Lineages report for Scotland

This report gives summaries of lineages sampled in Scotland for week 2020-06-05. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-05-26. The analysis (eg time since last sample) is therefore undertaken from this date. 2693 sequences from Scotland have been included in this analysis. 794 lineages have been recorded, 614 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 795 and the maximum is 997

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

758 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 36 that remain: 17 are pending extinction, ie last seen three weeks ago. 9 lineages have gone quiet, ie haven't been seen this week. 4 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 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	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK5098	Mar-16, May-12	240	B.1.p73	14	0.017
UK36	Mar-20, May-21	237	B.1	5	0.0387
UK40	Mar-13, May-12	135	B, B.16	14	0.0302
UK2464	Mar-19, May-25	103	B.1.p11	1	0.1774
UK39	Mar-12, May-10	86	A.2	16	0.0434
UK5	Mar-13, May-26	75	B.1.1.1	0	active today

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK225	Mar-14, Apr-10	51	B.2	46	0.0105
UK88	Mar-22, May-12	47	B.1	14	0.0775
UK44	Mar-17, May-01	46	B	25	0.0391
UK82	Mar-25, May-13	39	B.1.1, B.1.1.p11	13	0.0966

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.

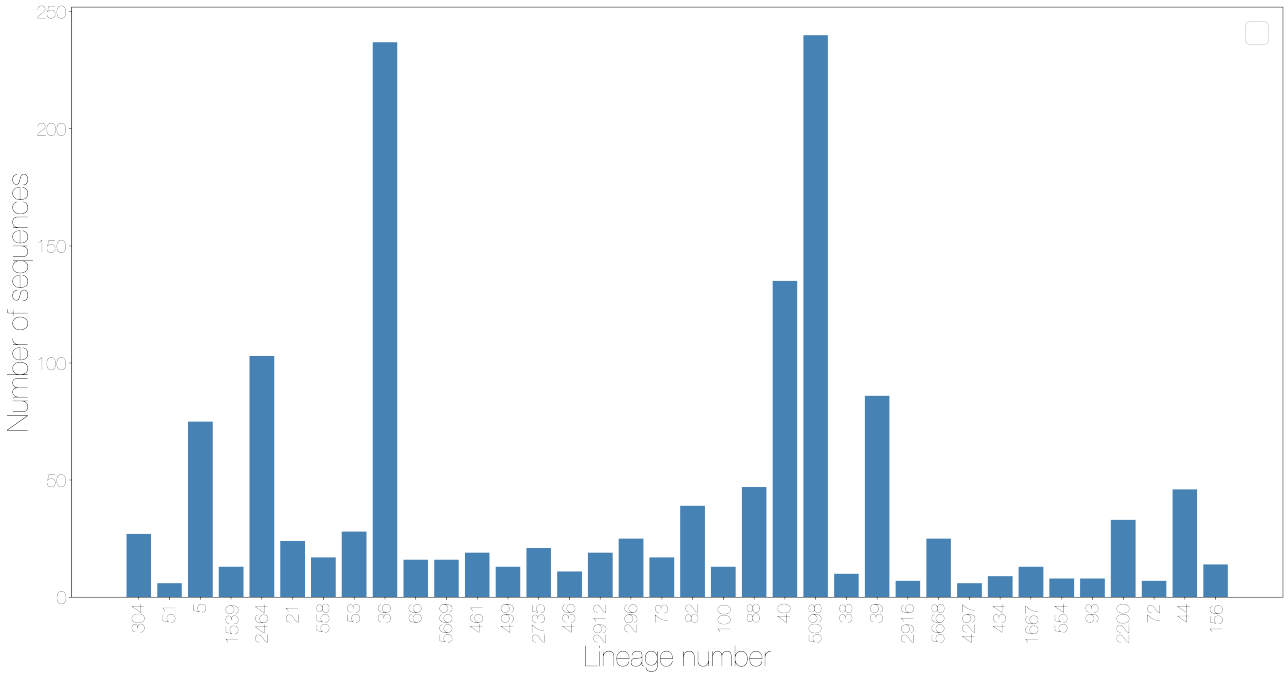

Figure 1: Number of sequences sampled in a lineage by country

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.

