UK lineages summary report

This report gives summaries of lineages sampled in Scotland for week 2020-05-29. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-05-24. The analysis (eg time since last sample) is therefore undertaken from this date. 2520 sequences from Scotland have been included in this analysis. 725 lineages have been recorded, 554 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 726 and the maximum is 928

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

664 are lineages which were sampled less than five times in Scotland, 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 61 that remain: 22 are pending extinction, ie last seen three weeks ago. 20 have not been seen for more than one month, and so are viewed as extinct, but will continue to be monitored. 10 lineages have gone quiet, ie haven't been seen this week. 3 lineages have reactivated. 6 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
UK36	Mar-20, May-11	216	B.1	13	0.0153
UK52	Mar-16, May-12	175	B.1.p73	12	0.0268
UK40	Mar-13, May-12	126	B, B.16	12	0.0376
UK2464	Mar-19, May-12	92	B.1.p11	12	0.0133
UK39	Mar-12, May-10	80	A.2	14	0.0533

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK5	Mar-13, May-18	70	B.1.1.1	6	0.0334
UK175	Mar-22, May-04	67	B.1	20	0.0326
UK225	Mar-14, Apr-07	50	B.2	47	0.0112
UK88	Mar-22, May-12	40	B.1	12	0.1062
UK44	Mar-17, May-01	39	B	23	0.0502

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.

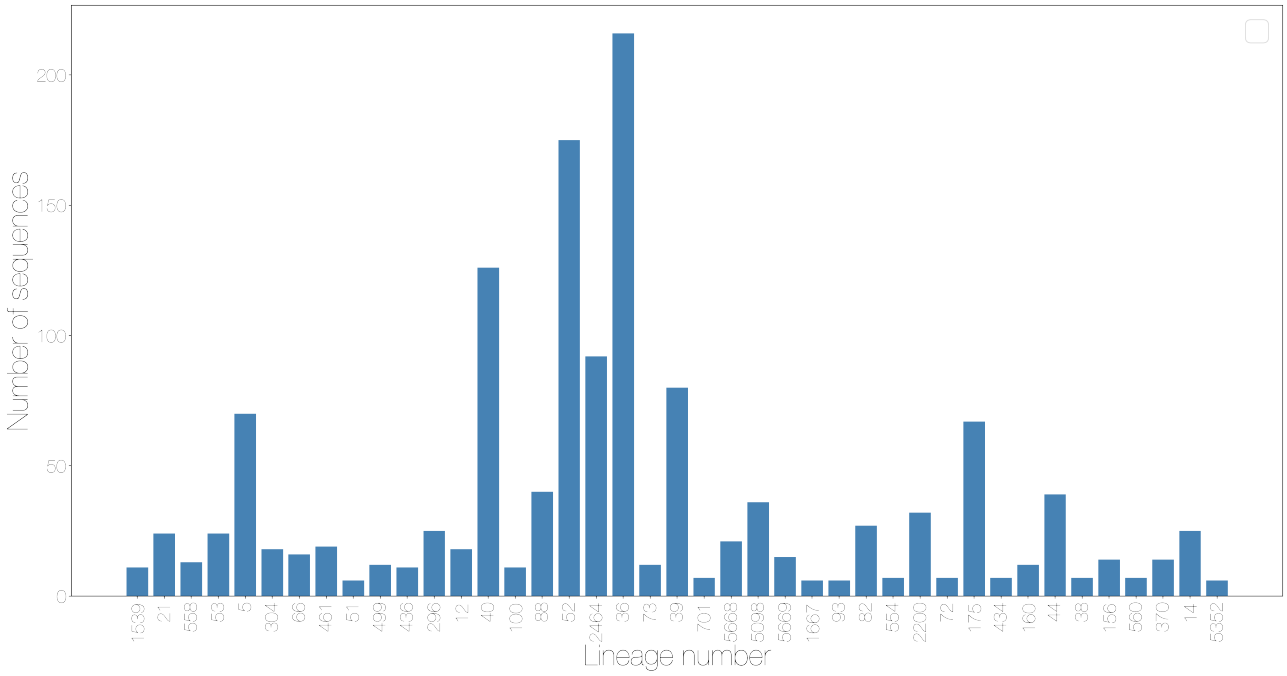

Figure 1: Number of sequences sampled in a lineage by country

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.

