Lineages report for PHWC

This report gives summaries of UK specific lineages sequenced by PHWC for week 2020-06-05. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-05-17. The analysis (eg time since last sample) is therefore undertaken from this date. 3208 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 846 and the maximum is 1184

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

765 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 58 that remain: 22 are pending extinction, ie last seen three weeks ago. 13 lineages have gone quiet, ie haven't been seen this week. 4 lineages have reactivated. 19 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 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 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	373 (100.0%)	Mar-10, May-15	373	B.3	2	0.0887
UK158	193 (100.0%)	Mar-20, May-17	193	B.1.1.2	0	active today
UK632	145 (100.0%)	Mar-25, May-15	145	B.1.1	2	0.1771
UK5	140 (100.0%)	Mar-04, May-17	140	B.1.1.1	0	active today
UK3021	139 (100.0%)	Mar-29, May-16	139	B.1	1	0.3478
UK42	135 (100.0%)	Mar-07, May-16	135	B.1, B.1.35	1	0.5224
UK19	95 (100.0%)	Mar-17, May-15	95	B.1.44, B.1	2	0.3138
UK2464	77 (100.0%)	Mar-26, May-11	77	B.1.p11	6	0.1009
UK495	58 (100.0%)	Apr-01, May-14	58	B.1.p11	3	0.2515
UK2916	52 (100.0%)	Mar-25, May-10	52	B.1	7	0.1289

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.

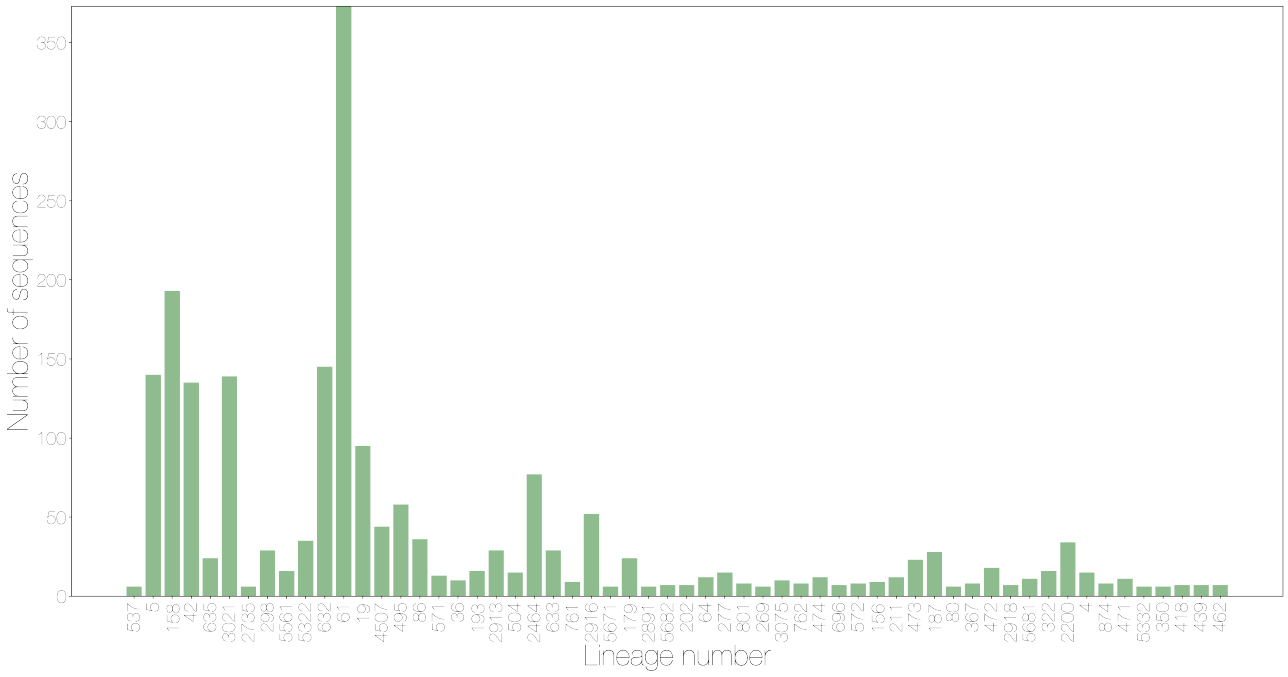

Figure 1: Number of sequences sampled in a lineage by country

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 lag for this sequencing centre is 19 days`

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.

