UK lineages summary report

This report gives summaries of UK specific lineages sequenced by PHWC 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 in the UK from the sequencing centre PHWC have been included in this analysis.

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 6130 and the maximum is 9084

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

716 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 66 that remain: 15 are pending extinction, ie last seen three weeks ago. 13 lineages have gone quiet, ie haven't been seen this week. 11 lineages have reactivated. 27 lineages have been continuously circulating.

The following table contains information about the ten largest lineages lineages and the number of sequences the dataset. Information about other lienages 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 expect 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	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK61	340 (100.0%)	Mar-10, May-01	340	B.3	2	0.0767
UK158	142 (100.0%)	Mar-20, May-02	142	B.1.1.2	1	0.305
UK5	125 (100.0%)	Mar-04, May-03	125	B.1.1, B.1.1.1	0	active today
UK42	112 (100.0%)	Mar-07, Apr-27	112	B.1.35, B.1	6	0.0766
UK632	97 (100.0%)	Mar-25, May-02	97	B.1.1	1	0.3958
UK74	96 (100.0%)	Mar-30, May-03	96	B.1	0	active today

Lineage name	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK19	81 (100.0%)	Mar-17, May-02	81	B.1, B.1.44	1	0.575
UK2464	70 (100.0%)	Mar-26, May-02	70	B.1.p11	1	0.5362
UK495	65 (100.0%)	Apr-01, May-02	65	B.1.p11	1	0.4844
UK701	43 (100.0%)	Mar-25, May-01	43	B.1	2	0.4405

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.

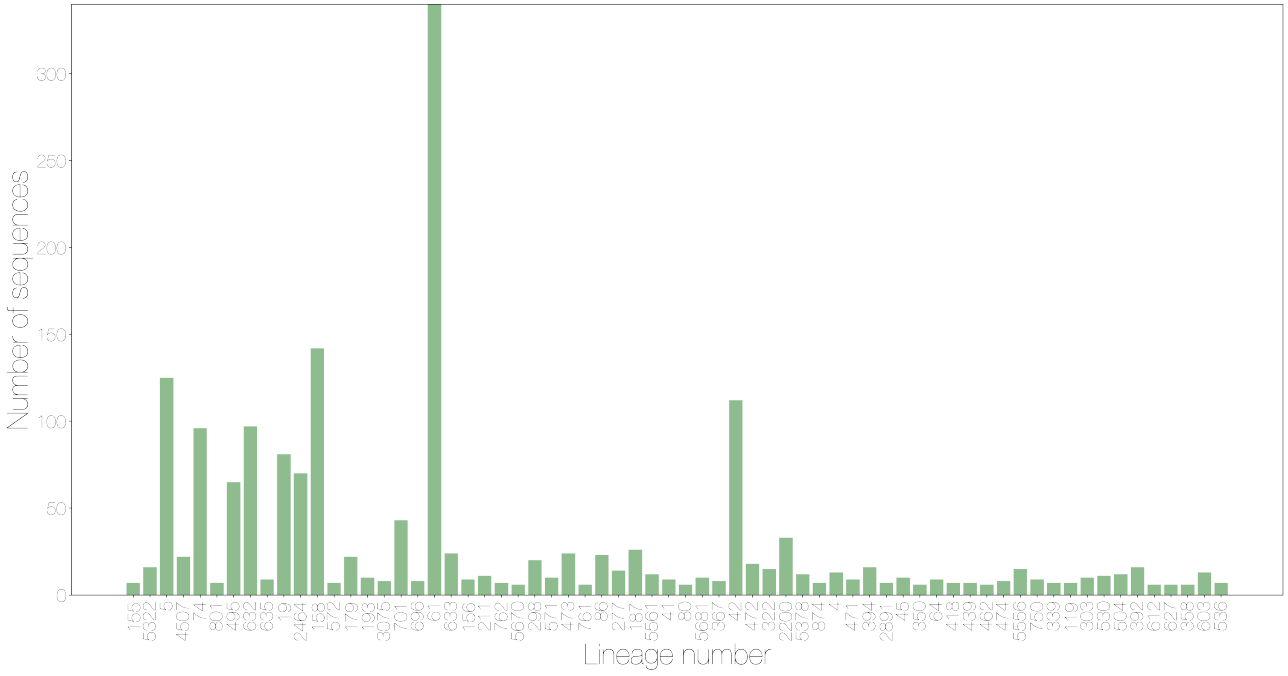


Figure 1: Number of sequences sampled in a lineage by country

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.