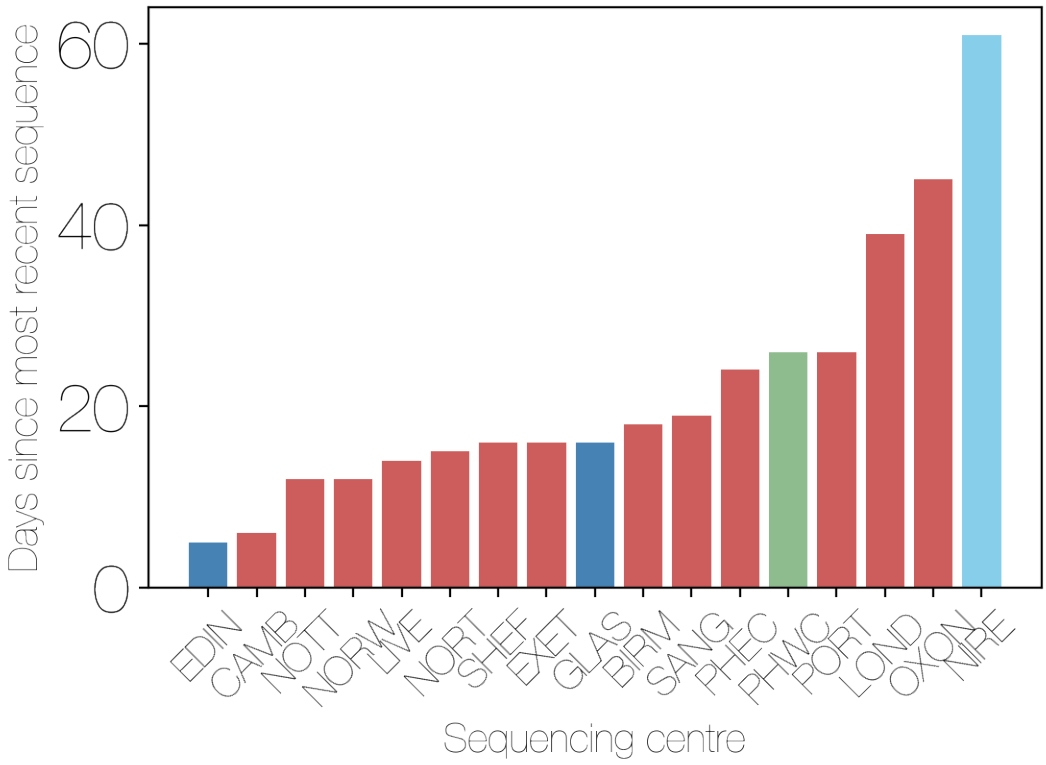

Figure 2: Lag since the most recent sequence from each sequencing centre to most current date

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

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.