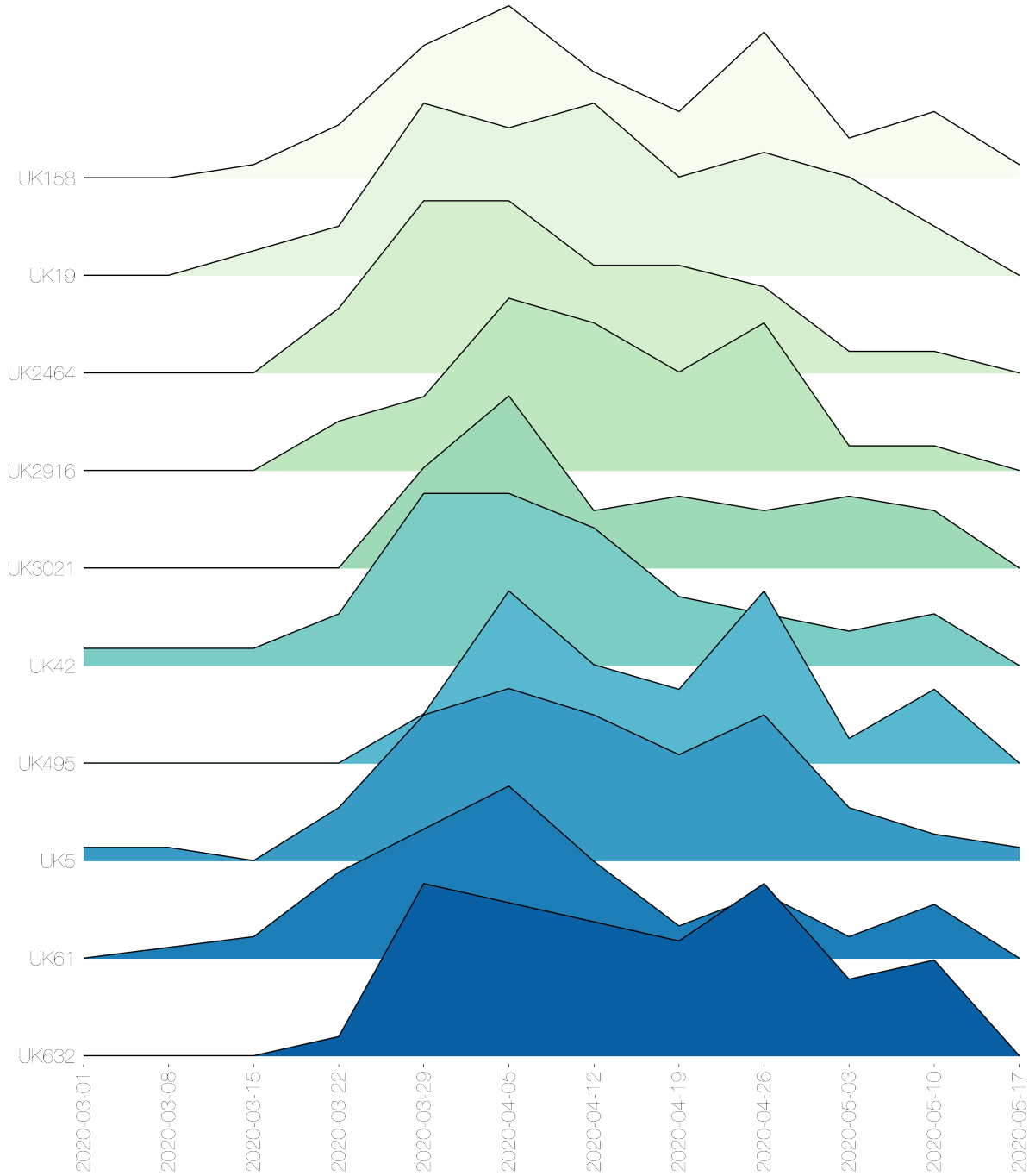

Figure 2: Lineages by number of adm2 regions present by epiweek

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.