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.

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

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.

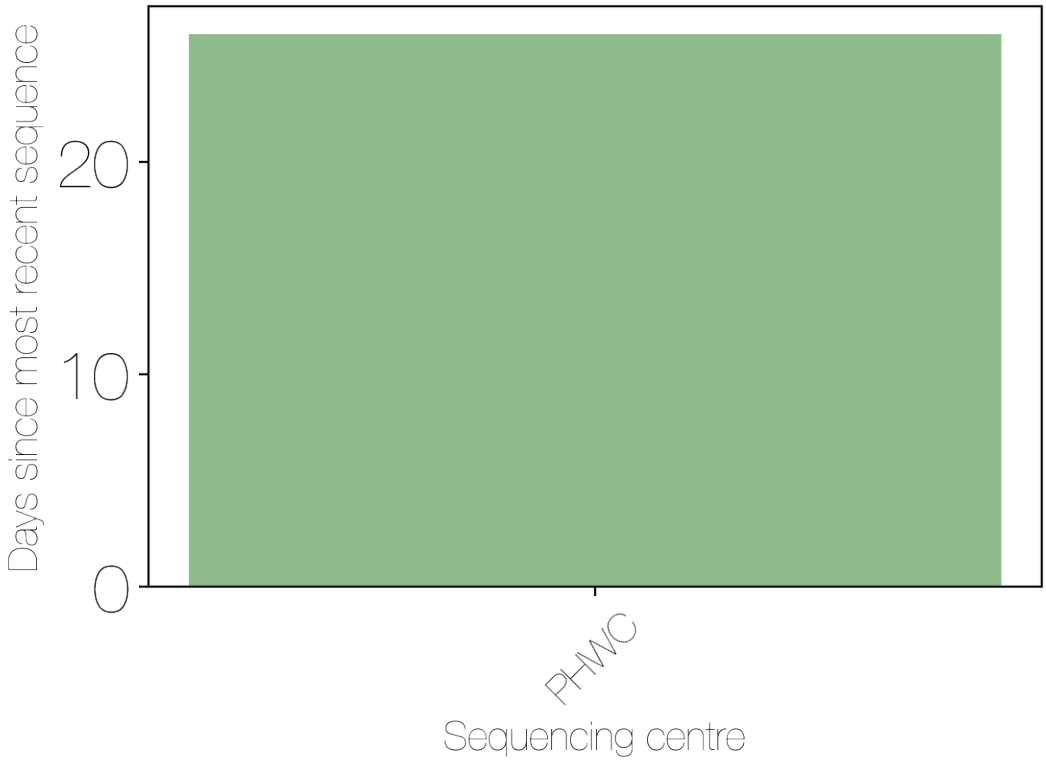


Figure 2: Lag since the most recent sequence from each sequencing centre to most current date

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.