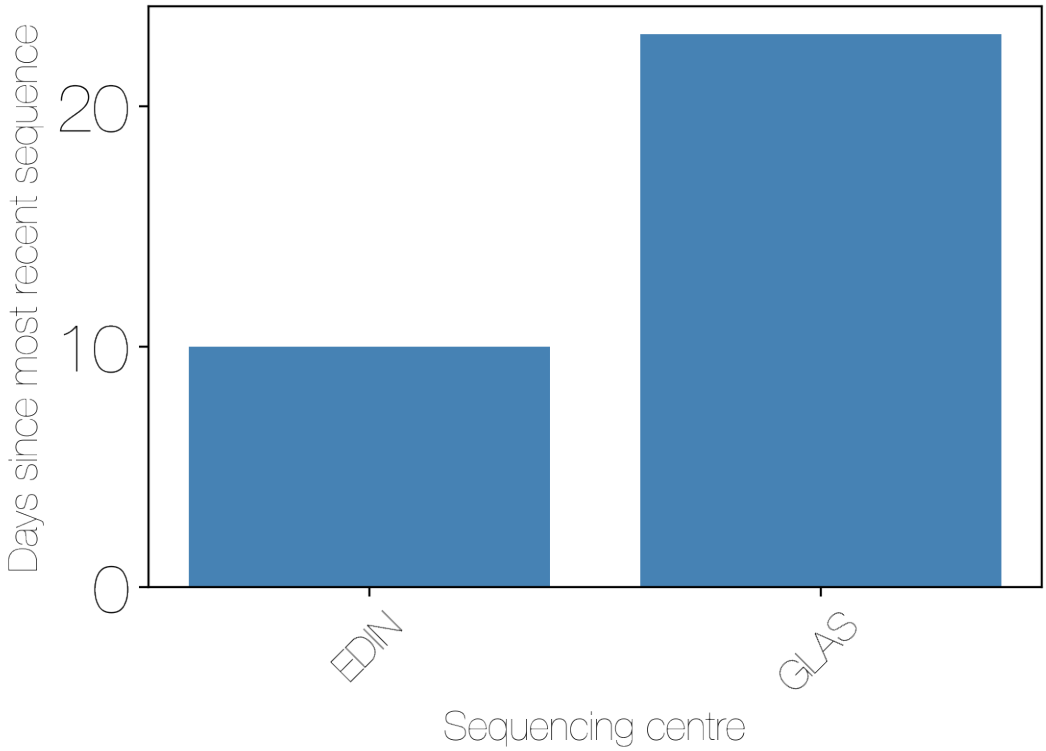

Figure 2: Lag since the most recent sequence from each sequencing centre to most current date

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.