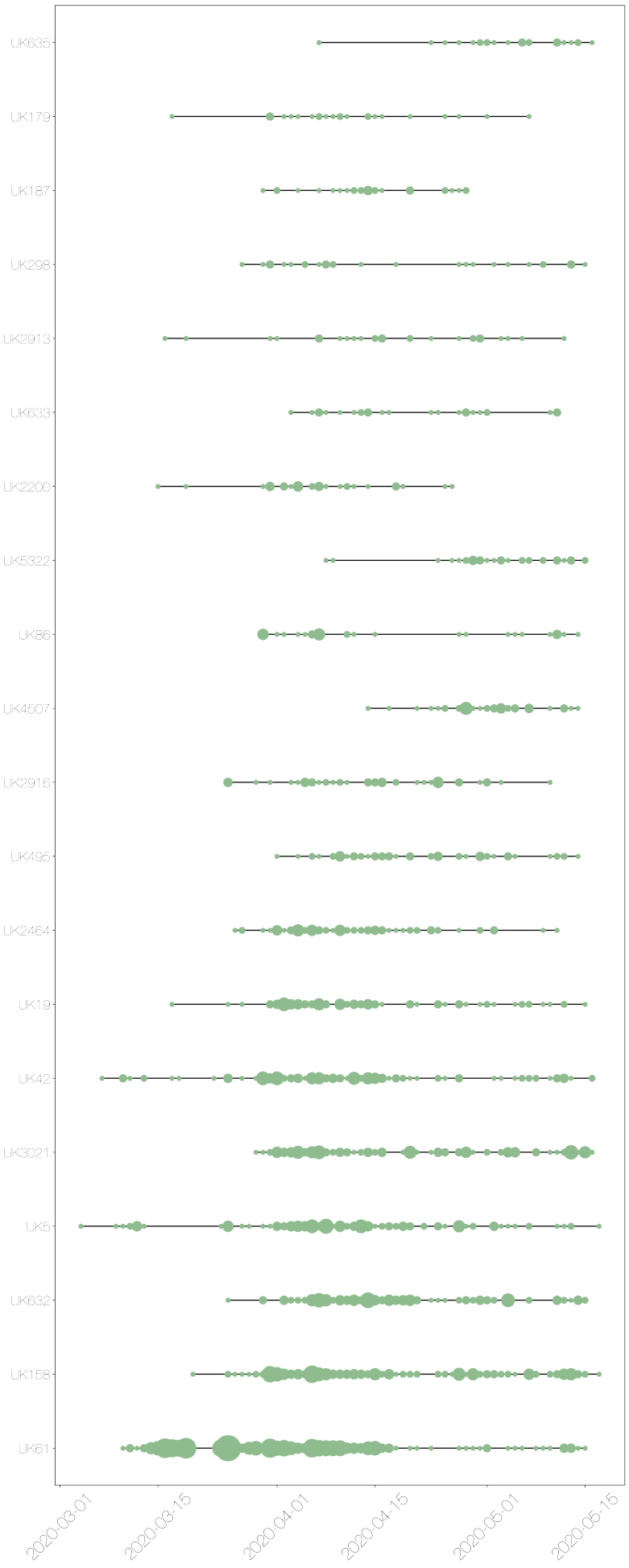

Figure 3: Timeline of lineages, sized by number of sequences from each country.