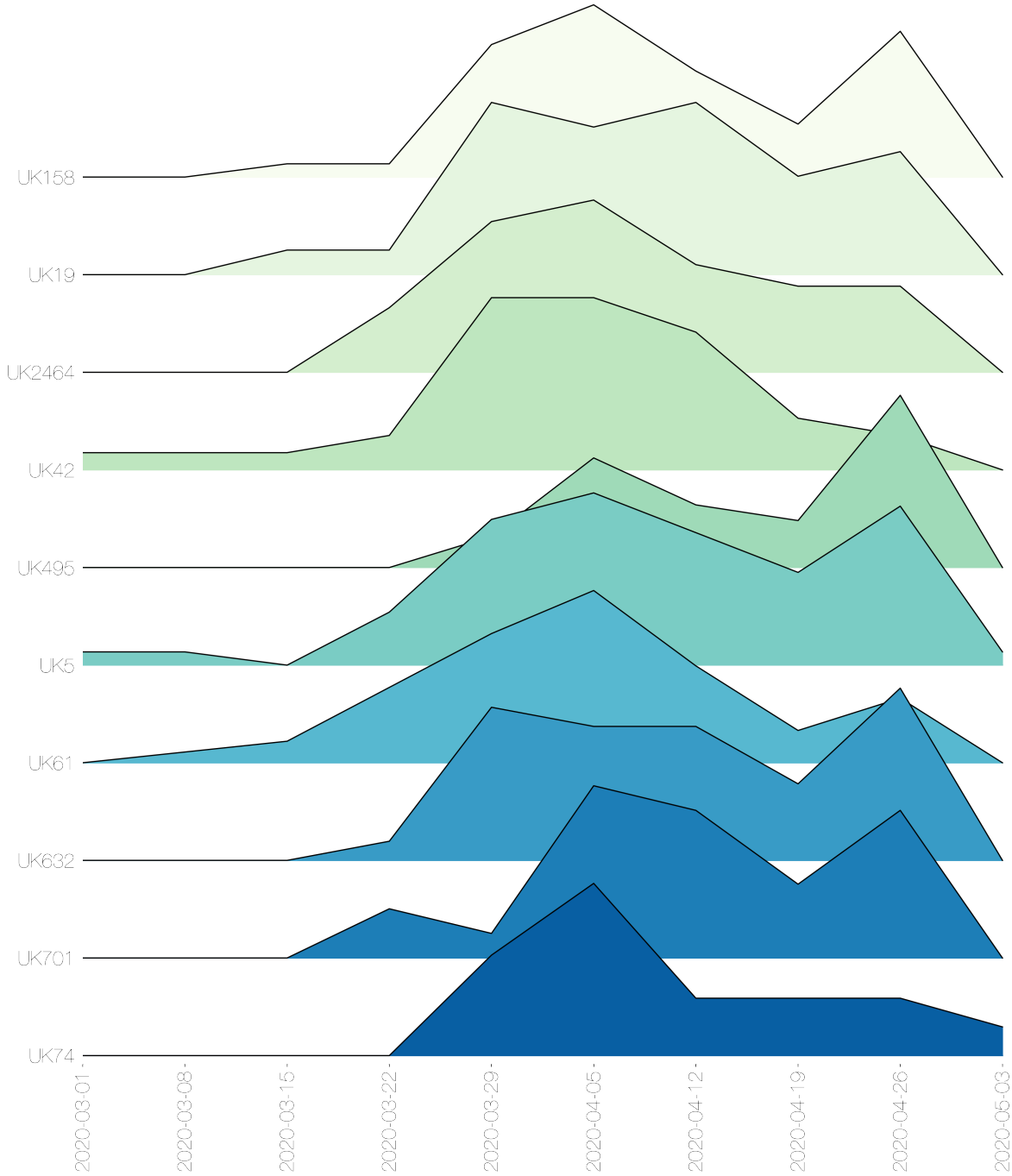


Figure 3: Lineages by number of adm2 regions present by epiweek

