

# Lineages report for Scotland

This report gives summaries of lineages sampled in Scotland for week 2020-07-03. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-06-22. The analysis (eg time since last sample) is therefore undertaken from this date. 4151 sequences from Scotland have been included in this analysis. 255 lineages have been recorded, 136 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 221 and the maximum is 1384

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

224 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 31 that remain: 18 are pending extinction, ie last seen three weeks ago. 4 lineages have gone quiet, ie haven't been seen this week. 5 lineages have reactivated. 4 lineages have been continuously circulating.

The following table contains information about the ten largest 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 expected 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-19, Jun-06	536	B.1	16	0.0083
UK5	Feb-28, Jun-21	435	B.1.1.13, B.1.1, B.1.1.p11, B.1.1.4, B.1.1.p12, B.1.1.1, B.1.1.14	1	0.0152
UK5098	Mar-01, Jun-05	433	B.1, B.1.p73	17	0.0129
UK109	Mar-12, Jun-12	244	B.1.5.5, B.1.5	10	0.0232
UK199	Mar-05, Jun-22	227	B.1, B.1.5, B.1.p73	0	active today
UK40	Mar-13, Jun-08	215	B.16, B	14	0.028
UK42	Mar-04, Jun-21	205	B.1.p11, B.1, B.1.p73, B.1.71, B.1.5	1	0.0863
UK2464	Mar-19, Jun-18	180	B.1.p11	4	0.0455
UK5676	Mar-12, May-27	150	B.2	26	0.0061
UK39	Mar-12, May-29	145	A.2	24	0.0221

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.

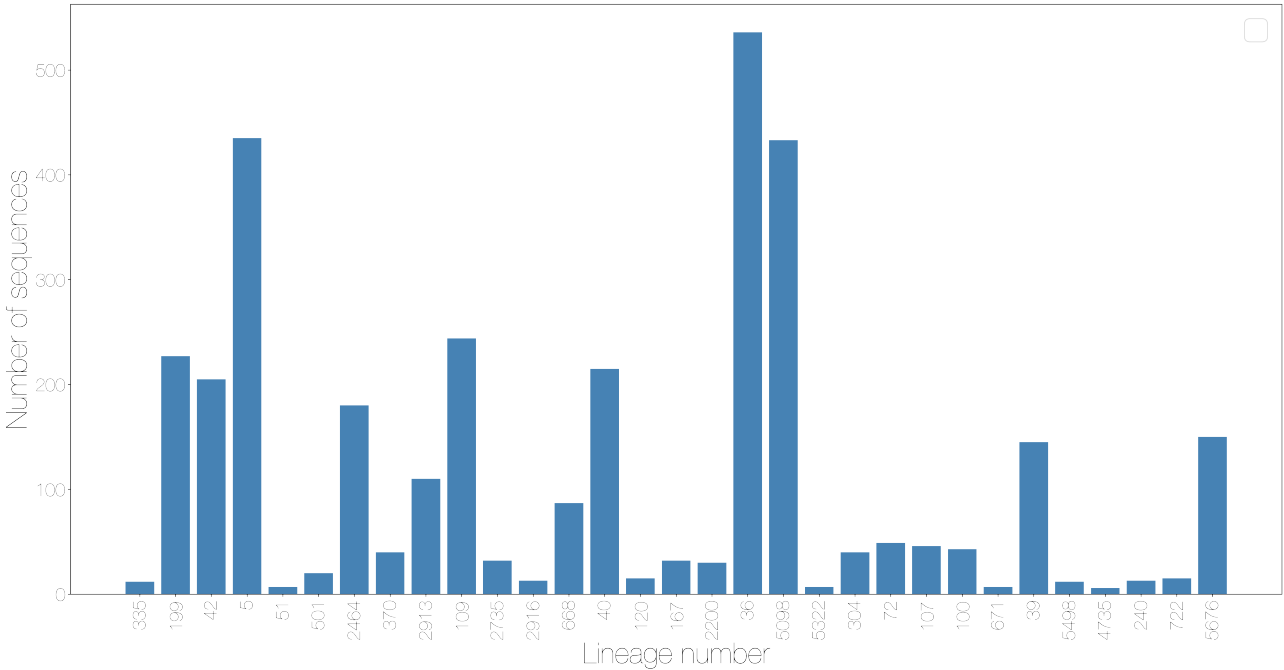


Figure 1: Number of sequences sampled in a lineage by country

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.