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

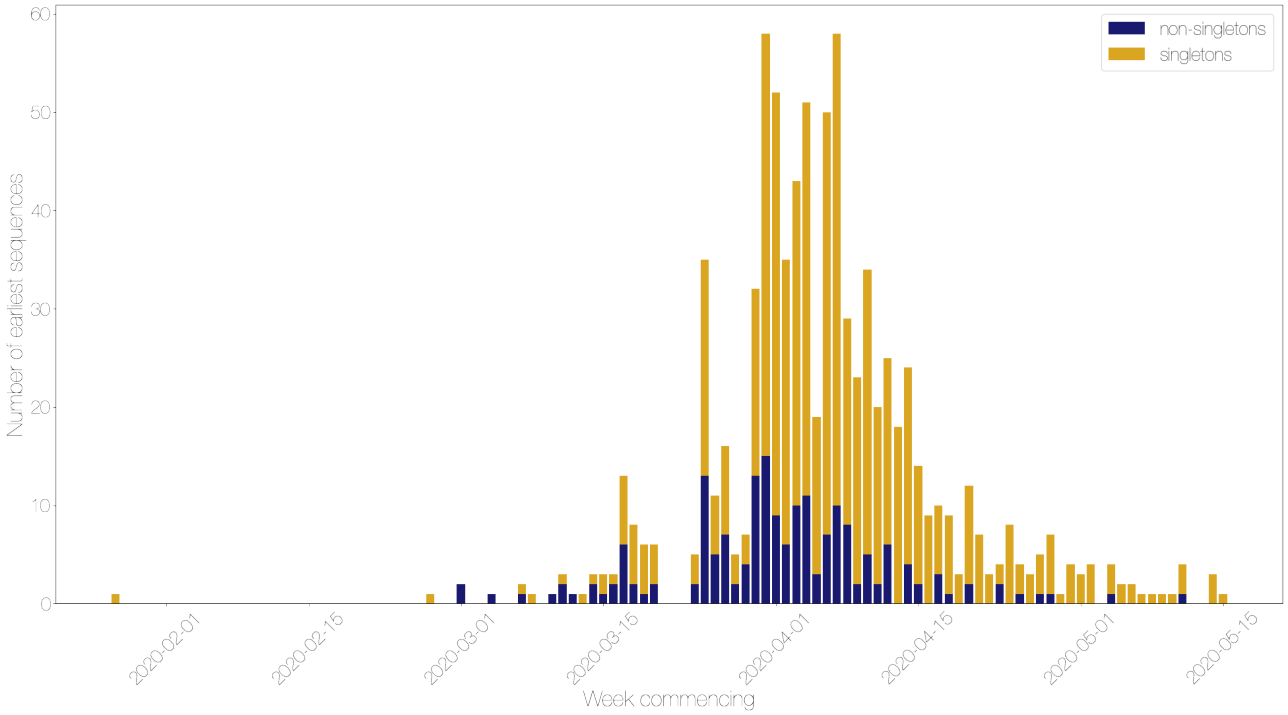

Figure 4: Lineage starts per week, split by singletons and non-singletons

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.

