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 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	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK61	Mar-10, May-01	340	B.3	2	0.072
UK158	Mar-20, May-02	142	B.1.1.2	1	0.259
UK5	Mar-04, May-03	125	B.1.1.1, B.1.1	0	-0.0035
UK42	Mar-07, Apr-27	112	B.1, B.1.35	6	-0.123
UK632	Mar-25, May-02	97	B.1.1	1	-0.0262

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK74	Mar-30, May-03	96	B.1	0	active today
UK19	Mar-17, May-02	81	B.1.44, B.1	1	-0.0408
UK2464	Mar-26, May-02	70	B.1.p11	1	-0.0177
UK495	Apr-01, May-02	65	B.1.p11	1	0.5455
UK701	Mar-25, May-01	43	B.1	2	-0.046

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.

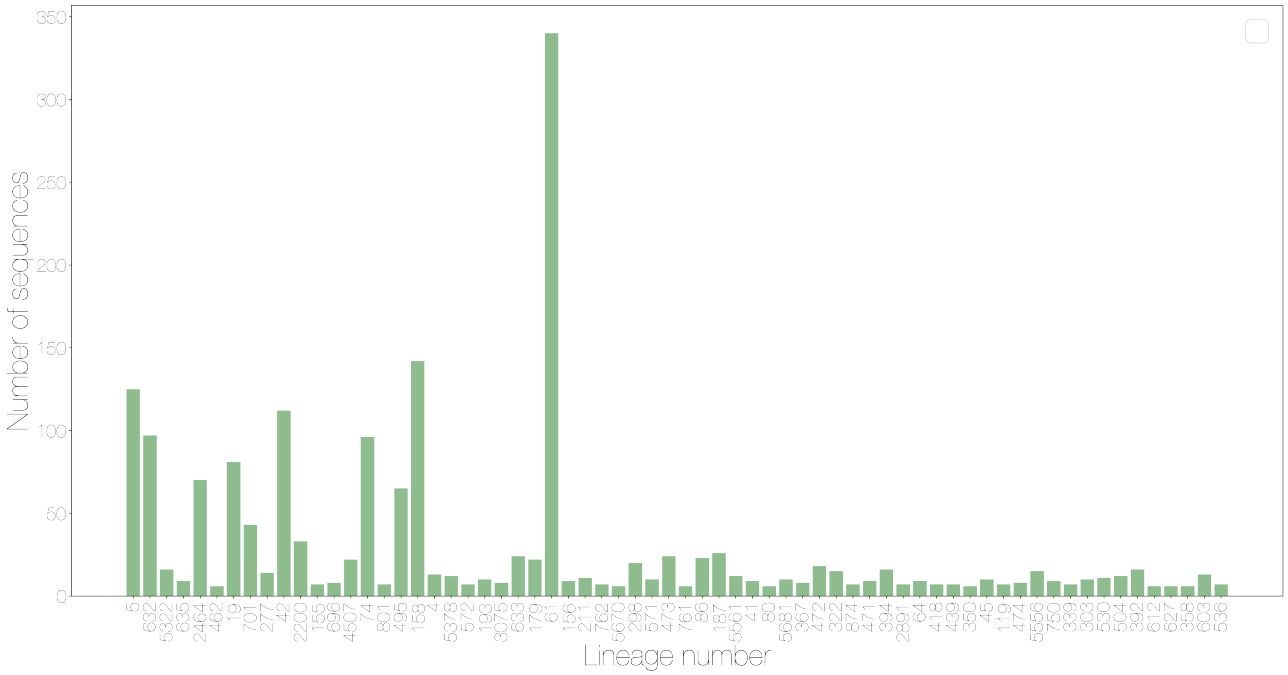

Figure 1: Number of sequences sampled in a lineage by country

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.