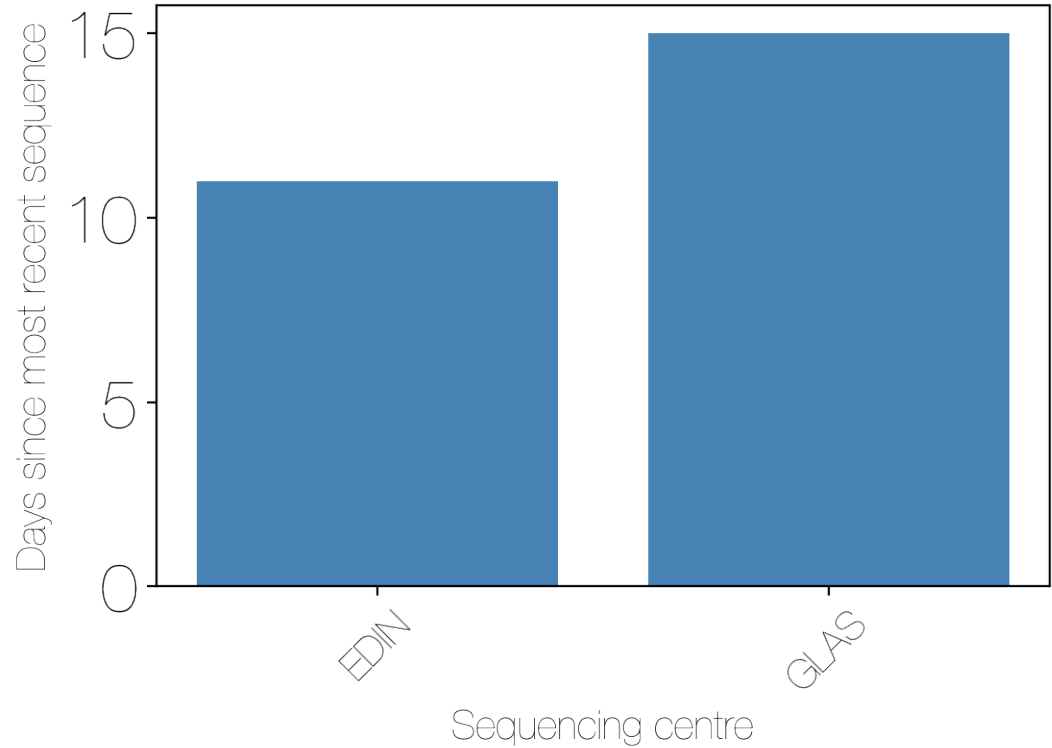


Figure 2: Lag since the most recent sequence from each sequencing centre to most current date

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.