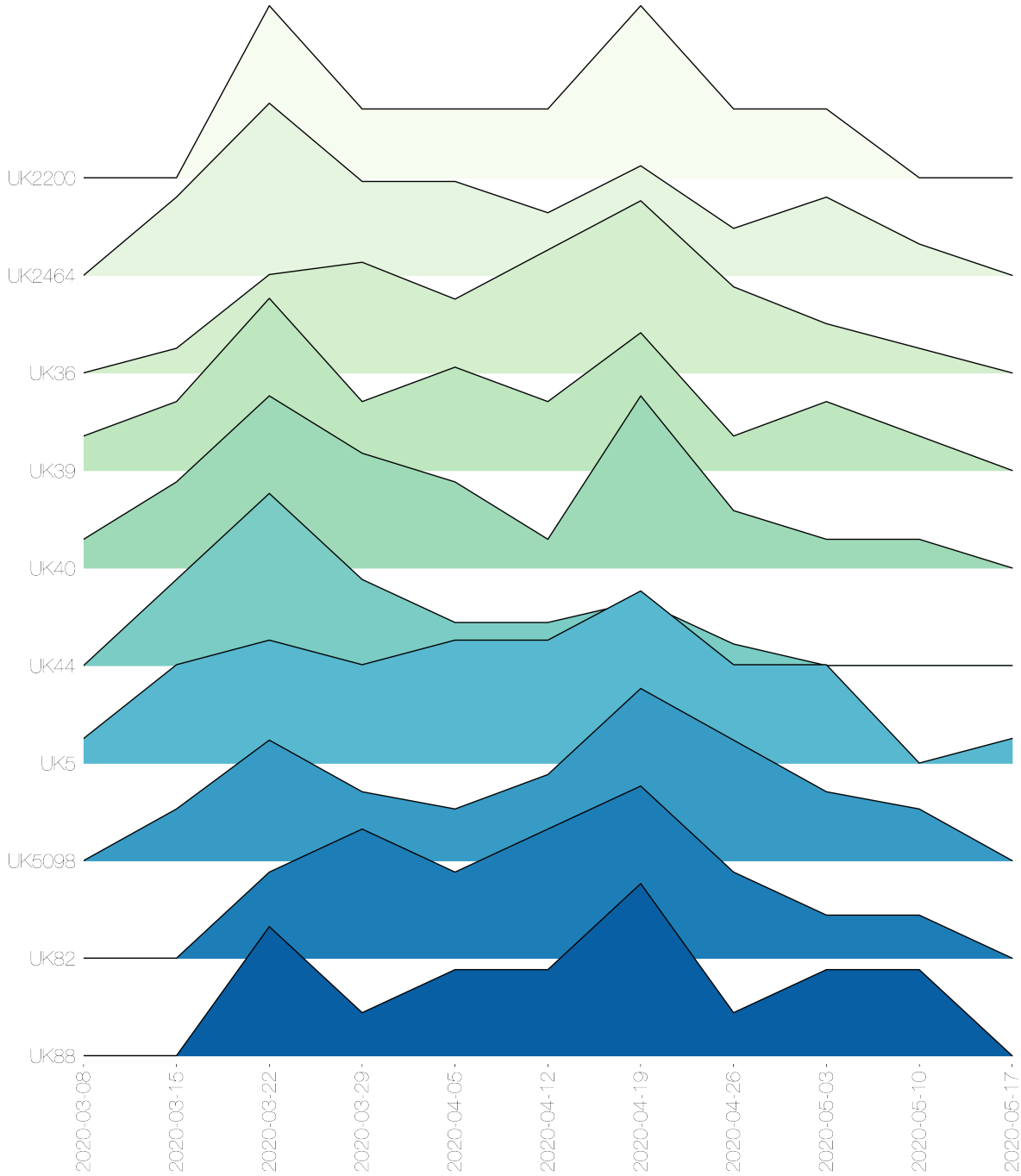

Figure 3: Lineages by number of adm2 regions present by epiweek

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.