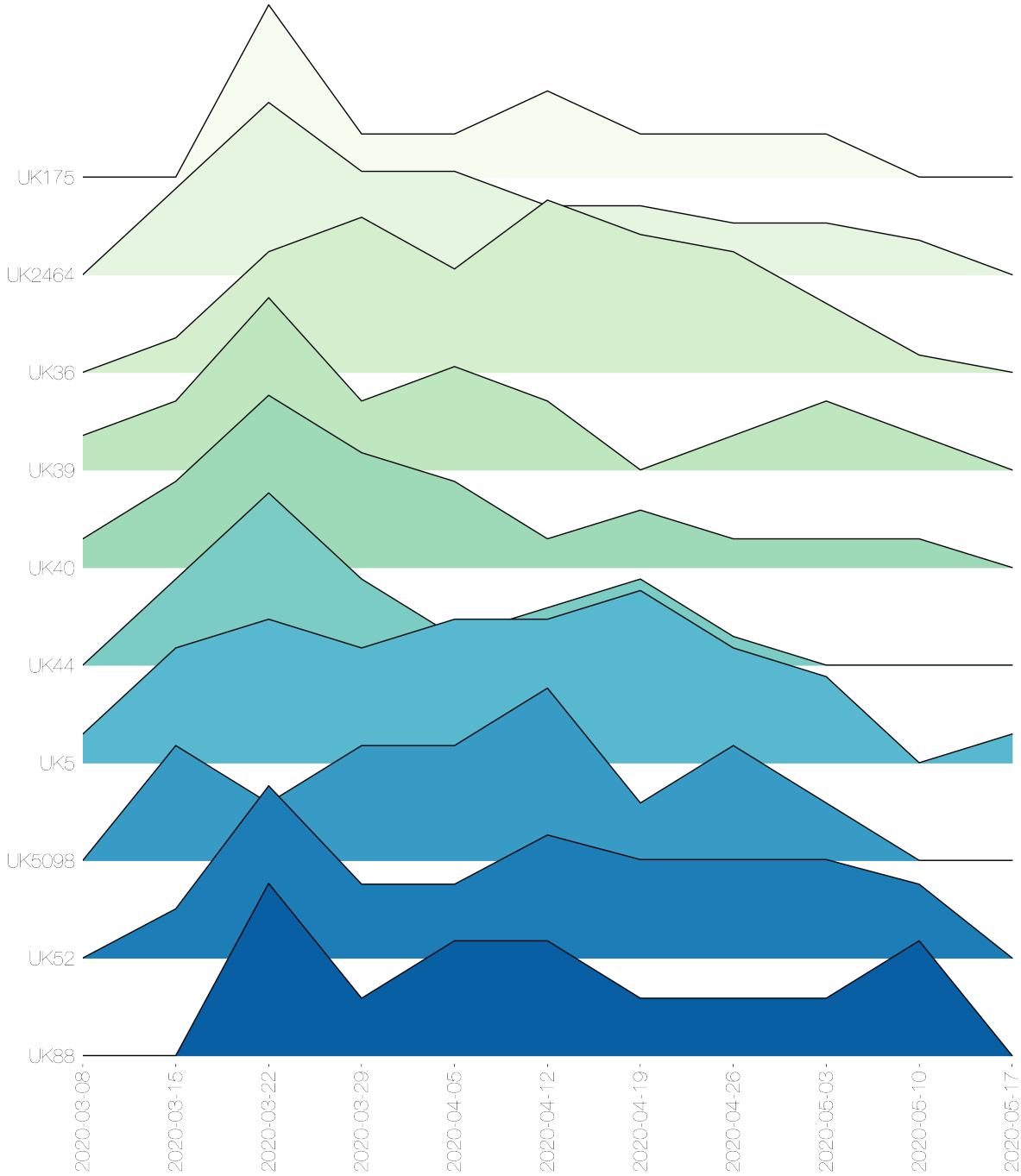

Figure 3: Lineages by number of adm2 regions present by epiweek

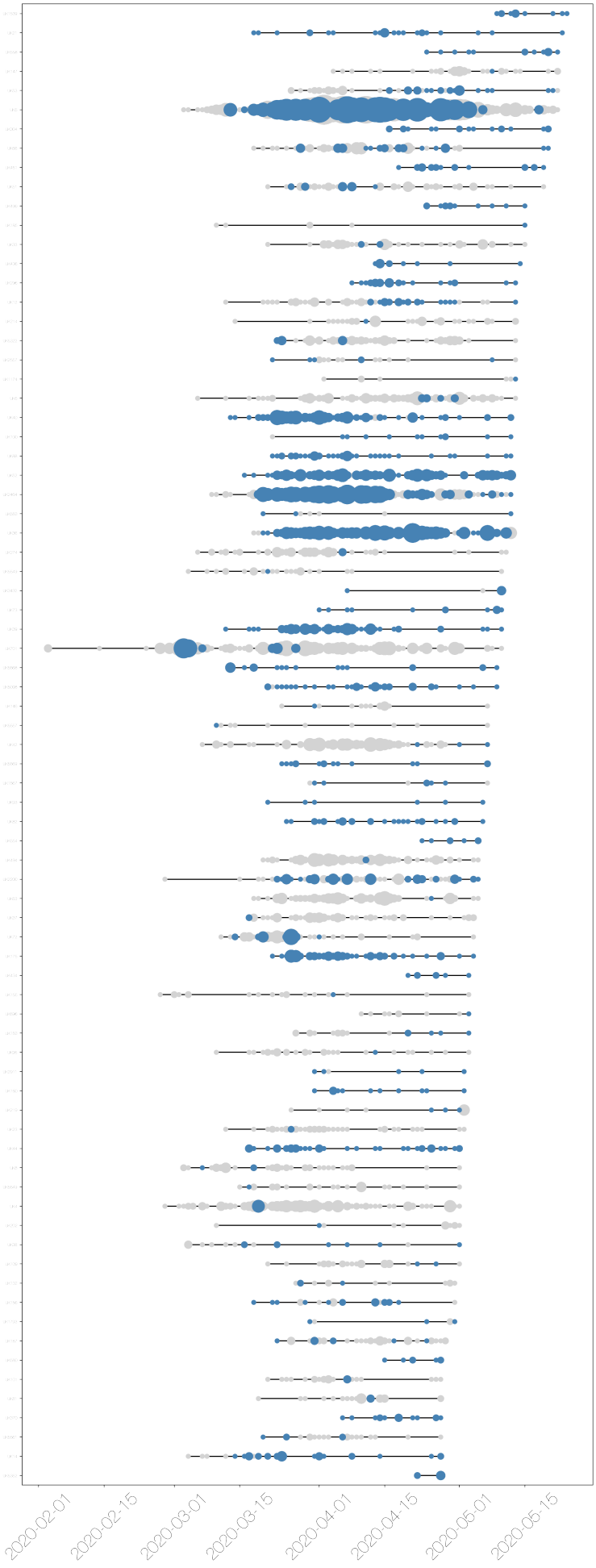

Figure 4: Timeline of lineages, sized by number of sequences from each country.