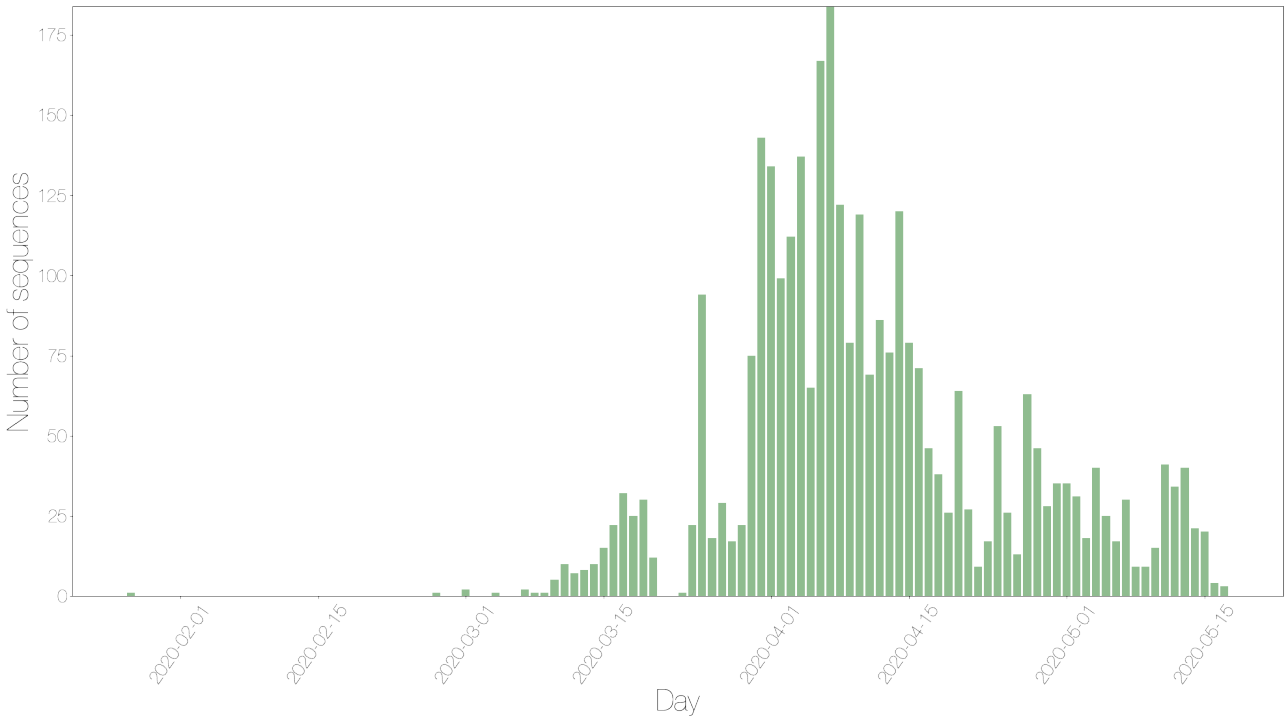

Figure 5: Sequences taken on each day by country

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.

```
There are 565 sequences without enough geographical information to map
```

from this centre.