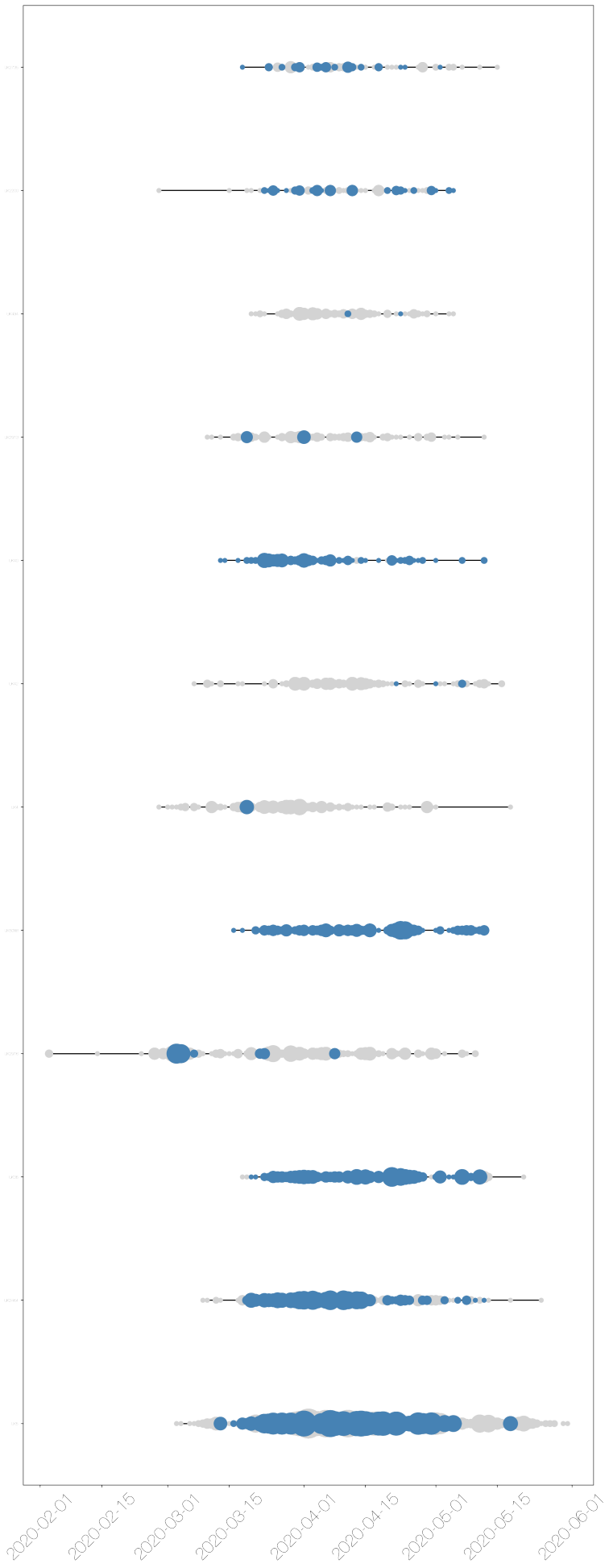

Figure 4: Timeline of lineages, sized by number of sequences from each country.

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

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

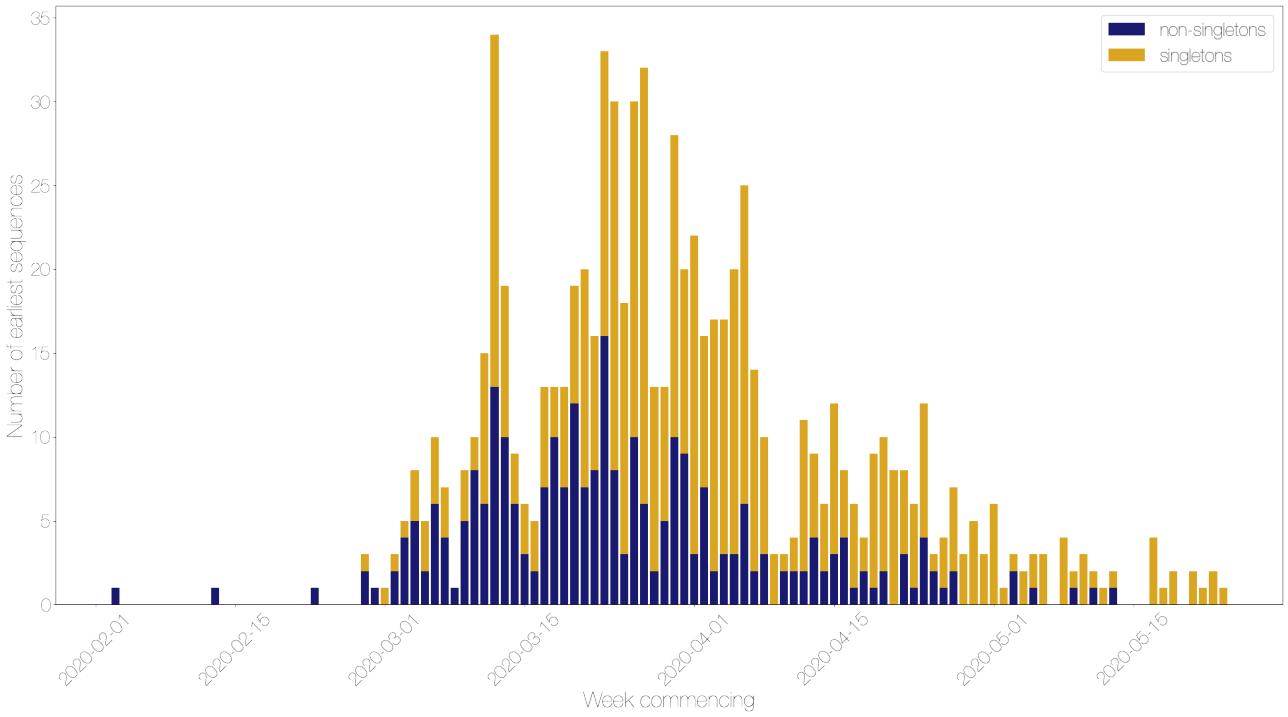

Figure 5: Lineage starts per week, split by singletons and non-singletons

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.