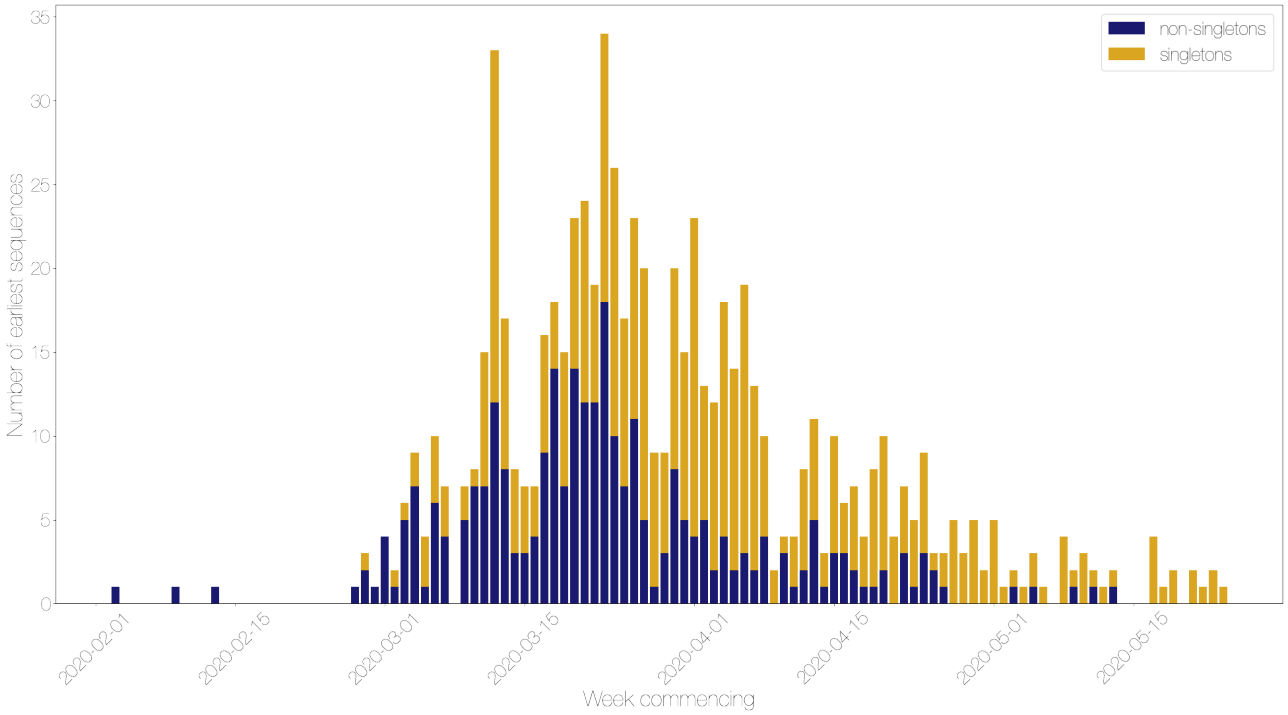

Figure 5: Lineage starts per week, split by singletons and non-singletons

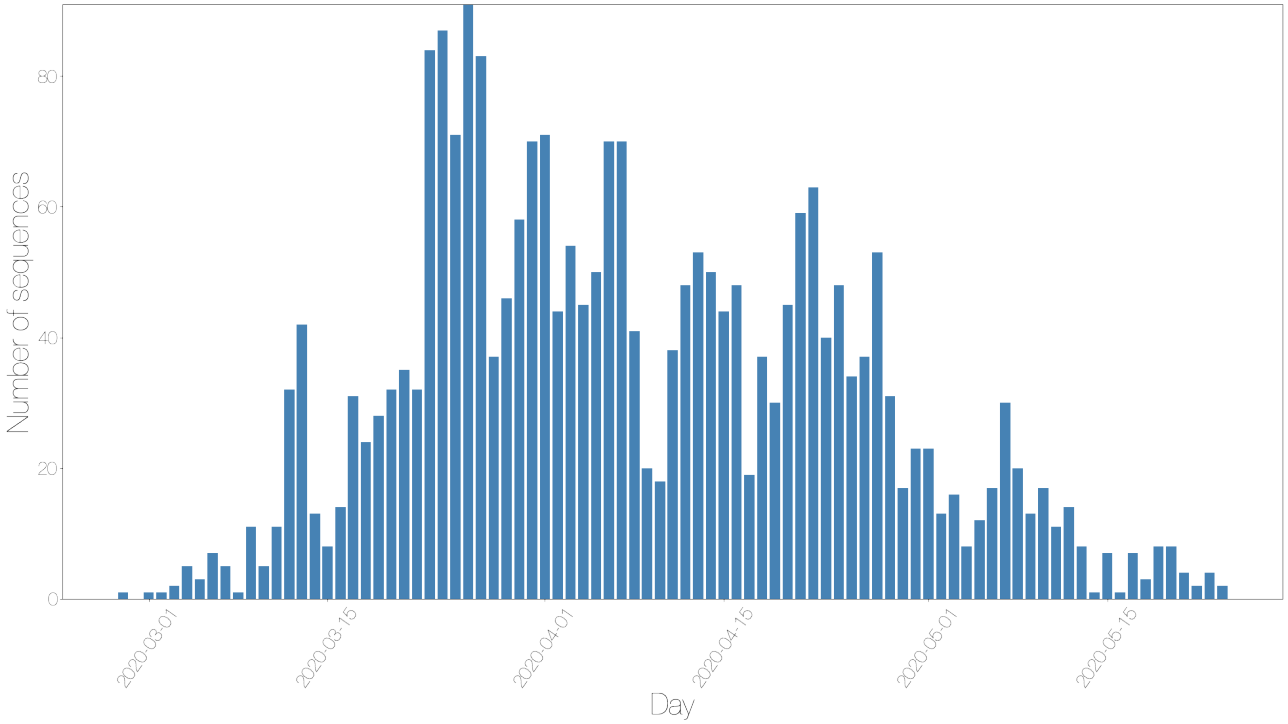

Figure 6: Sequences taken on each day by country

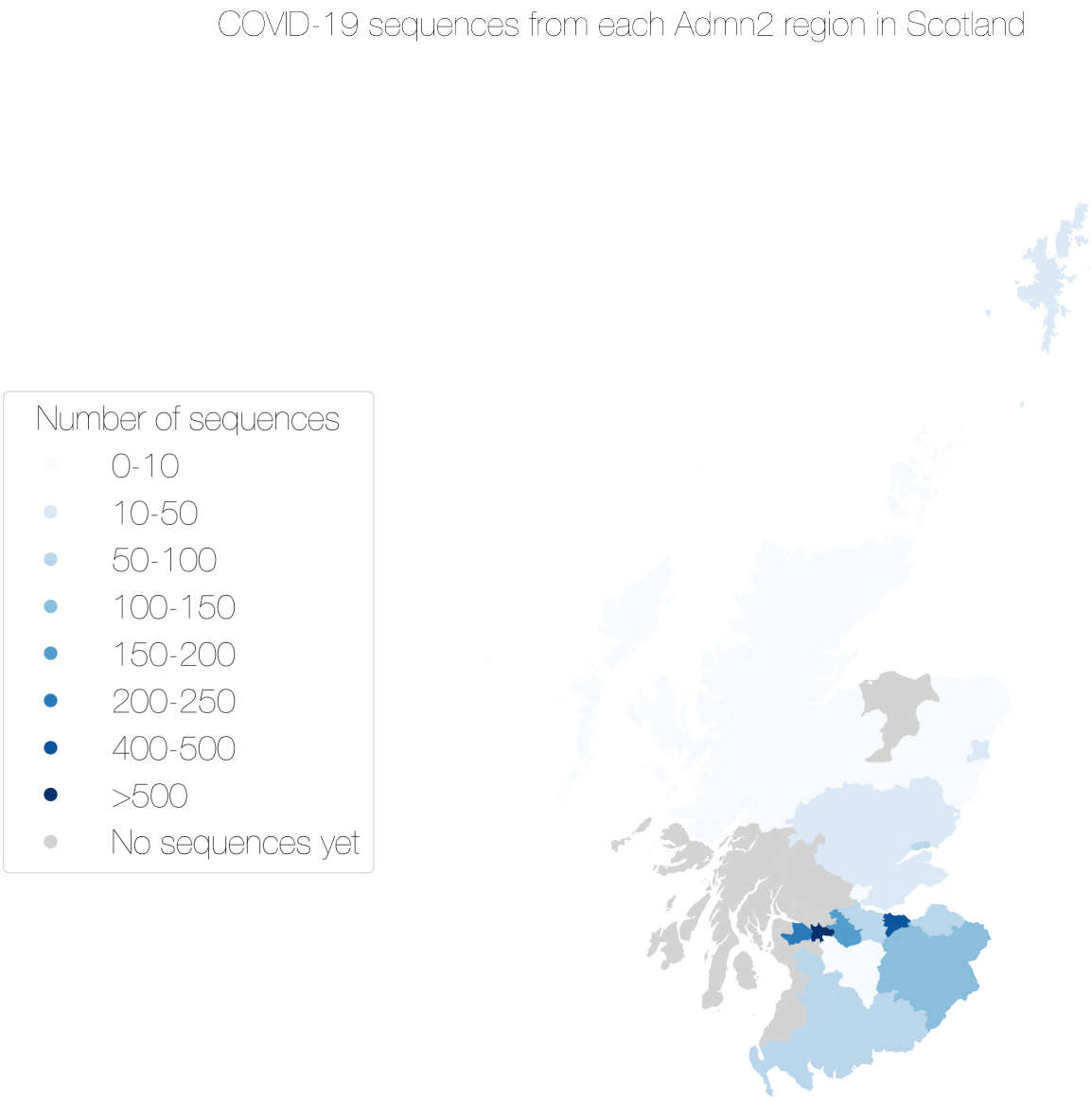

Figure 7: Map showing the number of sequences sampled by adm2 region

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
UK36	Mar-20, May-11	216	B.1	13	0.0153