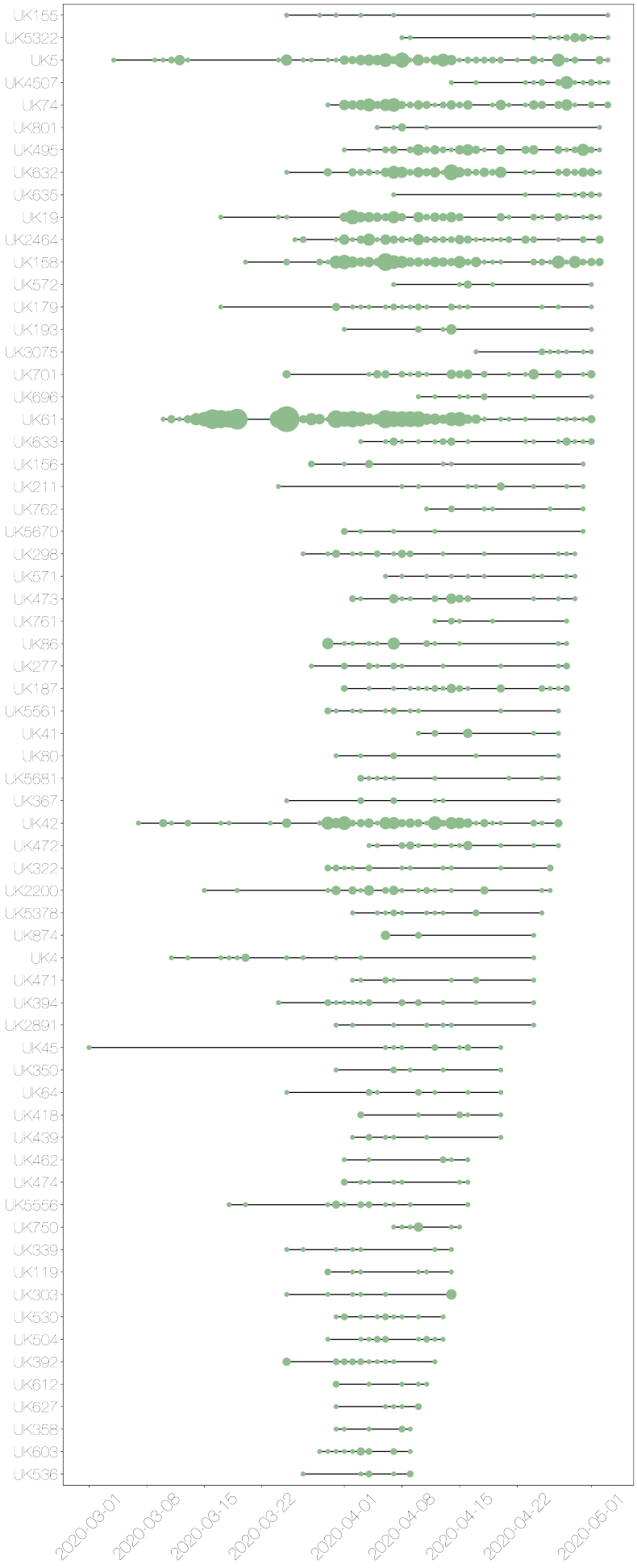


Figure 4: Timeline of lineages, sized by number of sequences from each country.