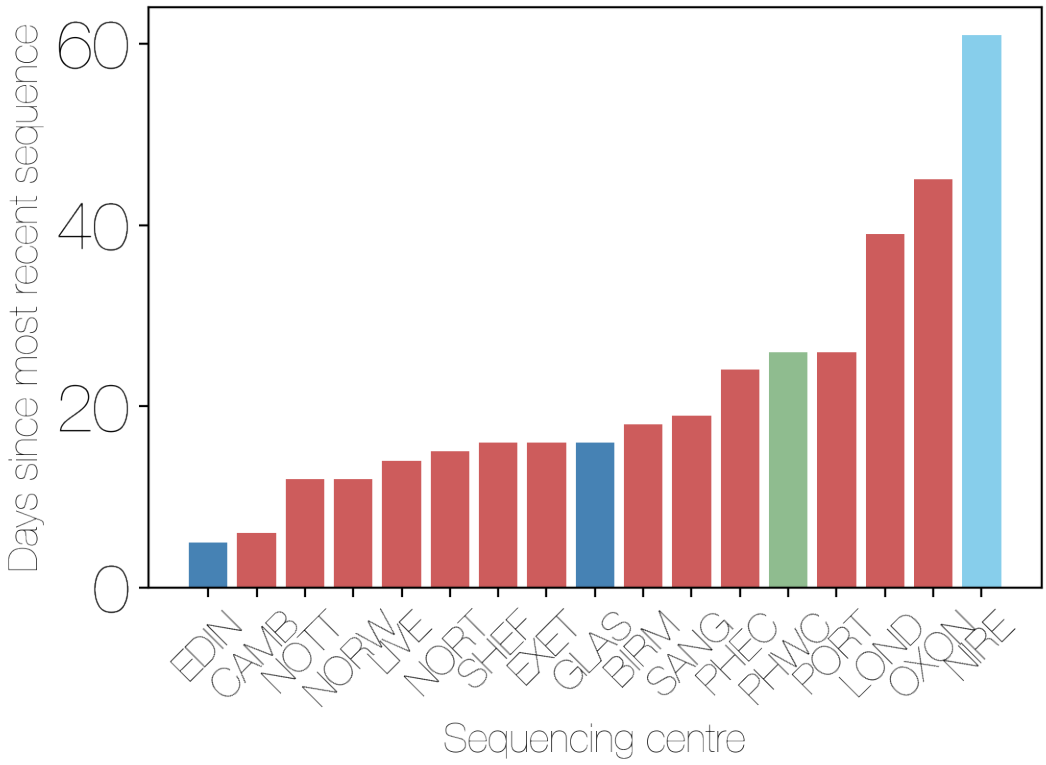

Figure 2: Lag since the most recent sequence from each sequencing centre to most current date

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

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.

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 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.