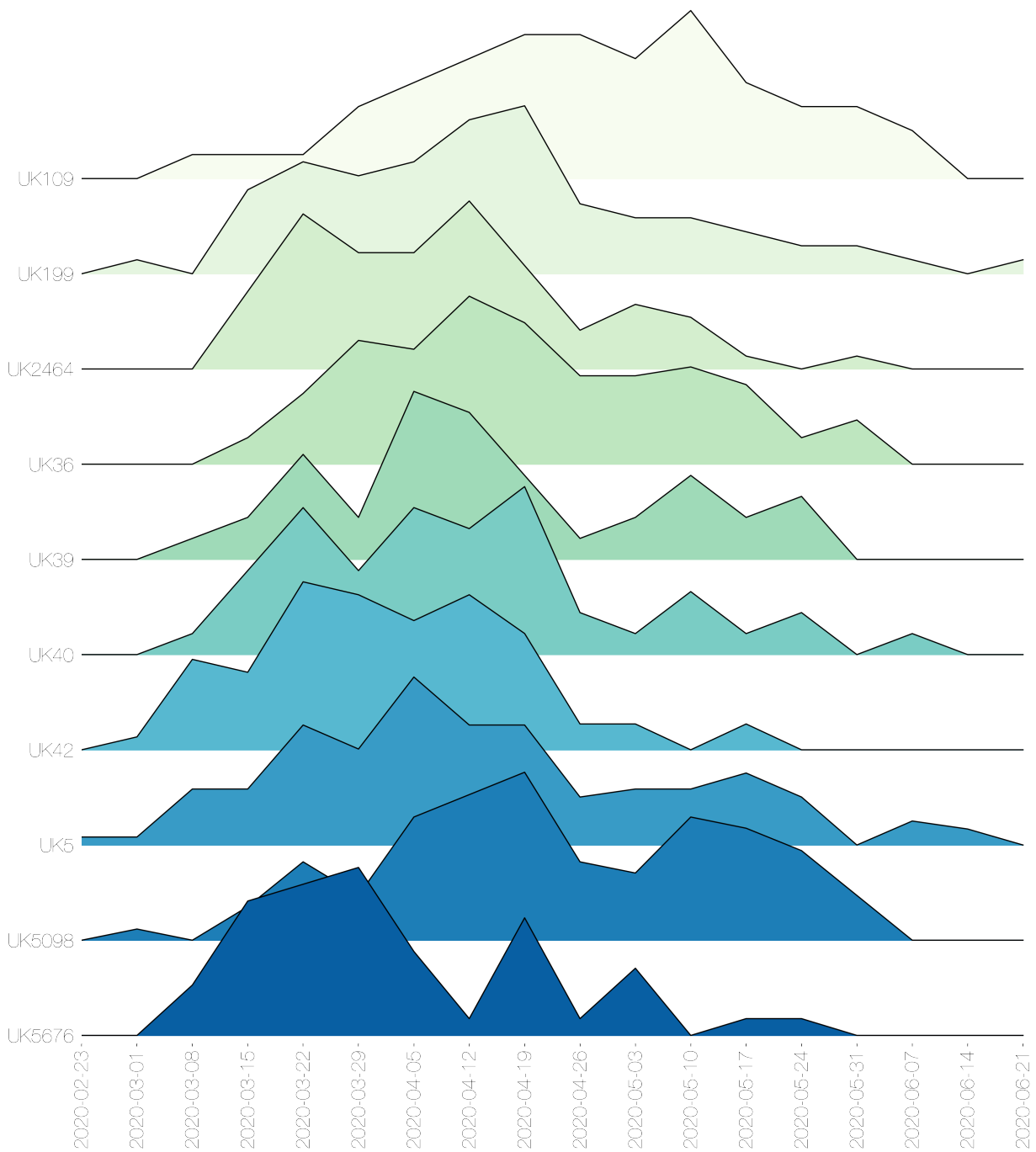


Figure 3: Lineages by number of adm2 regions present by epiweek

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.