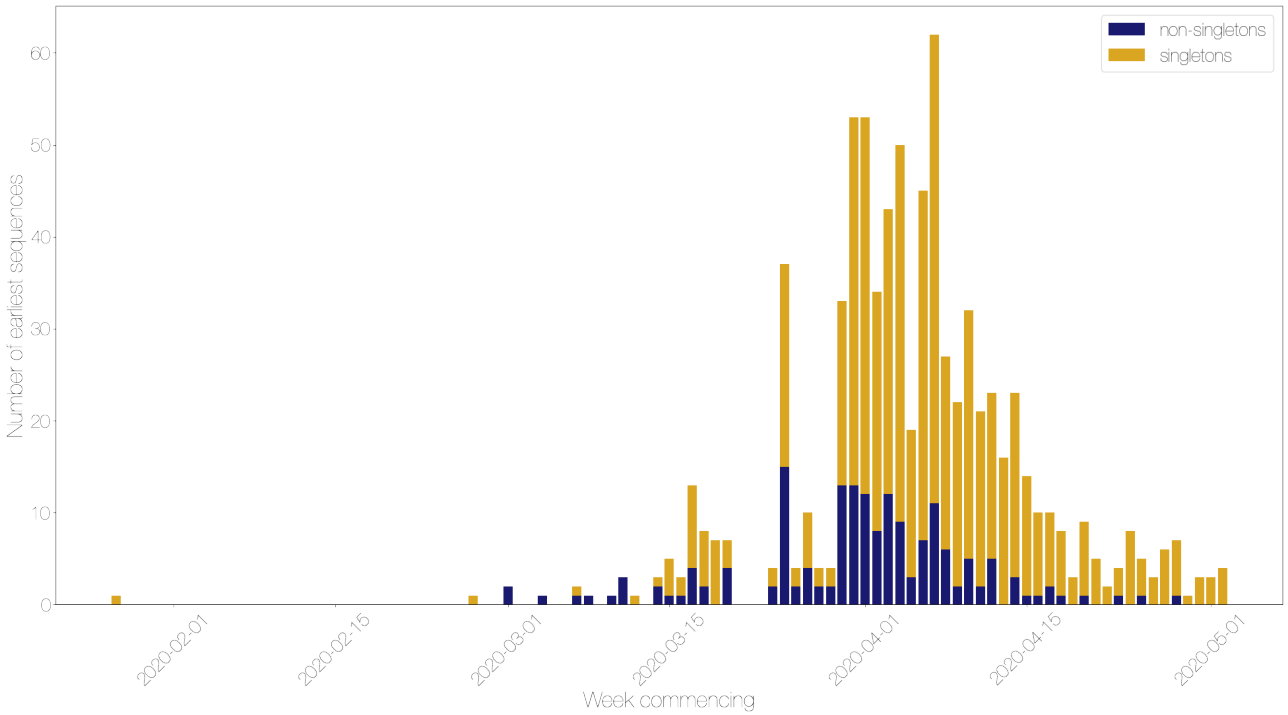


Figure 5: Lineage starts per week, split by singletons and non-singletons

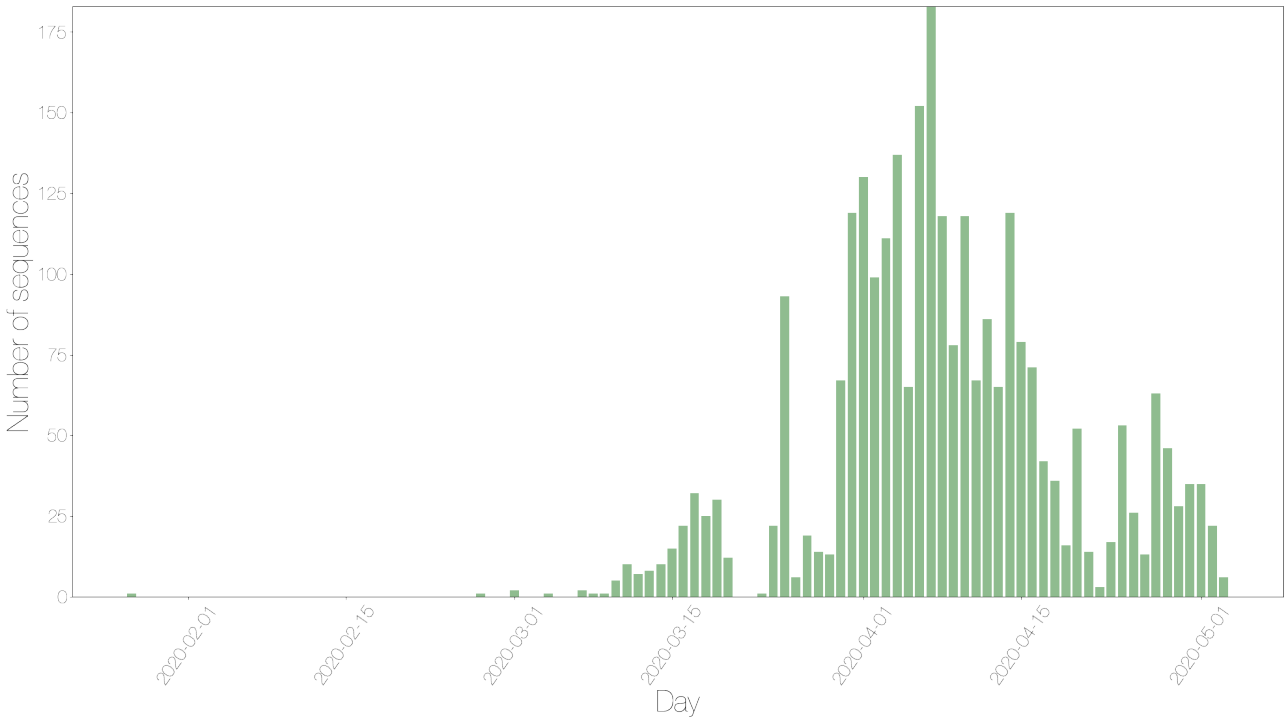
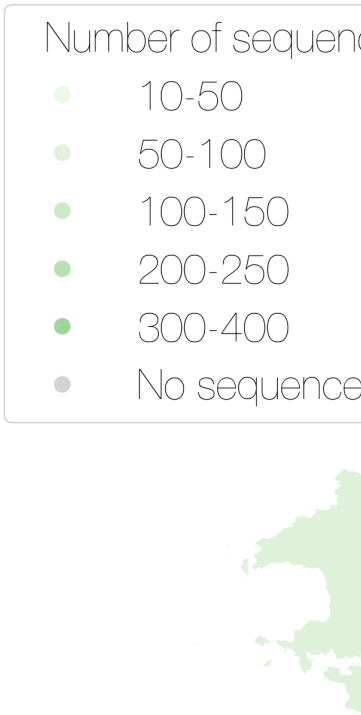


Figure 6: Sequences taken on each day by country



There are 545 sequences without enough geographical information to map from this centre.