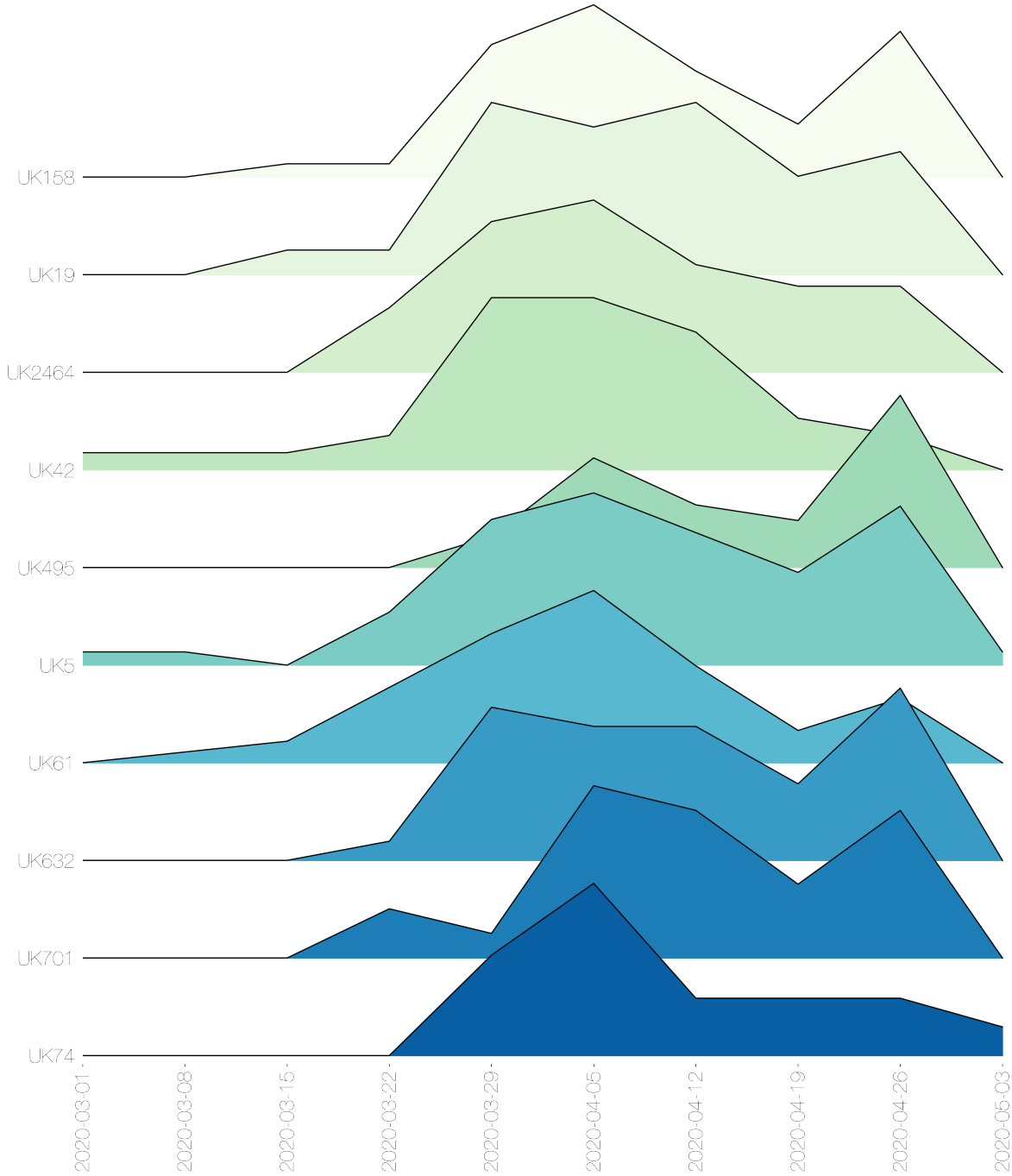

Figure 3: Lineages by number of adm2 regions present by epiweek

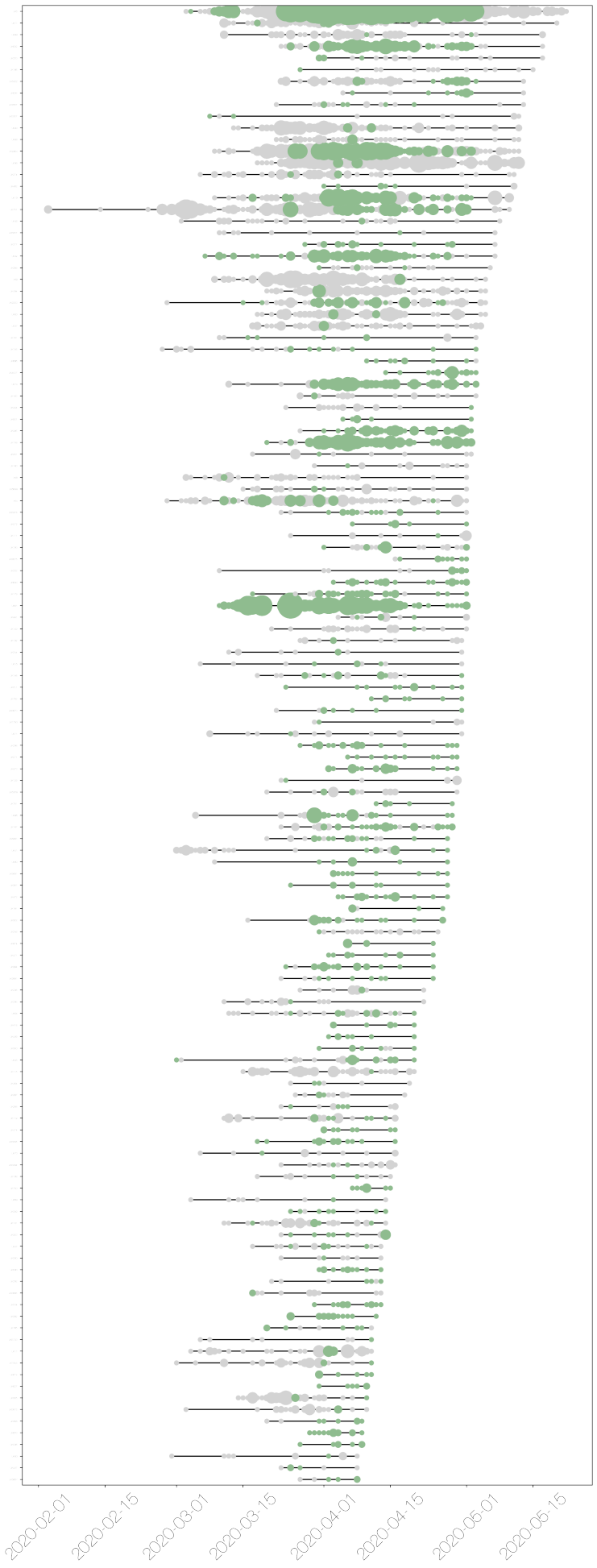

Figure 4: Timeline of lineages, sized by number of sequences from each country.