UK52	Mar-16, May-12	175	B.1.p73	12	0.0268
UK40	Mar-13, May-12	126	B, B.16	12	0.0376
UK2464	Mar-19, May-12	92	B.1.p11	12	0.0133
UK39	Mar-12, May-10	80	A.2	14	0.0533
UK5	Mar-13, May-18	70	B.1.1.1	6	0.0334
UK175	Mar-22, May-04	67	B.1	20	0.0326
UK225	Mar-14, Apr-07	50	B.2	47	0.0112
UK88	Mar-22, May-12	40	B.1	12	0.1062
UK44	Mar-17, May-01	39	B	23	0.0502
UK5098	Mar-21, May-09	36	B.1.p73	15	0.0907
UK2200	Mar-23, May-05	32	B.1.5, B.1.5.6	19	0.0396
UK43	Mar-23, Apr-26	28	A.5	28	0.0574
UK82	Mar-25, May-06	27	B.1.1.p11, B.1.1	18	0.0864
UK296	Apr-08, May-13	25	B.1.5	11	0.1326
UK14	Mar-14, Apr-27	25	B	27	0.0645
UK21	Mar-18, May-23	24	B.1.40	1	2.8696
UK53	Apr-16, May-21	24	B.1.1.4	3	0.5816
UK5668	Mar-13, May-09	21	B.2	15	0.19
UK87	Mar-13, Apr-20	20	B.1.70	34	0.0588
UK150	Mar-21, Apr-22	19	B.1.1.p12	32	0.0556