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	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK61	340 (100.0%)	Mar-10, May-01	340	B.3	2	0.0767
UK158	142 (100.0%)	Mar-20, May-02	142	B.1.1.2	1	0.305
UK5	125 (100.0%)	Mar-04, May-03	125	B.1.1, B.1.1.1	0	active today
UK42	112 (100.0%)	Mar-07, Apr-27	112	B.1.35, B.1	6	0.0766
UK632	97 (100.0%)	Mar-25, May-02	97	B.1.1	1	0.3958
UK74	96 (100.0%)	Mar-30, May-03	96	B.1	0	active today
UK19	81 (100.0%)	Mar-17, May-02	81	B.1, B.1.44	1	0.575
UK2464	70 (100.0%)	Mar-26, May-02	70	B.1.p11	1	0.5362
UK495	65 (100.0%)	Apr-01, May-02	65	B.1.p11	1	0.4844
UK701	43 (100.0%)	Mar-25, May-01	43	B.1	2	0.4405
UK2200	33 (100.0%)	Mar-15, Apr-26	33	B.1.5, B.1.5.6	7	0.1875
UK187	26 (100.0%)	Apr-01, Apr-28	26	B.1	5	0.216
UK633	24 (100.0%)	Apr-03, May-01	24	B.1.1.p16, B.1.1.16	2	0.6087
UK473	24 (100.0%)	Apr-02, Apr-29	24	B.1.1	4	0.2935
UK86	23 (100.0%)	Mar-30, Apr-28	23	B.1	5	0.2636
UK179	22 (100.0%)	Mar-17, May-01	22	B.1.1.p11	2	1.0714
UK4507	22 (100.0%)	Apr-14, May-03	22	B.1	0	active today
UK298	20 (100.0%)	Mar-27, Apr-29	20	B.1.1	4	0.4342
UK472	18 (100.0%)	Apr-04, Apr-27	18	B.1.1.p11, B.1.1	6	0.2255
UK394	16 (100.0%)	Mar-24, Apr-24	16	B.1.1, B.1.1.10	9	0.2296
UK392	16 (100.0%)	Mar-25, Apr-12	16	B.1.67	21	0.0571

Lineage name	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK5322	16 (100.0%)	Apr-08, May-03	16	B.1.1	0	active today
UK5556	15 (100.0%)	Mar-18, Apr-16	15	B.2.2	17	0.1218
UK322	15 (100.0%)	Mar-30, Apr-26	15	B.1	7	0.2755
UK277	14 (100.0%)	Mar-28, Apr-28	14	B.1.1	5	0.4769
UK603	13 (100.0%)	Mar-29, Apr-09	13	B.1.1	24	0.0382
UK4	13 (100.0%)	Mar-11, Apr-24	13	B	9	0.4074
UK5561	12 (100.0%)	Mar-30, Apr-27	12	B.2.2	6	0.4242
UK5378	12 (100.0%)	Apr-02, Apr-25	12	B.1.1	8	0.2614
UK504	12 (100.0%)	Mar-30, Apr-13	12	B.1.1	20	0.0636
UK530	11 (100.0%)	Mar-31, Apr-13	11	B.1.1	20	0.065
UK211	11 (100.0%)	Mar-24, Apr-30	11	B.1.5	3	1.2333
UK303	10 (100.0%)	Mar-25, Apr-14	10	B.1.1	19	0.117
UK193	10 (100.0%)	Apr-01, May-01	10	B.1.1	2	1.6667
UK571	10 (100.0%)	Apr-06, Apr-29	10	B.1.1	4	0.6389
UK45	10 (100.0%)	Mar-01, Apr-20	10	B.1.1	13	0.4274
UK5681	10 (100.0%)	Apr-03, Apr-27	10	B.2	6	0.4444
UK64	9 (100.0%)	Mar-25, Apr-20	9	B.1	13	0.25
UK471	9 (100.0%)	Apr-02, Apr-24	9	B.1.1	9	0.3056
UK156	9 (100.0%)	Mar-28, Apr-30	9	B.1.71	3	1.375
UK750	9 (100.0%)	Apr-07, Apr-15	9	B.1	18	0.0556
UK635	9 (100.0%)	Apr-07, May-02	9	B.1.1	1	3.125
UK41	9 (100.0%)	Apr-10, Apr-27	9	B.1	6	0.3542
UK3075	8 (100.0%)	Apr-17, May-01	8	B.1.1	2	1.0
UK696	8 (100.0%)	Apr-10, May-01	8	B.1.5, B.1	2	1.5