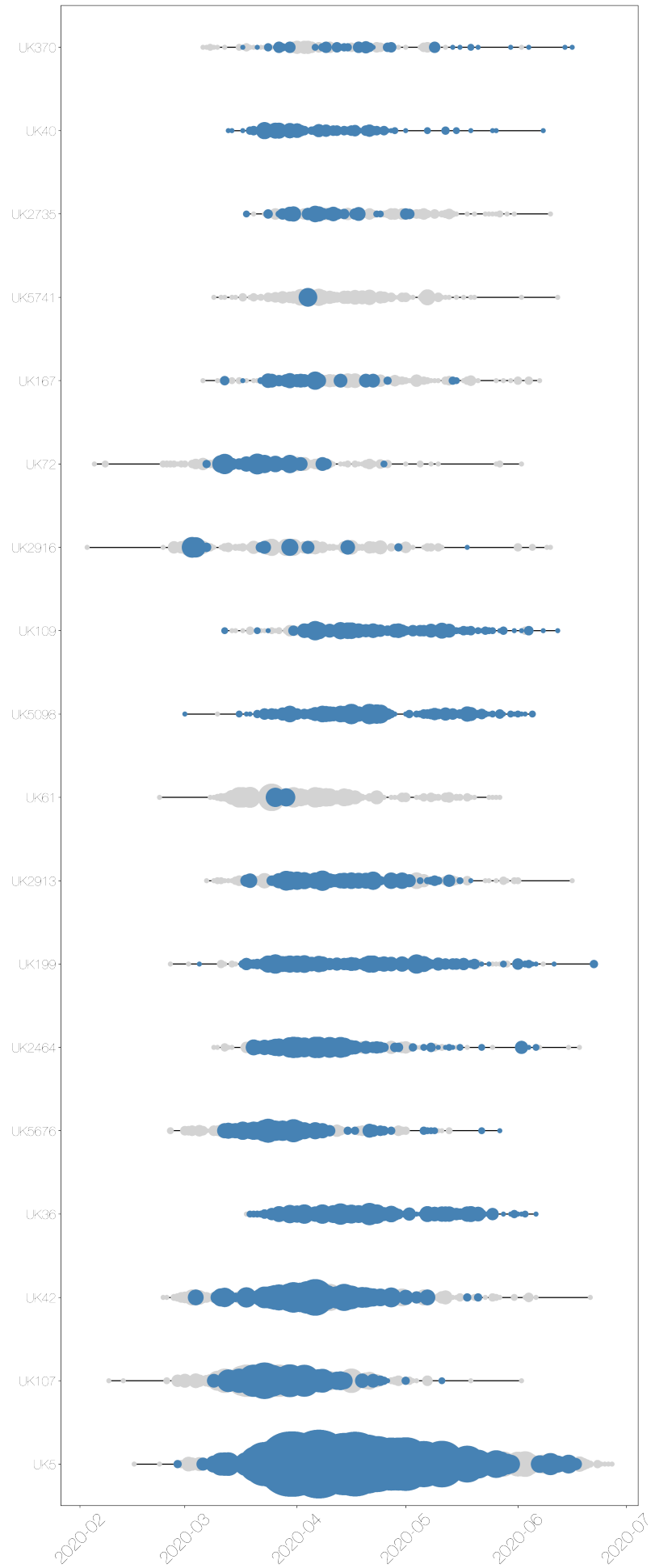


Figure 4: Timeline of lineages, sized by number of sequences from each country.