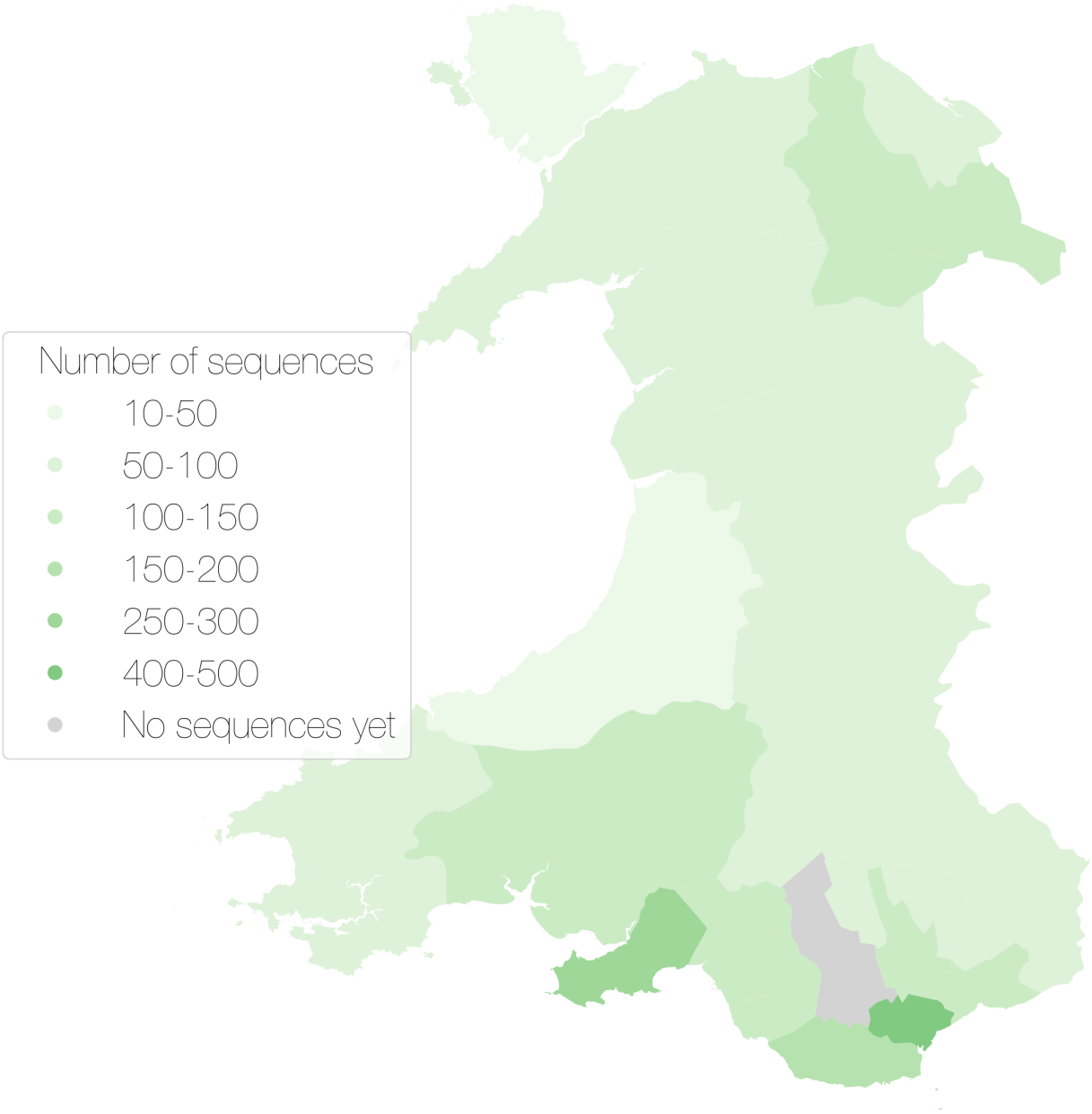

Figure 6: Map showing the number of sequences sampled by adm2 region

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	373 (100.0%)	Mar-10, May-15	373	B.3	2	0.0887
UK158	193 (100.0%)	Mar-20, May-17	193	B.1.1.2	0	active today
UK632	145 (100.0%)	Mar-25, May-15	145	B.1.1	2	0.1771
UK5	140 (100.0%)	Mar-04, May-17	140	B.1.1.1	0	active today
UK3021	139 (100.0%)	Mar-29, May-16	139	B.1	1	0.3478
UK42	135 (100.0%)	Mar-07, May-16	135	B.1, B.1.35	1	0.5224
UK19	95 (100.0%)	Mar-17, May-15	95	B.1.44, B.1	2	0.3138
UK2464	77 (100.0%)	Mar-26, May-11	77	B.1.p11	6	0.1009
UK495	58 (100.0%)	Apr-01, May-14	58	B.1.p11	3	0.2515
UK2916	52 (100.0%)	Mar-25, May-10	52	B.1	7	0.1289
UK4507	44 (100.0%)	Apr-14, May-14	44	B.1	3	0.2326
UK86	36 (100.0%)	Mar-30, May-14	36	B.1	3	0.4286
UK5322	35 (100.0%)	Apr-08, May-15	35	B.1.1	2	0.5441
UK2200	34 (100.0%)	Mar-15, Apr-26	34	B.1.5, B.1.5.6	21	0.0606
UK298	29 (100.0%)	Mar-27, May-15	29	B.1.1	2	0.875
UK2913	29 (100.0%)	Mar-16, May-12	29	B.1.p11	5	0.4071
UK633	29 (100.0%)	Apr-03, May-11	29	B.1.1.16, B.1.1.p16	6	0.2262
UK187	28 (100.0%)	Mar-30, Apr-28	28	B.1	19	0.0565
UK179	24 (100.0%)	Mar-17, May-07	24	B.1.1.p11	10	0.2217
UK635	24 (100.0%)	Apr-07, May-16	24	B.1.1	1	1.6957
UK473	23 (100.0%)	Apr-02, Apr-29	23	B.1.1	18	0.0682
UK394	23 (100.0%)	Mar-17, Apr-17	23	B.1.1	30	0.047
UK472	18 (100.0%)	Apr-04, Apr-27	18	B.1.1.p11, B.1.1	20	0.0676
UK5556	17 (100.0%)	Mar-18, Apr-16	17	B.2.2	31	0.0585
UK392	16 (100.0%)	Mar-25, Apr-12	16	B.1.67	35	0.0343
UK5561	16 (100.0%)	Mar-30, May-15	16	B.2.2	2	1.5333