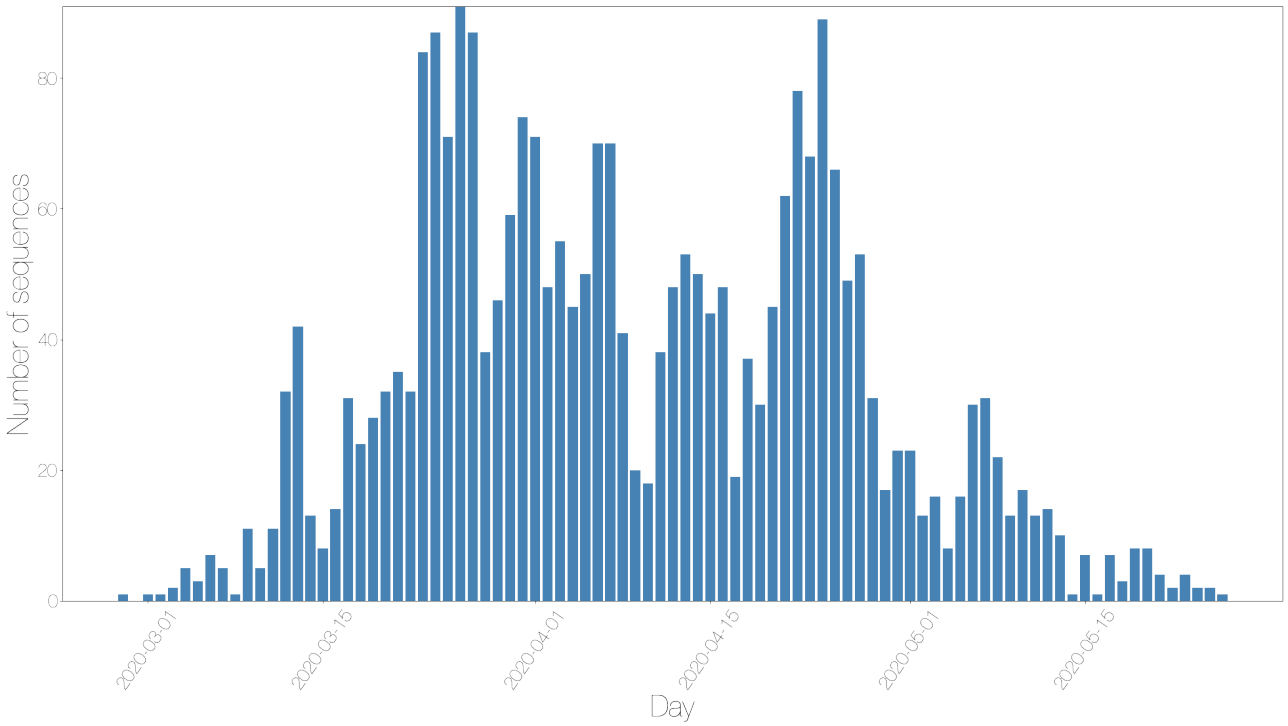

Figure 6: Sequences taken on each day by country

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.