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

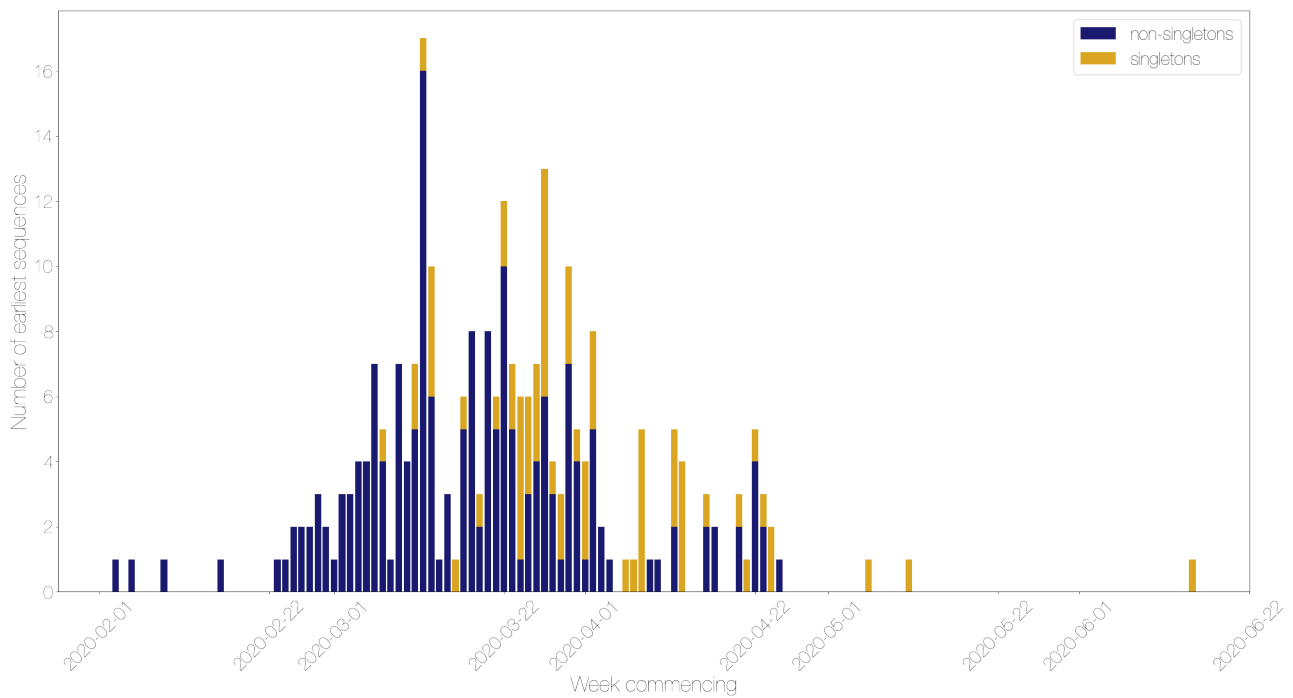


Figure 5: Lineage starts per week, split by singletons and non-singletons

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.

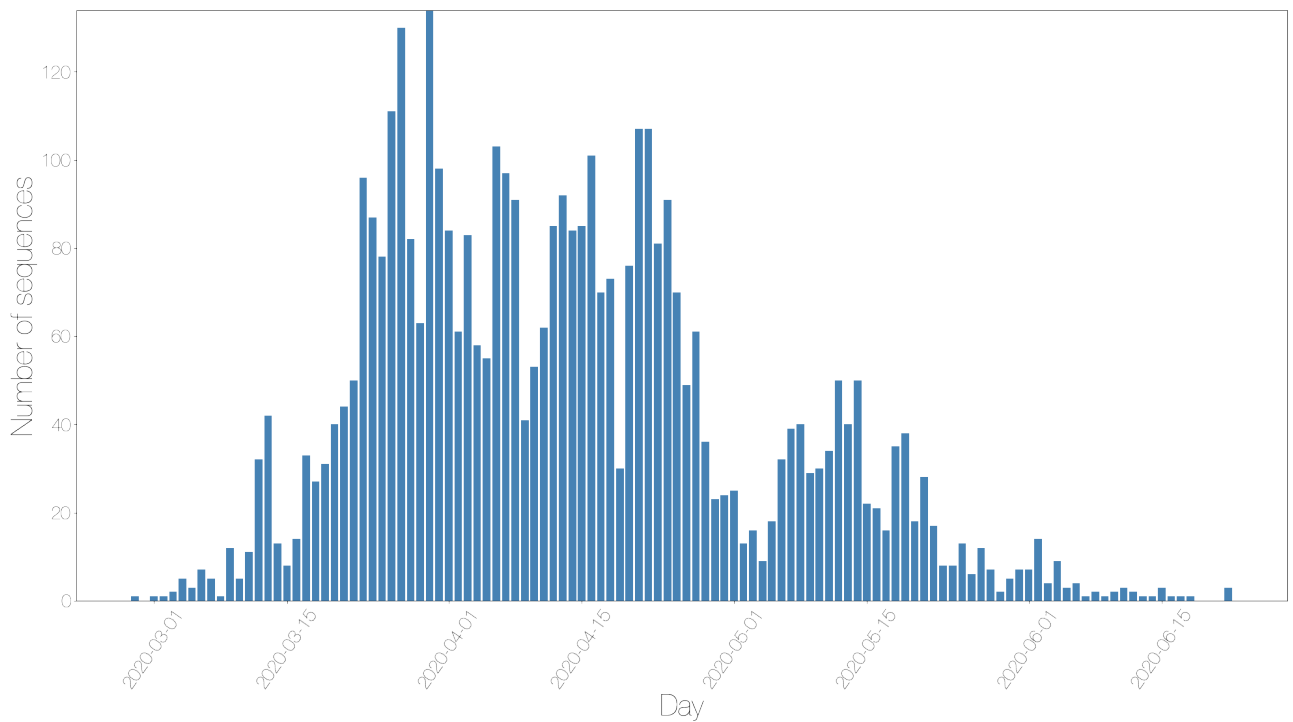


Figure 6: Sequences taken on each day by country

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.