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK461	Apr-18, May-19	19	B.1.5	5	0.3444
UK304	Apr-16, May-20	18	B.1.1.14	4	0.5
UK502	Mar-06, Mar-30	18	B.1.69	55	0.0257
UK12	Apr-12, May-13	18	B.1.p11	11	0.1044
UK66	Mar-28, May-20	16	B.1.1.8	4	0.1875
UK264	Mar-29, Apr-22	15	B.1.p11	32	0.0536
UK5669	Mar-24, May-07	15	B.2	17	0.1849
UK370	Apr-06, Apr-27	14	B.1.1.10	27	0.0598
UK156	Mar-18, Apr-18	14	B.1.71	36	0.0814
UK558	Apr-24, May-22	13	B.1.5	2	1.1667
UK160	Mar-31, May-02	12	B.1.1	22	0.1322
UK261	Mar-15, Apr-08	12	A.3	46	0.0474
UK73	Apr-01, May-10	12	B.1.p11	14	0.2532
UK499	Apr-24, May-15	12	B.1.5	9	0.2121
UK436	Apr-13, May-14	11	B.1.5	10	0.31
UK1539	May-09, May-24	11	B.1.5	0	active today
UK562	Mar-27, Apr-25	11	B.1	29	0.1
UK100	Apr-06, May-12	11	B.1.5	12	0.3864
UK414	Apr-05, Apr-22	10	B.1.5	32	0.059
UK137	Mar-10, Mar-31	10	B.1.1	54	0.0754
UK1548	Apr-13, Apr-24	9	B.1.5, B.1.5.5	30	0.0458
UK72	Mar-14, Apr-01	7	B.10	53	0.0355
UK699	Mar-23, Apr-24	7	B.1.5	30	0.1778
UK560	Apr-15, Apr-27	7	B.1.5	27	0.0741

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK434	Apr-20, May-03	7	B.1.5	21	0.1032
UK554	Apr-23, May-05	7	B.1.5	19	0.1053
UK133	Mar-22, Apr-25	7	B.1	29	0.1466
UK320	Mar-24, Apr-21	7	B.1.5, B.1	33	0.1414
UK701	Mar-03, Mar-27	7	B.1	58	0.023
UK282	Mar-23, Apr-22	7	B.1.1	32	0.1562
UK271	Apr-15, Apr-26	7	B.1	28	0.1224
UK58	Mar-12, Apr-07	7	B.1	47	0.0519
UK38	Mar-16, May-01	7	B.2.1	23	0.1576
UK51	Mar-26, Apr-13	6	B.1.36	41	0.2107
UK870	Mar-18, Mar-25	6	B.1.5	60	0.0233
UK198	Mar-18, Apr-15	6	B.1.5, A	39	0.0798
UK1667	Mar-31, Apr-28	6	B.1.p9	26	0.2794
UK93	Mar-21, May-06	6	B.1.1	18	0.5111
UK5352	Apr-22, Apr-27	6	B.1.1.14	27	0.037
UK555	Apr-13, Apr-25	6	B.1.5	29	0.0828

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

Week commencing	UK36	UK52	UK40	UK2464	UK39	UK5	UK175	UK88	UK44	UK5098
2020-03-08	0	0	1	0	1	1	0	0	0	0
2020-03-15	2	2	3	5	2	4	0	0	3	2
2020-03-22	7	7	6	10	5	5	4	3	6	1
2020-03-29	9	3	4	6	2	4	1	1	3	2
2020-04-05	6	3	3	6	3	5	1	2	1	2
2020-04-12	10	5	1	4	2	5	2	2	2	3
2020-04-19	8	4	2	4	0	6	1	1	3	1
2020-04-26	7	4	1	3	1	4	1	1	1	2
2020-05-03	4	4	1	3	2	3	1	1	0	1
2020-05-10	1	3	1	2	1	0	0	2	0	0
2020-05-17	0	0	0	0	0	1	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
2020-02-03	0	1	1
2020-02-09	0	1	1
2020-02-13	0	1	1
2020-02-27	0	1	1
2020-02-28	1	2	3
2020-02-29	0	1	1
2020-03-01	0	4	4
2020-03-02	1	1	2
2020-03-03	1	5	6
2020-03-04	2	7	9
2020-03-05	3	1	4
2020-03-06	4	6	10
2020-03-07	3	4	7
2020-03-09	2	5	7
2020-03-10	1	7	8
2020-03-11	8	7	15
2020-03-12	21	12	33
2020-03-13	9	8	17
2020-03-14	5	3	8
2020-03-15	4	3	7
2020-03-16	3	4	7
2020-03-17	7	9	16
2020-03-18	4	14	18
2020-03-19	8	7	15
2020-03-20	9	14	23
2020-03-21	12	12	24
2020-03-22	7	12	19
2020-03-23	16	18	34
2020-03-24	16	10	26
2020-03-25	10	7	17
2020-03-26	12	11	23
2020-03-27	15	5	20
2020-03-28	8	1	9
2020-03-29	6	3	9
2020-03-30	12	8	20
2020-03-31	10	5	15
2020-04-01	19	4	23
2020-04-02	8	5	13
2020-04-03	10	2	12
2020-04-04	14	4	18
2020-04-05	12	2	14
2020-04-06	16	3	19
2020-04-07	11	2	13
2020-04-08	6	4	10
2020-04-09	2	0	2
2020-04-10	1	3	4

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

Table S5 Raw data for figure seven showing the number of sequences taken over time.