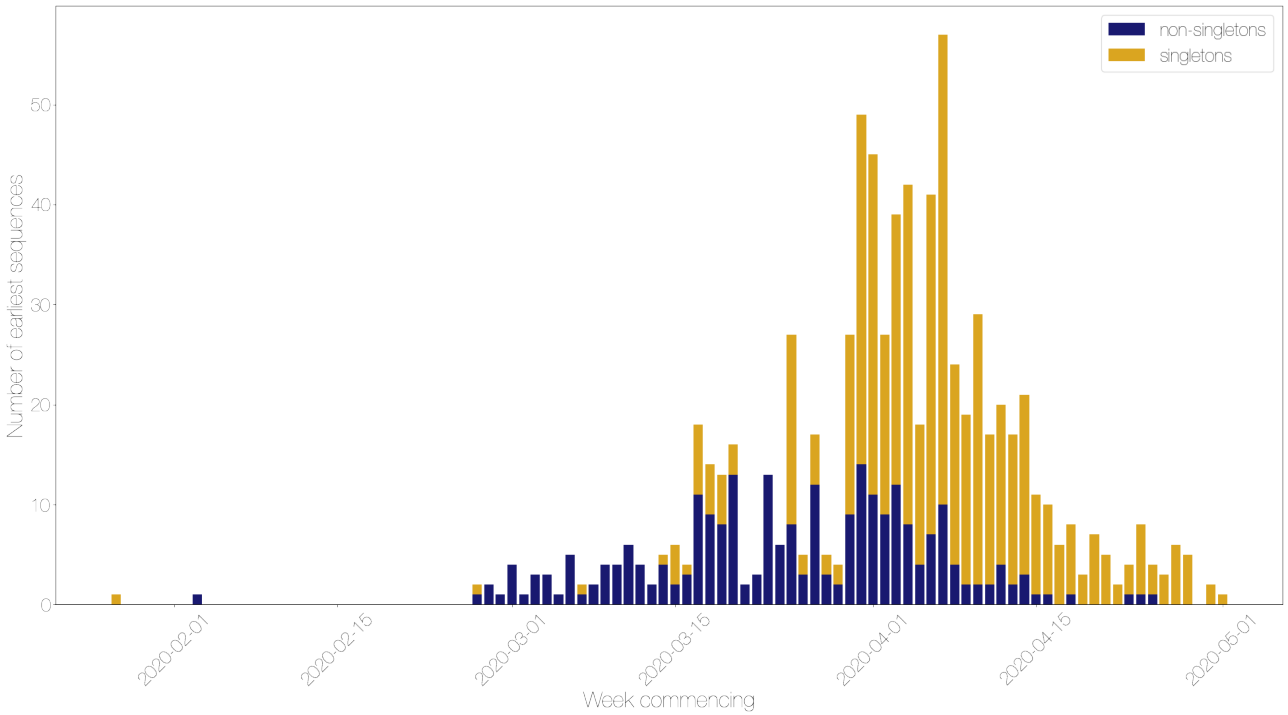

Figure 5: Lineage starts per week, split by singletons and non-singletons

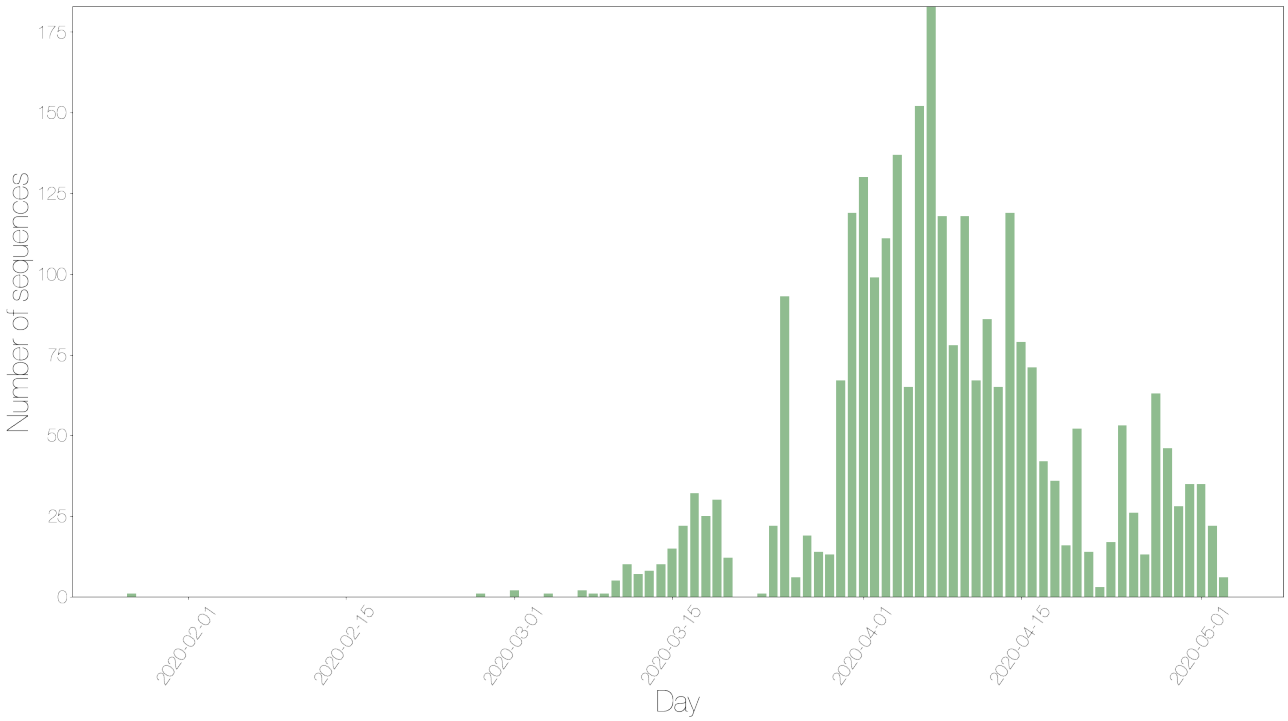

Figure 6: Sequences taken on each day by country

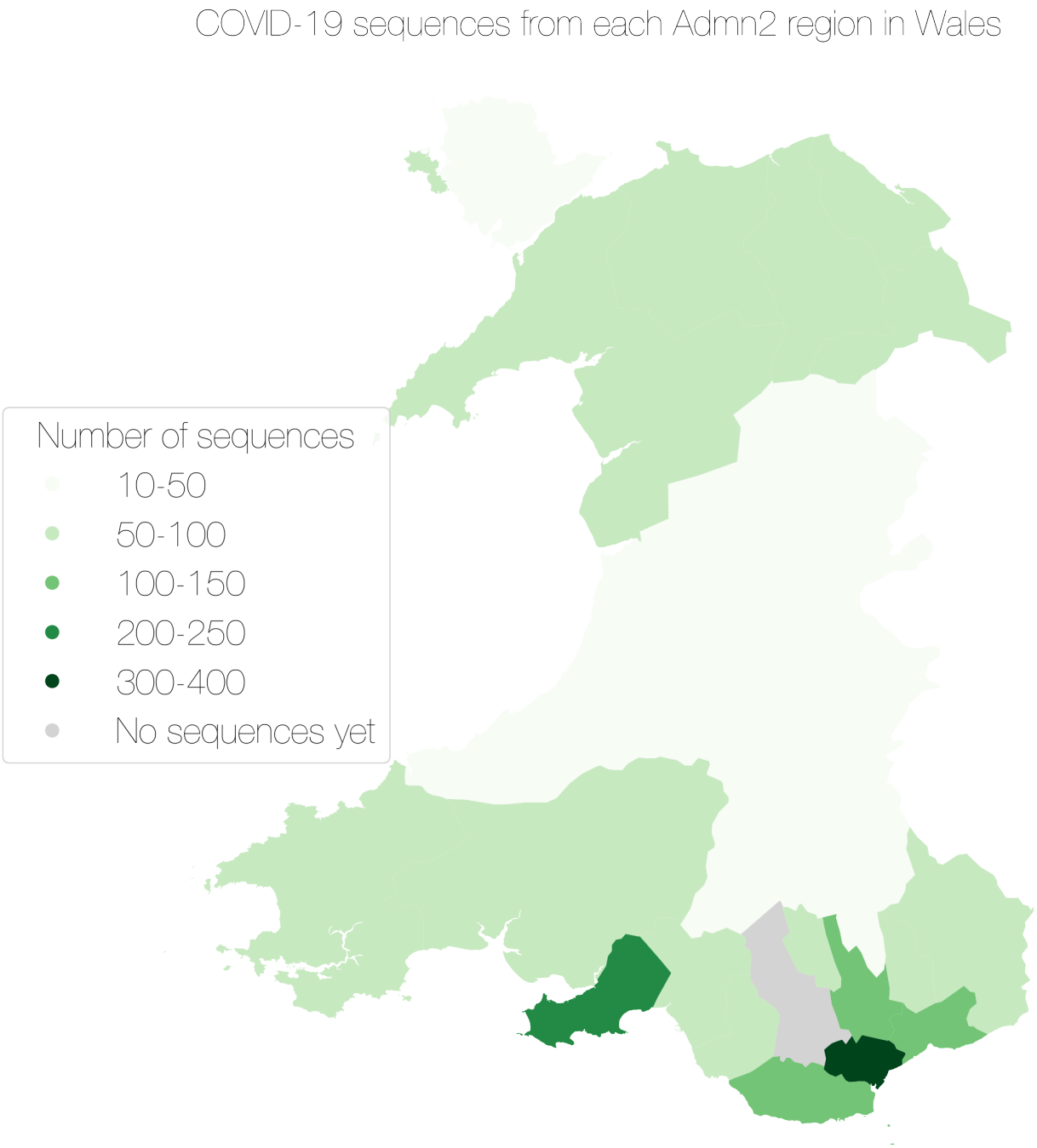

Figure 7: Map showing the number of sequences sampled by adm2 region

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	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK61	Mar-10, May-01	340	B.3	2	0.072