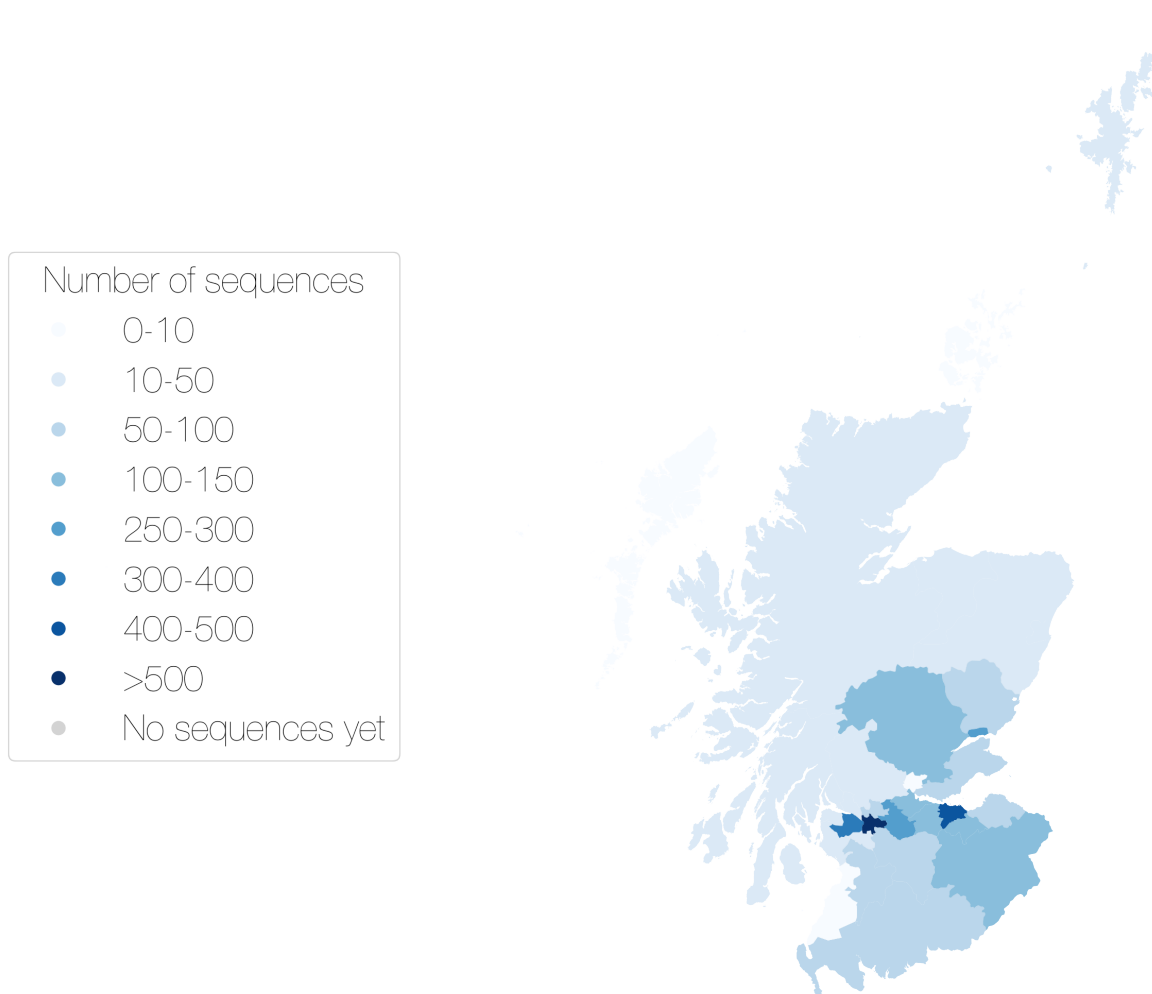


Figure 7: Map showing the number of sequences sampled by adm2 region

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
UK36	Mar-19, Jun-06	536	B.1	16	0.0083
UK5	Feb-28, Jun-21	435	B.1.1.13, B.1.1, B.1.1.p11, B.1.1.4, B.1.1.p12, B.1.1.1, B.1.1.14	1	0.0152
UK5098	Mar-01, Jun-05	433	B.1, B.1.p73	17	0.0129
UK109	Mar-12, Jun-12	244	B.1.5.5, B.1.5	10	0.0232
UK199	Mar-05, Jun-22	227	B.1, B.1.5, B.1.p73	0	active today
UK40	Mar-13, Jun-08	215	B.16, B	14	0.028
UK42	Mar-04, Jun-21	205	B.1.p11, B.1, B.1.p73, B.1.71, B.1.5	1	0.0863
UK2464	Mar-19, Jun-18	180	B.1.p11	4	0.0455
UK5676	Mar-12, May-27	150	B.2	26	0.0061
UK39	Mar-12, May-29	145	A.2	24	0.0221
UK2913	Mar-18, Jun-16	110	B.1.p11	6	0.0317
UK668	Mar-20, Jun-10	87	B.1	12	0.0594
UK72	Mar-07, Jun-02	49	B	20	0.0174
UK107	Mar-09, Jun-02	46	B.2.1	20	0.004
UK100	Mar-22, Jun-01	43	B.1, B.1.5	21	0.0805
UK304	Apr-16, Jun-02	40	B.1.1.14	20	0.0603
UK370	Mar-17, Jun-16	40	B.1.1.10	6	0.0895
UK44	Mar-17, Apr-23	37	B	60	0.015
UK15	Mar-07, May-06	36	B.1.1	47	0.0077
UK14	Mar-14, May-21	33	B	32	0.0625
UK43	Mar-18, Apr-26	33	A.5	57	0.0239
UK87	Mar-13, Apr-24	33	B.1.70	59	0.0222
UK167	Mar-12, Jun-07	32	B.1	15	0.0195
UK2735	Mar-18, Jun-10	32	B.1.1	12	0.0258
UK21	Mar-18, May-23	31	B.1.40	30	0.0733
