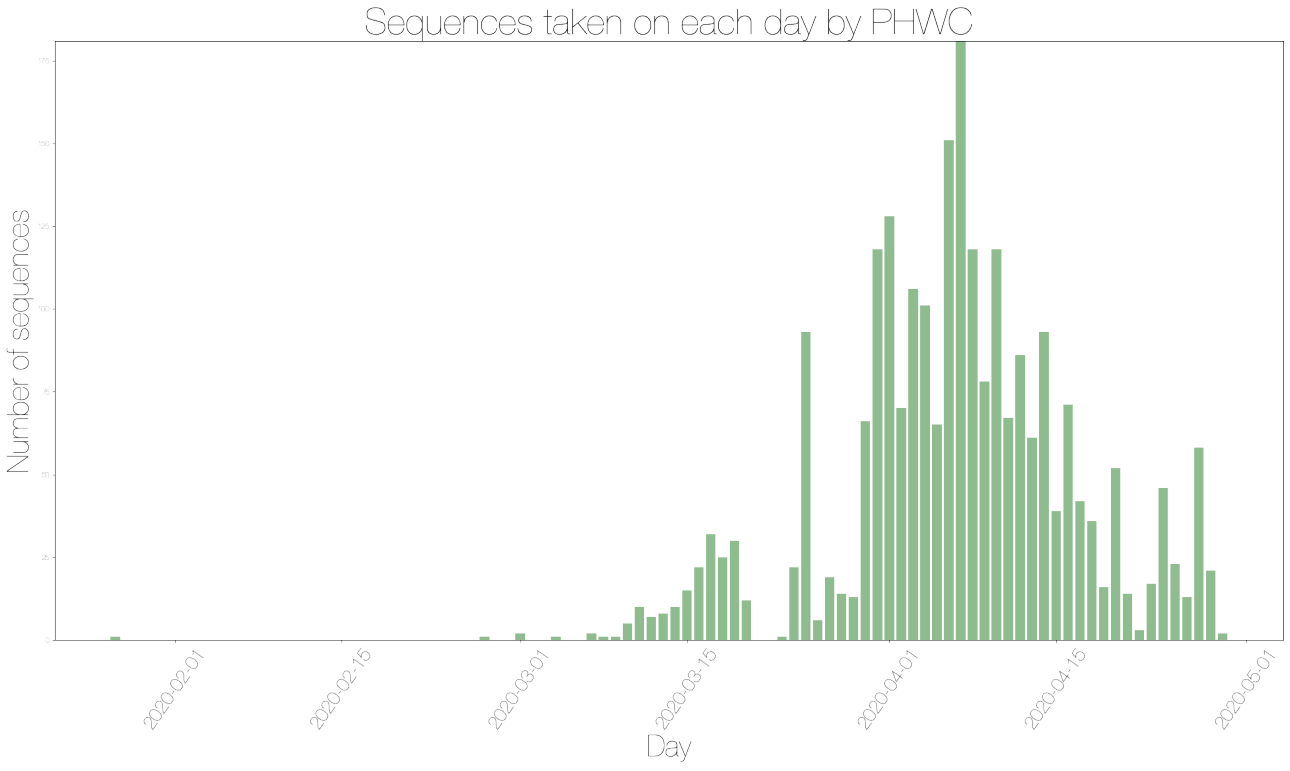
The date of first sequence in the cluster is shown below for every cluster with date information.



Day	Number of singleton starts	Number of non-singleton starts	Total
2020-01-27	1	0	1
2020-02-27	1	0	1
2020-03-01	0	2	2
2020-03-04	0	1	1
2020-03-07	1	1	2
2020-03-08	0	1	1
2020-03-10	0	1	1
2020-03-11	1	2	3
2020-03-12	0	1	1
2020-03-14	1	2	3
2020-03-15	1	2	3
2020-03-16	0	2	2
2020-03-17	6	4	10
2020-03-18	6	1	7
2020-03-19	4	1	5
2020-03-20	5	2	7
2020-03-24	4	1	5
2020-03-25	20	14	34
2020-03-26	1	1	2
2020-03-27	4	4	8
2020-03-28	2	2	4
2020-03-29	2	2	4
2020-03-30	19	13	32
2020-03-31	36	12	48
2020-04-01	38	9	47
2020-04-02	20	5	25
2020-04-03	28	9	37
2020-04-04	28	8	36

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-04-05	16	4	20
2020-04-06	37	7	44
2020-04-07	49	9	58
2020-04-08	21	6	27
2020-04-09	21	1	22
2020-04-10	30	4	34
2020-04-11	17	2	19
2020-04-12	15	5	20
2020-04-13	15	0	15
2020-04-14	16	3	19
2020-04-15	5	0	5
2020-04-16	9	1	10
2020-04-17	6	1	7
2020-04-18	7	1	8
2020-04-19	4	0	4
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	5	0	5
2020-04-25	2	1	3
2020-04-26	4	0	4
2020-04-27	5	0	5
2020-04-28	3	0	3