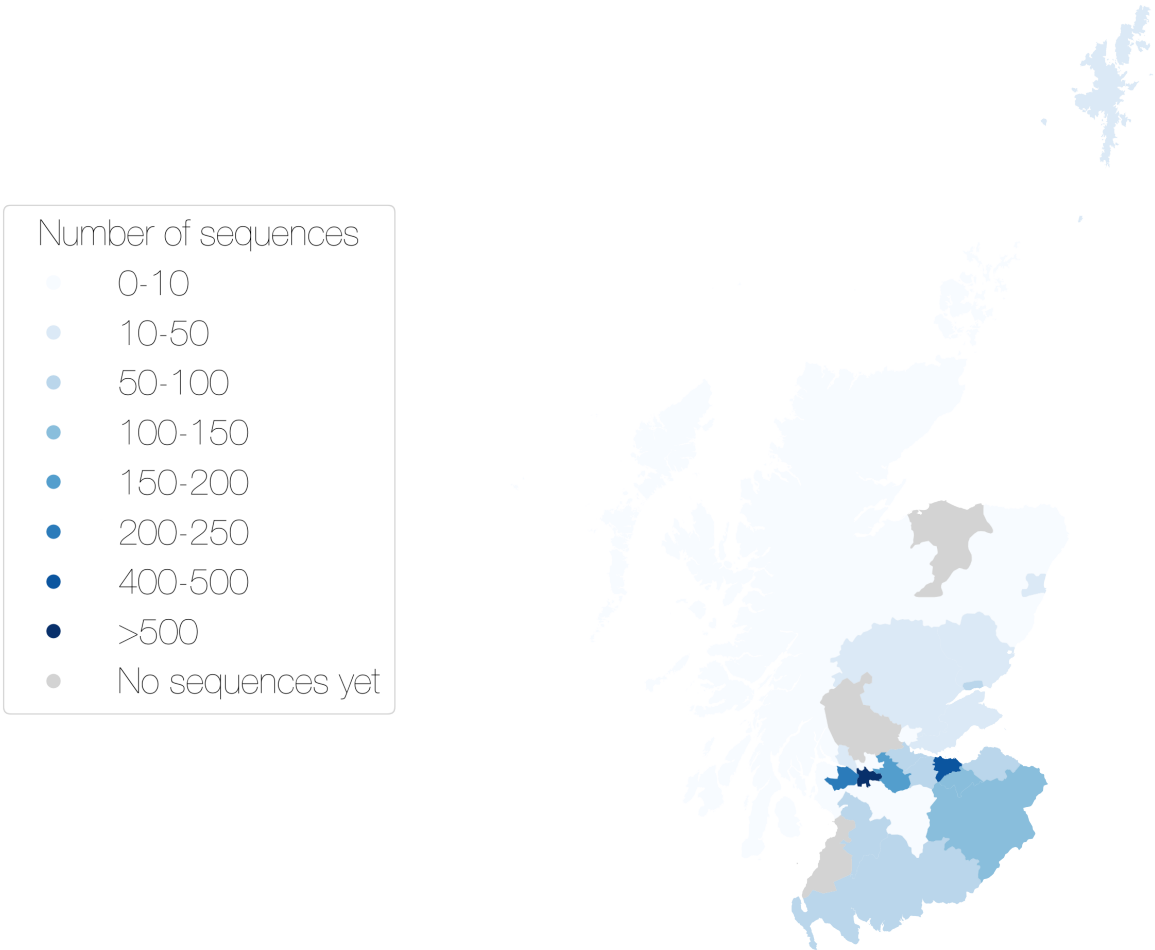

Figure 7: Map showing the number of sequences sampled by adm2 region

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.

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
UK5098	Mar-16, May-12	240	B.1.p73	14	0.017
UK36	Mar-20, May-21	237	B.1	5	0.0387
UK40	Mar-13, May-12	135	B, B.16	14	0.0302
UK2464	Mar-19, May-25	103	B.1.p11	1	0.1774
UK39	Mar-12, May-10	86	A.2	16	0.0434
UK5	Mar-13, May-26	75	B.1.1.1	0	active today
UK225	Mar-14, Apr-10	51	B.2	46	0.0105
UK88	Mar-22, May-12	47	B.1	14	0.0775
UK44	Mar-17, May-01	46	B	25	0.0391
UK82	Mar-25, May-13	39	B.1.1, B.1.1.p11	13	0.0966
UK2200	Mar-23, May-05	33	B.1.5, B.1.5.6	21	0.0351
UK43	Mar-23, Apr-26	28	A.5	30	0.0536
UK53	Apr-16, May-22	28	B.1.1.4	4	0.2639
UK304	Apr-16, May-26	27	B.1.1.14	0	active today
UK14	Mar-14, Apr-27	26	B	29	0.0587
UK5668	Mar-13, May-09	25	B.2	17	0.1341
UK296	Apr-08, May-13	25	B.1.5	13	0.1122
UK21	Mar-18, May-23	24	B.1.40	3	0.9565
UK2735	Mar-18, May-15	21	B.1.1	11	0.0586
UK2912	Apr-12, May-13	19	B.1.p11	13	0.1498
UK461	Apr-18, May-19	19	B.1.5	7	0.2331