UK2200	Mar-17, Jun-06	30	B.1.5, B.1.5.6	16	0.0578

Lineage name	Date range	Number of sequences	Global lineage	Time since last sample (days)	Activity score
UK4493	Apr-23, May-19	26	B.1	34	0.0306
UK1667	Mar-31, May-18	25	B.1.p9	35	0.0483
UK501	Mar-19, Jun-18	20	B.1	4	0.2878
UK502	Mar-06, Mar-30	18	B.1.69	84	0.0168
UK58	Mar-12, Apr-24	17	B.1	59	0.0331
UK66	Mar-28, May-20	17	B.1.1.8	33	0.0156
UK120	Mar-02, Jun-07	15	B	15	0.0935
UK722	Mar-23, May-27	15	B.1.5	26	0.1786
UK261	Mar-15, Apr-10	14	A.3	73	0.0274
UK137	Mar-09, Mar-31	13	B.1.1	83	0.0204
UK240	Mar-22, May-27	13	B.2	26	0.0193
UK2916	Mar-03, Jun-10	13	B.1	12	0.0284
UK436	Mar-28, May-14	13	B.1.5	39	0.0927
UK335	Apr-15, Jun-22	12	B.1.1	0	active today
UK5498	Mar-12, May-28	12	B.2, B	25	0.0369
UK187	Mar-21, Apr-30	9	B.1	53	0.0148
UK601	Mar-14, May-11	9	B.10	42	0.015
UK151	Mar-23, Apr-24	8	B.1	59	0.0775
UK548	Mar-14, Mar-30	8	B.2.1	84	0.0272
UK594	Apr-20, May-01	8	B	52	0.0302
UK5322	Mar-22, Jun-04	7	B.1.1	18	0.0437
UK5561	Mar-10, May-24	7	B.2.2, B.2	29	0.0154
UK133	Mar-22, Apr-25	7	B.1	58	0.0651
UK671	Apr-17, May-31	7	B.1.p73	22	0.3333
UK271	Apr-15, Apr-26	7	B.1	57	0.0602
UK51	Mar-26, Jun-20	7	B.1.36	2	0.4789
UK4735	Apr-22, May-27	6	B.1.1	26	0.2692
UK697	Mar-31, Apr-24	6	B.1	59	0.0814