Lineage name	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK474	8 (100.0%)	Apr-01, Apr-16	8	B.1.1	17	0.1261
UK367	8 (100.0%)	Mar-25, Apr-27	8	B.1	6	0.7857
UK762	7 (100.0%)	Apr-11, Apr-30	7	B.1.1	3	1.0556
UK418	7 (100.0%)	Apr-03, Apr-20	7	B.1.1.10	13	0.2179
UK874	7 (100.0%)	Apr-06, Apr-24	7	B.1	9	0.3333
UK2891	7 (100.0%)	Mar-31, Apr-24	7	B.1.1	9	0.4444
UK339	7 (100.0%)	Mar-25, Apr-14	7	B.3	19	0.1754
UK801	7 (100.0%)	Apr-05, May-02	7	B.1	1	4.5
UK119	7 (100.0%)	Mar-30, Apr-14	7	B.2.5	19	0.1316
UK572	7 (100.0%)	Apr-07, May-01	7	B.1.1	2	2.0
UK439	7 (100.0%)	Apr-02, Apr-20	7	B.1.1	13	0.2308
UK536	7 (100.0%)	Mar-27, Apr-09	7	B.1.1	24	0.0903
UK155	7 (100.0%)	Mar-25, May-03	7	B.1	0	active today
UK612	6 (100.0%)	Mar-31, Apr-11	6	B.2.1	22	0.1
UK5670	6 (100.0%)	Apr-01, Apr-30	6	B.2	3	1.9333
UK627	6 (100.0%)	Mar-31, Apr-10	6	B.1	23	0.087
UK358	6 (100.0%)	Mar-31, Apr-09	6	B.2.1	24	0.075
UK761	6 (100.0%)	Apr-12, Apr-28	6	B.1.1	5	0.64
UK350	6 (100.0%)	Mar-31, Apr-20	6	B.1.1	13	0.3077
UK451	6 (100.0%)	Mar-25, Apr-05	6	B.2.1	28	0.0786
UK462	6 (100.0%)	Apr-01, Apr-16	6	B.1	17	0.1765
UK80	6 (100.0%)	Mar-31, Apr-27	6	B.1.1.p15	6	0.9

Table S2 Raw data for figure three showing the number of admin2 regions a lineage is present in over time

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
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

Table S3 is not appropriate for this report and so has been omitted.

Table S4 Raw data for figure six 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	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	0	3	3
2020-03-12	1	0	1
2020-03-14	1	2	3
2020-03-15	4	1	5
2020-03-16	2	1	3
2020-03-17	9	4	13
2020-03-18	6	2	8
2020-03-19	7	0	7
2020-03-20	3	4	7
2020-03-24	2	2	4
2020-03-25	22	15	37
2020-03-26	2	2	4
2020-03-27	6	4	10
2020-03-28	2	2	4
2020-03-29	2	2	4
2020-03-30	20	13	33
2020-03-31	40	13	53
2020-04-01	41	12	53
2020-04-02	26	8	34
2020-04-03	31	12	43
2020-04-04	41	9	50
2020-04-05	16	3	19
2020-04-06	38	7	45
2020-04-07	51	11	62
2020-04-08	21	6	27
2020-04-09	20	2	22
2020-04-10	27	5	32
2020-04-11	19	2	21
2020-04-12	18	5	23
2020-04-13	16	0	16
2020-04-14	20	3	23
2020-04-15	13	1	14
2020-04-16	9	1	10
2020-04-17	8	2	10
2020-04-18	7	1	8
2020-04-19	3	0	3
2020-04-20	8	1	9
2020-04-21	5	0	5
2020-04-22	2	0	2

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-04-23	3	1	4
2020-04-24	8	0	8
2020-04-25	4	1	5
2020-04-26	3	0	3
2020-04-27	6	0	6
2020-04-28	6	1	7
2020-04-29	1	0	1
2020-04-30	3	0	3
2020-05-01	3	0	3
2020-05-02	4	0	4

Table S5 Raw data for figure seven 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	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	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

Day	Wales
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

Table S6 Raw data for the map with the number of sequences assigned to each admin2 region.

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	367	300-400
CARMARTHENSHIRE	Wales	79	50-100
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	52	50-100
NEATH PORT TALBOT	Wales	94	50-100
NEWPORT	Wales	121	100-150
PEMBROKESHIRE	Wales	62	50-100
POWYS	Wales	46	10-50
SWANSEA	Wales	223	200-250
TORFAEN	Wales	76	50-100
VALE OF GLAMORGAN	Wales	137	100-150
WREXHAM	Wales	73	50-100