Lineage name	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK193	16 (100.0%)	Apr-01, May-13	16	B.1.1	4	0.7
UK322	16 (100.0%)	Mar-29, Apr-26	16	B.1	21	0.0889
UK4	15 (100.0%)	Mar-11, Apr-24	15	B	23	0.1366
UK504	15 (100.0%)	Mar-30, May-12	15	B.1.1	5	0.6143
UK277	15 (100.0%)	Mar-28, May-05	15	B.1.1	12	0.2262
UK339	13 (100.0%)	Mar-14, Apr-14	13	B.3	33	0.0783
UK603	13 (100.0%)	Mar-29, Apr-09	13	B.1.1	38	0.0241
UK571	13 (100.0%)	Apr-06, May-13	13	B.1.1	4	0.7708
UK211	12 (100.0%)	Mar-24, Apr-30	12	B.1.5	17	0.1979
UK64	12 (100.0%)	Mar-25, May-05	12	B.1	12	0.3106
UK474	12 (100.0%)	Apr-01, May-02	12	B.1.1	15	0.1879
UK471	11 (100.0%)	Mar-26, Apr-24	11	B.1.1	23	0.1261
UK5681	11 (100.0%)	Apr-03, Apr-27	11	B.2, B.1.1	20	0.12
UK3075	10 (100.0%)	Apr-17, May-04	10	B.1.1	13	0.1453
UK36	10 (100.0%)	Apr-04, May-13	10	B.1	4	1.0833
UK303	10 (100.0%)	Mar-25, Apr-14	10	B.1.1	33	0.0673
UK750	9 (100.0%)	Apr-07, Apr-15	9	B.1	32	0.0312
UK761	9 (100.0%)	Apr-12, May-10	9	B.1.1	7	0.5
UK156	9 (100.0%)	Mar-28, Apr-30	9	B.1.71	17	0.2426
UK874	8 (100.0%)	Mar-26, Apr-24	8	B.1	23	0.1801
UK801	8 (100.0%)	Apr-05, May-05	8	B.1	12	0.3571
UK367	8 (100.0%)	Mar-25, Apr-27	8	B.1	20	0.2357
UK572	8 (100.0%)	Apr-07, May-01	8	B.1.1	16	0.2143
UK762	8 (100.0%)	Apr-11, May-04	8	B.1.1	13	0.2527
UK696	7 (100.0%)	Apr-10, May-01	7	B.1.5, B.1	16	0.2188
UK5682	7 (100.0%)	Apr-08, May-06	7	B.2, B.1.1	11	0.4242
UK536	7 (100.0%)	Mar-27, Apr-09	7	B.1.1	38	0.057
UK462	7 (100.0%)	Apr-01, Apr-20	7	B.1	27	0.1173
UK439	7 (100.0%)	Apr-02, Apr-20	7	B.1.1	27	0.1111

Lineage name	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK418	7 (100.0%)	Apr-03, Apr-20	7	B.1.1.10	27	0.1049
UK2918	7 (100.0%)	Apr-12, Apr-27	7	B.1	20	0.125
UK119	7 (100.0%)	Mar-30, Apr-14	7	B.2.5	33	0.0758
UK202	7 (100.0%)	Apr-28, May-05	7	B.1.1	12	0.0972
UK2891	6 (100.0%)	Mar-27, May-06	6	B.1.1	11	0.7273
UK269	6 (100.0%)	Mar-31, May-04	6	B.1.1	13	0.5231
UK537	6 (100.0%)	Apr-07, May-17	6	B.1.1	0	active today
UK5332	6 (100.0%)	Mar-01, Apr-20	6	B.1.1	27	0.3704
UK80	6 (100.0%)	Mar-31, Apr-27	6	B.1.1.p15	20	0.27
UK530	6 (100.0%)	Mar-31, Apr-08	6	B.1.1	39	0.041
UK451	6 (100.0%)	Mar-25, Apr-05	6	B.2.1	42	0.0524
UK350	6 (100.0%)	Mar-31, Apr-20	6	B.1.1	27	0.1481
UK5671	6 (100.0%)	Mar-31, May-09	6	B.2, B.1.1	8	0.975
UK358	6 (100.0%)	Mar-31, Apr-09	6	B.2.1	38	0.0474
UK612	6 (100.0%)	Mar-31, Apr-11	6	B.2.1	36	0.0611
UK2735	6 (100.0%)	Mar-30, May-15	6	B.1.1	2	4.6