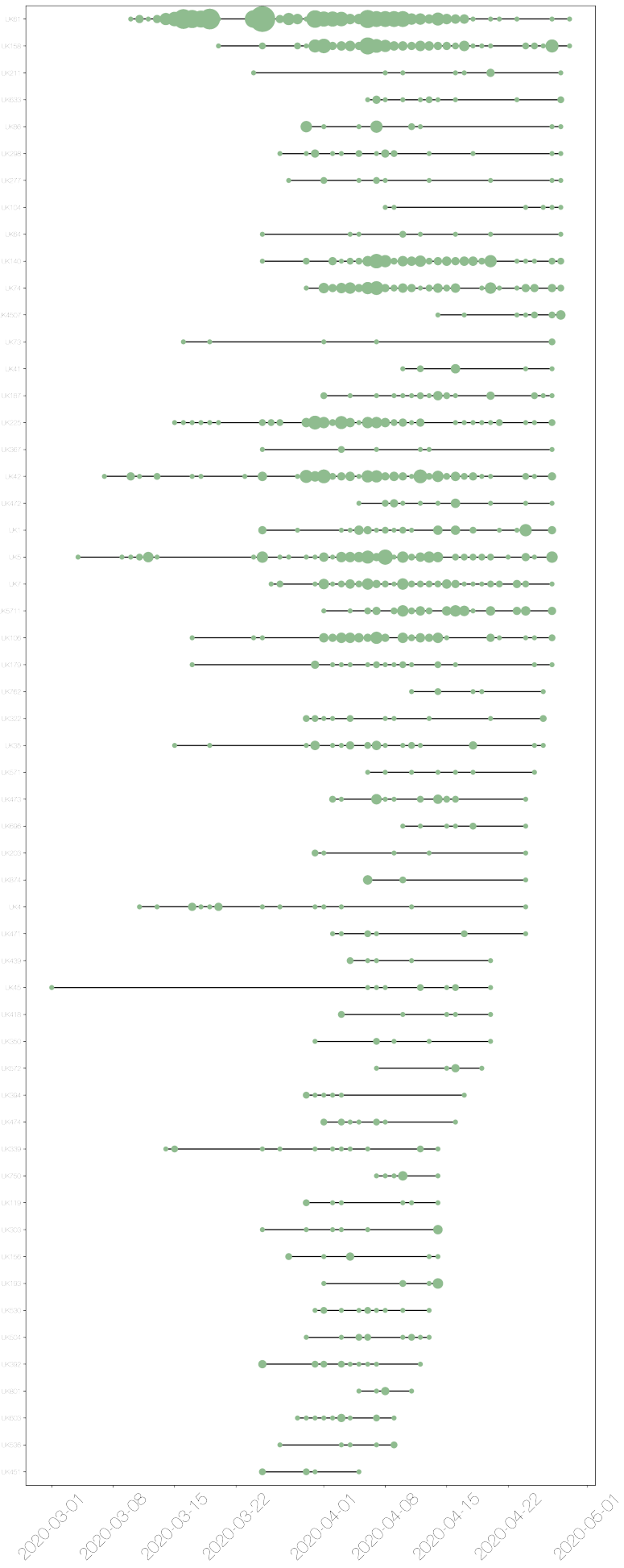
For comparison, here is a plot of the day that every sequence was taken, coloured by country. Note that sequences without dates were not included.



Day	Wales
2020-01-27	21
2020-02-27	58
2020-03-01	46
2020-03-04	17
2020-03-07	42
2020-03-08	2
2020-03-09	23
2020-03-10	13
2020-03-11	36
2020-03-12	1
2020-03-13	52
2020-03-14	16
2020-03-15	3
2020-03-16	93
2020-03-17	71
2020-03-18	39
2020-03-19	61
2020-03-20	14
2020-03-23	181
2020-03-24	151
2020-03-25	86
2020-03-26	101
2020-03-27	78
2020-03-28	106
2020-03-29	118
2020-03-30	67
2020-03-31	70
2020-04-01	118
2020-04-02	128
2020-04-03	118
2020-04-04	65
2020-04-05	13
2020-04-06	14
2020-04-07	19
2020-04-08	66
2020-04-09	6
2020-04-10	22
2020-04-11	93
2020-04-12	2
2020-04-13	1
2020-04-14	10
2020-04-15	5
2020-04-16	1
2020-04-17	1
2020-04-18	1
2020-04-19	2
2020-04-20	1