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK87	Mar-17, Apr-24	19	B.1.70	32	0.066
UK502	Mar-06, Mar-30	18	B.1.69	57	0.0248
UK558	Apr-24, May-22	17	B.1.5	4	0.4375
UK73	Apr-01, May-13	17	B.1.p11	13	0.2019
UK66	Mar-28, May-20	16	B.1.1.8	6	0.125
UK150	Mar-21, Apr-22	16	B.1.1.p12	34	0.0627
UK5669	Mar-24, May-19	16	B.2	7	0.5
UK264	Mar-29, Apr-22	15	B.1.p11	34	0.0504
UK156	Mar-18, Apr-30	14	B.1.71	26	0.0752
UK370	Apr-06, Apr-27	14	B.1.1.10	29	0.0557
UK1539	May-09, May-25	13	B.1.5	1	1.3333
UK1667	Mar-31, May-07	13	B.1.p9	19	0.1333
UK100	Apr-06, May-12	13	B.1.5	14	0.2143
UK499	Apr-24, May-15	13	B.1.5	11	0.1591
UK261	Mar-15, Apr-08	12	A.3	48	0.0455
UK436	Apr-13, May-14	11	B.1.5	12	0.3409
UK58	Mar-12, Apr-24	11	B.1	32	0.084
UK562	Mar-27, Apr-25	10	B.1	31	0.1039
UK414	Apr-05, Apr-22	10	B.1.5	34	0.0556
UK38	Mar-16, May-11	10	B.2.1	15	0.2386
UK699	Mar-23, Apr-24	9	B.1.5	32	0.125
UK1548	Apr-13, Apr-24	9	B.1.5, B.1.5.5	32	0.043
UK434	Apr-20, May-07	9	B.1.5	19	0.1118
UK93	Mar-21, May-06	8	B.1.1	20	0.3286

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK554	Apr-23, May-06	8	B.1.5	20	0.0929
UK271	Apr-15, Apr-26	7	B.1	30	0.1143
UK133	Mar-22, Apr-25	7	B.1	31	0.1371
UK560	Apr-15, Apr-27	7	B.1.5	29	0.069
UK72	Mar-14, May-04	7	B.10	22	0.0303
UK2916	Mar-03, May-10	7	B.1	16	0.0189
UK282	Mar-23, Apr-22	7	B.1.1	34	0.1261
UK187	Mar-23, Apr-28	7	B.1	28	0.0299
UK870	Mar-18, Mar-25	6	B.1.5	62	0.0226
UK51	Mar-26, May-26	6	B.1.36	0	active today
UK198	Mar-18, Apr-15	6	B.1.5, A	41	0.0759
UK555	Apr-13, Apr-25	6	B.1.5	31	0.0774
UK137	Mar-13, Apr-11	6	B.1.1	45	0.0768
UK4297	Mar-26, May-08	6	B.1.1	18	0.4778