UK158	Mar-20, May-02	142	B.1.1.2	1	0.259
UK5	Mar-04, May-03	125	B.1.1.1, B.1.1	0	-0.0035
UK42	Mar-07, Apr-27	112	B.1, B.1.35	6	-0.123
UK632	Mar-25, May-02	97	B.1.1	1	-0.0262
UK74	Mar-30, May-03	96	B.1	0	active today
UK19	Mar-17, May-02	81	B.1.44, B.1	1	-0.0408
UK2464	Mar-26, May-02	70	B.1.p11	1	-0.0177
UK495	Apr-01, May-02	65	B.1.p11	1	0.5455
UK701	Mar-25, May-01	43	B.1	2	-0.046
UK2200	Mar-15, Apr-26	33	B.1.5.6, B.1.5	7	-0.3764
UK187	Apr-01, Apr-28	26	B.1	5	0.1846
UK473	Apr-02, Apr-29	24	B.1.1	4	0.2935
UK633	Apr-03, May-01	24	B.1.1.p16, B.1.1.16	2	0.6087
UK86	Mar-30, Apr-28	23	B.1	5	0.2842
UK4507	Apr-14, May-03	22	B.1	0	active today
UK179	Mar-17, May-01	22	B.1.1.p11	2	0.6429
UK298	Mar-27, Apr-29	20	B.1.1	4	0.4342
UK472	Apr-04, Apr-27	18	B.1.1.p11, B.1.1	6	0.2255
UK392	Mar-25, Apr-12	16	B.1.67	21	0.0571
UK5322	Apr-08, May-03	16	B.1.1	0	-0.0689