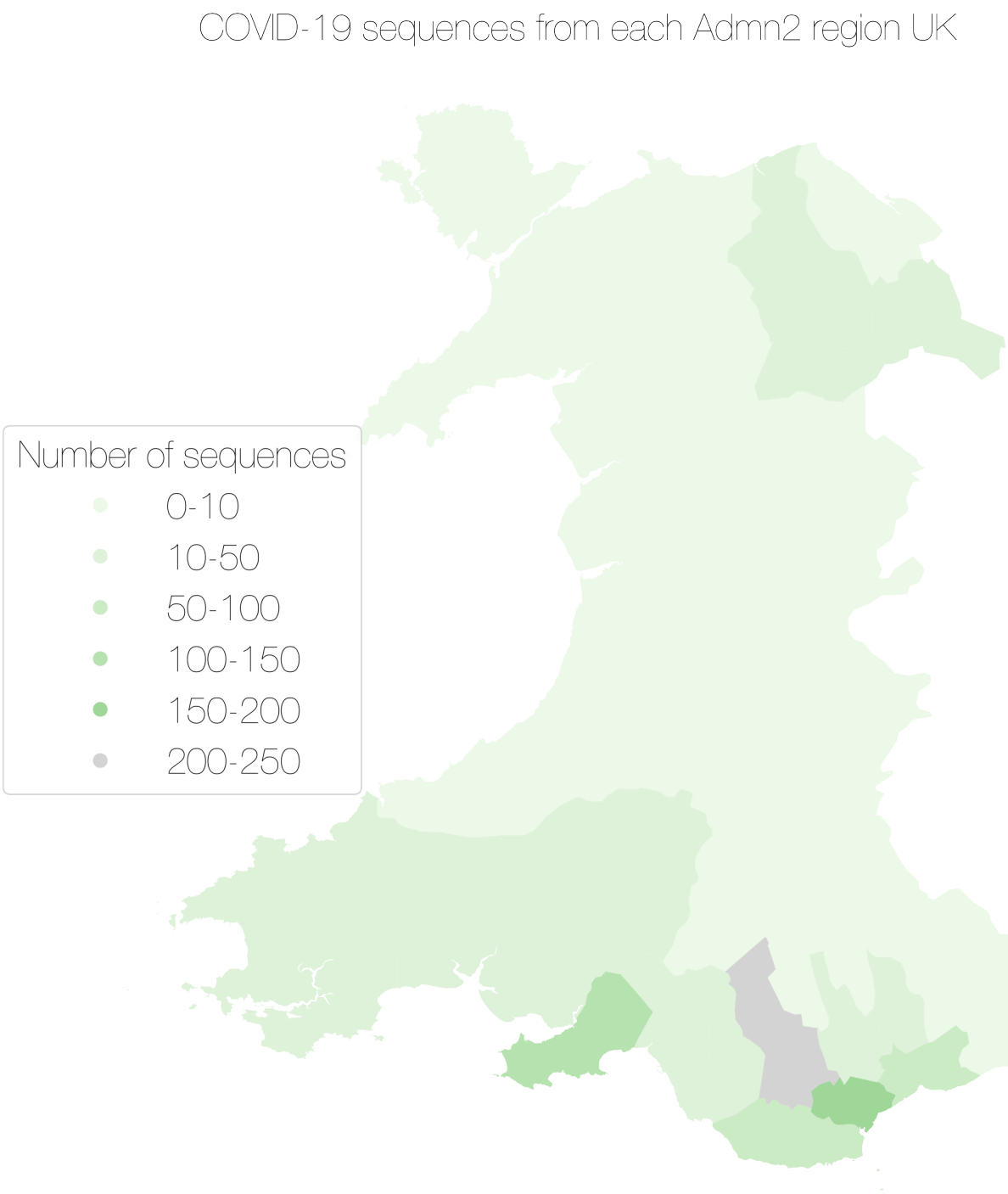
Day	Wales
2020-04-21	12
2020-04-22	30
2020-04-23	25
2020-04-24	32
2020-04-25	22
2020-04-26	15
2020-04-27	10
2020-04-28	8
2020-04-29	7

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.


2020-03-01
2020-03-08
2020-03-15
2020-03-22
2020-04-01
2020-04-08
2020-04-15
2020-04-22
2020-05-01

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.

There are 535 sequences without enough geographical information to map from this centre.



Admin2	Country	Number of sequences	Sequence group
ANGLESEY	Wales	18	10-50
BLAENAU GWENT	Wales	42	10-50
BRIDGEND	Wales	83	50-100

Admin2	Country	Number of sequences	Sequence group
CAERPHILLY	Wales	97	50-100
CARDIFF	Wales	309	300-400
CARMARTHENSHIRE	Wales	74	50-100
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
SWANSEA	Wales	200	200-250
TORFAEN	Wales	71	50-100
VALE OF GLAMORGAN	Wales	118	100-150
WREXHAM	Wales	64	50-100

Other results modules for UK lineage analysis can be added in here if required.