Day	Scotland
2020-02-28	1
2020-03-01	1
2020-03-02	1
2020-03-03	2
2020-03-04	5
2020-03-05	3
2020-03-06	7
2020-03-07	5
2020-03-08	1
2020-03-09	11
2020-03-10	5
2020-03-11	11
2020-03-12	32
2020-03-13	42
2020-03-14	13
2020-03-15	8
2020-03-16	14
2020-03-17	31
2020-03-18	24
2020-03-19	28
2020-03-20	32
2020-03-21	35
2020-03-22	32
2020-03-23	84
2020-03-24	87
2020-03-25	71
2020-03-26	91
2020-03-27	83
2020-03-28	37
2020-03-29	46
2020-03-30	58
2020-03-31	70
2020-04-01	71
2020-04-02	44
2020-04-03	54
2020-04-04	45
2020-04-05	50
2020-04-06	70
2020-04-07	70
2020-04-08	41
2020-04-09	20
2020-04-10	18
2020-04-11	38
2020-04-12	48
2020-04-13	53
2020-04-14	50
2020-04-15	44

Day	Scotland
2020-04-16	48
2020-04-17	19
2020-04-18	37
2020-04-19	30
2020-04-20	45
2020-04-21	59
2020-04-22	63
2020-04-23	40
2020-04-24	48
2020-04-25	34
2020-04-26	37
2020-04-27	53
2020-04-28	31
2020-04-29	17
2020-04-30	23
2020-05-01	23
2020-05-02	13
2020-05-03	16
2020-05-04	8
2020-05-05	12
2020-05-06	17
2020-05-07	30
2020-05-08	20
2020-05-09	13
2020-05-10	17
2020-05-11	11
2020-05-12	14
2020-05-13	8
2020-05-14	1
2020-05-15	7
2020-05-16	1
2020-05-17	7
2020-05-18	3
2020-05-19	8
2020-05-20	8
2020-05-21	4
2020-05-22	2
2020-05-23	4
2020-05-24	2

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

Admin2	Country	Number of sequences	Sequence group
ABERDEEN	Scotland	22	10-50
ABERDEENSHIRE	Scotland	5	1-10
ANGUS	Scotland	13	10-50
ARGYLL AND BUTE	Scotland	0	0
CLACKMANNANSHIRE	Scotland	2	1-10
DUMFRIES AND GALLOWAY	Scotland	54	50-100
DUNDEE	Scotland	93	50-100
EAST AYRSHIRE	Scotland	75	50-100
EAST DUNBARTONSHIRE	Scotland	0	0
EAST LOTHIAN	Scotland	54	50-100
EAST RENFREWSHIRE	Scotland	0	0
EDINBURGH	Scotland	412	400-500
EILEAN SIAR	Scotland	2	1-10
FALKIRK	Scotland	92	50-100
FIFE	Scotland	42	10-50
GLASGOW	Scotland	791	>500
HIGHLAND	Scotland	9	1-10
INVERCLYDE	Scotland	0	0
MIDLOTHIAN	Scotland	127	100-150
MORAY	Scotland	0	0
NORTH AYRSHIRE	Scotland	0	0
NORTH LANARKSHIRE	Scotland	158	150-200
ORKNEY ISLANDS	Scotland	1	1-10
PERTHSHIRE AND KINROSS	Scotland	48	10-50
RENFREWSHIRE	Scotland	209	200-250
SCOTTISH BORDERS	Scotland	128	100-150
SHETLAND ISLANDS	Scotland	14	10-50
SOUTH AYRSHIRE	Scotland	0	0
SOUTH LANARKSHIRE	Scotland	4	1-10
STIRLING	Scotland	0	0
WEST DUNBARTONSHIRE	Scotland	0	0
WEST LOTHIAN	Scotland	92	50-100