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	EDIN	10
1	GLAS	23

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

Week commencing	UK5098	UK36	UK40	UK2464	UK39	UK5	UK88	UK44	UK82	UK2200
2020-03-08	0	0	1	0	1	1	0	0	0	0
2020-03-15	3	2	3	5	2	4	0	4	0	0
2020-03-22	7	8	6	11	5	5	3	8	2	5
2020-03-29	4	9	4	6	2	4	1	4	3	2
2020-04-05	3	6	3	6	3	5	2	2	2	2
2020-04-12	5	10	1	4	2	5	2	2	3	2
2020-04-19	10	14	6	7	4	7	4	3	4	5
2020-04-26	7	7	2	3	1	4	1	1	2	2
2020-05-03	4	4	1	5	2	4	2	0	1	2
2020-05-10	3	2	1	2	1	0	2	0	1	0
2020-05-17	0	0	0	0	0	1	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
2020-02-03	0	1	1
2020-02-13	0	1	1
2020-02-23	0	1	1
2020-02-28	1	2	3
2020-02-29	0	1	1
2020-03-01	1	0	1
2020-03-02	1	2	3
2020-03-03	1	4	5
2020-03-04	3	5	8
2020-03-05	3	2	5
2020-03-06	4	6	10
2020-03-07	3	4	7
2020-03-08	0	1	1
2020-03-09	3	5	8
2020-03-10	2	8	10
2020-03-11	9	6	15
2020-03-12	21	13	34
2020-03-13	9	10	19
2020-03-14	3	6	9
2020-03-15	3	3	6
2020-03-16	3	2	5
2020-03-17	6	7	13
2020-03-18	3	10	13
2020-03-19	6	7	13
2020-03-20	7	12	19
2020-03-21	13	7	20
2020-03-22	8	8	16
2020-03-23	17	16	33
2020-03-24	22	8	30
2020-03-25	15	3	18
2020-03-26	20	10	30
2020-03-27	26	6	32
2020-03-28	11	2	13
2020-03-29	8	5	13
2020-03-30	18	10	28
2020-03-31	11	9	20
2020-04-01	19	3	22
2020-04-02	9	7	16
2020-04-03	15	2	17
2020-04-04	14	3	17
2020-04-05	17	3	20
2020-04-06	19	6	25
2020-04-07	12	2	14
2020-04-08	7	3	10
2020-04-09	3	0	3
2020-04-10	1	2	3

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-04-11	2	2	4
2020-04-12	9	2	11
2020-04-13	5	4	9
2020-04-14	4	2	6
2020-04-15	9	3	12
2020-04-16	4	4	8
2020-04-17	5	1	6
2020-04-18	2	2	4
2020-04-19	8	1	9
2020-04-20	8	2	10
2020-04-21	8	0	8
2020-04-22	5	3	8
2020-04-23	5	1	6
2020-04-24	8	4	12
2020-04-25	1	2	3
2020-04-26	3	1	4
2020-04-27	5	2	7
2020-04-28	3	0	3
2020-04-29	5	0	5
2020-04-30	3	0	3
2020-05-01	6	0	6
2020-05-02	1	0	1
2020-05-03	1	2	3
2020-05-04	2	0	2
2020-05-05	2	1	3
2020-05-06	3	0	3
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 S6 Raw data for figure six 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	87
2020-03-28	38
2020-03-29	46
2020-03-30	59
2020-03-31	74
2020-04-01	71
2020-04-02	48
2020-04-03	55
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	62
2020-04-22	78
2020-04-23	68
2020-04-24	89
2020-04-25	66
2020-04-26	49
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	16
2020-05-06	30
2020-05-07	31
2020-05-08	22
2020-05-09	13
2020-05-10	17
2020-05-11	13
2020-05-12	14
2020-05-13	10
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
2020-05-25	2
2020-05-26	1

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
ABERDEEN	Scotland	22	10-50
ABERDEENSHIRE	Scotland	5	1-10
ANGUS	Scotland	13	10-50
ARGYLL AND BUTE	Scotland	2	1-10
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	6	1-10
EAST LOTHIAN	Scotland	54	50-100
EAST RENFREWSHIRE	Scotland	5	1-10
EDINBURGH	Scotland	426	400-500
EILEAN SIAR	Scotland	2	1-10
FALKIRK	Scotland	92	50-100
FIFE	Scotland	43	10-50
GLASGOW	Scotland	791	>500
HIGHLAND	Scotland	9	1-10
INVERCLYDE	Scotland	3	1-10
MIDLOTHIAN	Scotland	131	100-150
MORAY	Scotland	0	0
NORTH AYRSHIRE	Scotland	2	1-10
NORTH LANARKSHIRE	Scotland	163	150-200
ORKNEY ISLANDS	Scotland	1	1-10
PERTHSHIRE AND KINROSS	Scotland	48	10-50
RENFREWSHIRE	Scotland	227	200-250
SCOTTISH BORDERS	Scotland	132	100-150
SHETLAND ISLANDS	Scotland	14	10-50
SOUTH AYRSHIRE	Scotland	0	0
SOUTH LANARKSHIRE	Scotland	9	1-10
STIRLING	Scotland	0	0
WEST DUNBARTONSHIRE	Scotland	10	10-50
WEST LOTHIAN	Scotland	95	50-100