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK394	Mar-24, Apr-24	16	B.1.1.10, B.1.1	9	0.1722
UK5556	Mar-18, Apr-16	15	B.2.2	17	0.1218
UK322	Mar-30, Apr-26	15	B.1	7	0.2662
UK277	Mar-28, Apr-28	14	B.1.1	5	-0.625
UK603	Mar-29, Apr-09	13	B.1.1	24	0.0382
UK4	Mar-11, Apr-24	13	B	9	0.2086
UK504	Mar-30, Apr-13	12	B.1.1	20	0.0636
UK5378	Apr-02, Apr-25	12	B.1.1	8	1.1471
UK5561	Mar-30, Apr-27	12	B.2.2	6	0.3167
UK211	Mar-24, Apr-30	11	B.1.5	3	1.2333
UK530	Mar-31, Apr-13	11	B.1.1	20	0.065
UK571	Apr-06, Apr-29	10	B.1.1	4	0.6389
UK45	Mar-01, Apr-20	10	B.1.1	13	0.1538
UK5681	Apr-03, Apr-27	10	B.2	6	0.4444
UK193	Apr-01, May-01	10	B.1.1	2	0.5556
UK303	Mar-25, Apr-14	10	B.1.1	19	0.0891
UK64	Mar-25, Apr-20	9	B.1	13	0.0938
UK471	Apr-02, Apr-24	9	B.1.1	9	0.3056
UK750	Apr-07, Apr-15	9	B.1	18	0.0556
UK41	Apr-10, Apr-27	9	B.1	6	0.2879
UK635	Apr-07, May-02	9	B.1.1	1	-0.2923
UK156	Mar-28, Apr-30	9	B.1.71	3	0.6515
UK367	Mar-25, Apr-27	8	B.1	6	0.7857
UK474	Apr-01, Apr-16	8	B.1.1	17	0.1261

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK3075	Apr-17, May-01	8	B.1.1	2	0.9375
UK696	Apr-10, May-01	8	B.1, B.1.5	2	active today
UK572	Apr-07, May-01	7	B.1.1	2	2.0
UK339	Mar-25, Apr-14	7	B.3	19	0.1316
UK2891	Mar-31, Apr-24	7	B.1.1	9	0.3951
UK874	Apr-06, Apr-24	7	B.1	9	0.3333
UK801	Apr-05, May-02	7	B.1	1	4.5
UK536	Mar-27, Apr-09	7	B.1.1	24	0.0903
UK119	Mar-30, Apr-14	7	B.2.5	19	0.0706
UK762	Apr-11, Apr-30	7	B.1.1	3	1.0556
UK155	Mar-25, May-03	7	B.1	0	active today
UK439	Apr-02, Apr-20	7	B.1.1	13	0.2308
UK418	Apr-03, Apr-20	7	B.1.1.10	13	0.2179
UK761	Apr-12, Apr-28	6	B.1.1	5	0.64
UK358	Mar-31, Apr-09	6	B.2.1	24	0.119
UK5670	Apr-01, Apr-30	6	B.2	3	1.8571
UK80	Mar-31, Apr-27	6	B.1.1.p15	6	1.0208
UK612	Mar-31, Apr-11	6	B.2.1	22	0.0833
UK462	Apr-01, Apr-16	6	B.1	17	-0.625
UK350	Mar-31, Apr-20	6	B.1.1	13	0.2198
UK627	Mar-31, Apr-10	6	B.1	23	0.087
UK451	Mar-25, Apr-05	6	B.2.1	28	0.0952

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
2020-02-03	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
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

Day	Number of singleton starts	Number of non-singleton starts	Total
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

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	368	300-400
CARMARTHENSHIRE	Wales	80	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	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