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	19

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

Week commencing	UK61	UK158	UK632	UK5	UK3021	UK42	UK19	UK2464	UK495	UK2916
2020-03-01	0	0	0	1	0	1	0	0	0	0
2020-03-08	1	0	0	1	0	1	0	0	0	0
2020-03-15	2	1	0	0	0	1	1	0	0	0
2020-03-22	8	4	1	4	0	3	2	3	0	2
2020-03-29	12	10	9	11	7	10	7	8	2	3
2020-04-05	16	13	8	13	12	10	6	8	7	7
2020-04-12	9	8	7	11	4	8	7	5	4	6
2020-04-19	3	5	6	8	5	4	4	5	3	4
2020-04-26	6	11	9	11	4	3	5	4	7	6
2020-05-03	2	3	4	4	5	2	4	1	1	1
2020-05-10	5	5	5	2	4	3	2	1	3	1
2020-05-17	0	1	0	1	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
2020-01-27	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	1	0	1
2020-03-10	0	1	1
2020-03-11	1	2	3
2020-03-12	0	1	1
2020-03-13	1	0	1
2020-03-14	1	2	3
2020-03-15	2	1	3
2020-03-16	1	2	3
2020-03-17	7	6	13
2020-03-18	6	2	8
2020-03-19	5	1	6
2020-03-20	4	2	6
2020-03-24	3	2	5
2020-03-25	22	13	35
2020-03-26	6	5	11
2020-03-27	9	7	16
2020-03-28	3	2	5
2020-03-29	3	4	7
2020-03-30	19	13	32
2020-03-31	43	15	58
2020-04-01	43	9	52
2020-04-02	29	6	35
2020-04-03	33	10	43
2020-04-04	40	11	51
2020-04-05	16	3	19
2020-04-06	43	7	50
2020-04-07	48	10	58
2020-04-08	21	8	29
2020-04-09	21	2	23
2020-04-10	29	5	34
2020-04-11	18	2	20
2020-04-12	19	6	25
2020-04-13	18	0	18
2020-04-14	20	4	24
2020-04-15	12	2	14
2020-04-16	9	0	9
2020-04-17	7	3	10
2020-04-18	8	1	9
2020-04-19	3	0	3
2020-04-20	10	2	12
2020-04-21	7	0	7
2020-04-22	3	0	3
2020-04-23	2	2	4
2020-04-24	8	0	8
2020-04-25	3	1	4
2020-04-26	3	0	3
2020-04-27	4	1	5
2020-04-28	6	1	7
2020-04-29	1	0	1
2020-04-30	4	0	4
2020-05-01	3	0	3
2020-05-02	4	0	4

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-05-04	3	1	4
2020-05-05	2	0	2
2020-05-06	2	0	2
2020-05-07	1	0	1
2020-05-08	1	0	1
2020-05-09	1	0	1
2020-05-10	1	0	1
2020-05-11	3	1	4
2020-05-14	3	0	3
2020-05-15	1	0	1

Table S6 Raw data for figure six 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	94
2020-03-26	18
2020-03-27	29
2020-03-28	17
2020-03-29	22
2020-03-30	75
2020-03-31	143
2020-04-01	134
2020-04-02	99
2020-04-03	112
2020-04-04	137
2020-04-05	65
2020-04-06	167
2020-04-07	184
2020-04-08	122
2020-04-09	79
2020-04-10	119
2020-04-11	69
2020-04-12	86
2020-04-13	76
2020-04-14	120
2020-04-15	79
2020-04-16	71
2020-04-17	46
2020-04-18	38
2020-04-19	26
2020-04-20	64
2020-04-21	27
2020-04-22	9
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

Day	Wales
2020-05-01	35
2020-05-02	31
2020-05-03	18
2020-05-04	40
2020-05-05	25
2020-05-06	17
2020-05-07	30
2020-05-08	9
2020-05-09	9
2020-05-10	15
2020-05-11	41
2020-05-12	34
2020-05-13	40
2020-05-14	21
2020-05-15	20
2020-05-16	4
2020-05-17	3

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	36	10-50
BLAENAU GWENT	Wales	52	50-100
BRIDGEND	Wales	104	100-150
CAERPHILLY	Wales	121	100-150
CARDIFF	Wales	429	400-500
CARMARTHENSHIRE	Wales	113	100-150
CEREDIGION	Wales	16	10-50
CONWY	Wales	88	50-100
DENBIGHSHIRE	Wales	115	100-150
FLINTSHIRE	Wales	79	50-100
GWYNEDD	Wales	69	50-100
MERTHYR TYDFIL	Wales	67	50-100
MONMOUTHSHIRE	Wales	62	50-100
NEATH PORT TALBOT	Wales	107	100-150
NEWPORT	Wales	144	100-150
PEMBROKESHIRE	Wales	67	50-100
POWYS	Wales	60	50-100
SWANSEA	Wales	252	250-300
TORFAEN	Wales	85	50-100
VALE OF GLAMORGAN	Wales	159	150-200
WREXHAM	Wales	102	100-150