**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	11
1	GLAS	15

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

Week commencing	UK36	UK5	UK5098	UK109	UK199	UK40	UK42	UK2464	UK5676	UK39
2020-02-23	0	1	0	0	0	0	0	0	0	0
2020-03-01	0	1	1	0	1	0	1	0	0	0
2020-03-08	0	7	0	1	0	1	7	0	3	1
2020-03-15	3	7	3	1	6	4	6	6	8	2
2020-03-22	8	15	7	1	8	7	13	12	9	5
2020-03-29	14	12	4	3	7	4	12	9	10	2
2020-04-05	13	21	11	4	8	7	10	9	5	8
2020-04-12	19	15	13	5	11	6	12	13	1	7
2020-04-19	16	15	15	6	12	8	9	8	7	4
2020-04-26	10	6	7	6	5	2	2	3	1	1
2020-05-03	10	7	6	5	4	1	2	5	4	2
2020-05-10	11	7	11	7	4	3	0	4	0	4
2020-05-17	9	9	10	4	3	1	2	1	1	2
2020-05-24	3	6	8	3	2	2	0	0	1	3
2020-05-31	5	0	4	3	2	0	0	1	0	0
2020-06-07	0	3	0	2	1	1	0	0	0	0
2020-06-14	0	2	0	0	0	0	0	0	0	0
2020-06-21	0	0	0	0	1	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
2020-02-03	0	1	1
2020-02-05	0	1	1
2020-02-09	0	1	1
2020-02-16	0	1	1
2020-02-23	0	1	1
2020-02-24	0	1	1
2020-02-25	0	2	2
2020-02-26	0	2	2
2020-02-27	0	2	2
2020-02-28	0	3	3
2020-02-29	0	2	2
2020-03-01	0	1	1
2020-03-02	0	3	3
2020-03-03	0	3	3
2020-03-04	0	4	4
2020-03-05	0	4	4
2020-03-06	0	7	7
2020-03-07	1	4	5
2020-03-08	0	1	1
2020-03-09	0	7	7
2020-03-10	0	4	4
2020-03-11	2	5	7
2020-03-12	1	16	17
2020-03-13	4	6	10
2020-03-14	0	1	1
2020-03-15	0	3	3
2020-03-16	1	0	1
2020-03-17	1	5	6
2020-03-18	0	8	8
2020-03-19	1	2	3
2020-03-20	0	8	8
2020-03-21	1	5	6
2020-03-22	2	10	12
2020-03-23	2	5	7
2020-03-24	5	1	6
2020-03-25	3	3	6
2020-03-26	3	4	7
2020-03-27	7	6	13
2020-03-28	1	3	4
2020-03-29	2	1	3
2020-03-30	3	7	10
2020-03-31	1	4	5
2020-04-01	3	1	4
2020-04-02	3	5	8
2020-04-03	0	2	2
2020-04-04	0	1	1
2020-04-06	1	0	1
2020-04-07	1	0	1
2020-04-08	5	0	5
2020-04-09	0	1	1
2020-04-10	0	1	1
2020-04-12	3	2	5
2020-04-13	4	0	4
2020-04-16	1	2	3
2020-04-17	0	2	2
2020-04-20	1	2	3
2020-04-21	1	0	1

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-04-22	1	4	5
2020-04-23	1	2	3
2020-04-24	2	0	2
2020-04-25	0	1	1
2020-05-06	1	0	1
2020-05-11	1	0	1
2020-06-15	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	12
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	33
2020-03-18	27
2020-03-19	31
2020-03-20	40
2020-03-21	44
2020-03-22	50
2020-03-23	96
2020-03-24	87
2020-03-25	78
2020-03-26	111
2020-03-27	130
2020-03-28	82
2020-03-29	63
2020-03-30	134
2020-03-31	98
2020-04-01	84
2020-04-02	61
2020-04-03	83
2020-04-04	58
2020-04-05	55
2020-04-06	103
2020-04-07	97
2020-04-08	91
2020-04-09	41
2020-04-10	53
2020-04-11	62
2020-04-12	85
2020-04-13	92
2020-04-14	84
2020-04-15	85
2020-04-16	101
2020-04-17	70
2020-04-18	73
2020-04-19	30
2020-04-20	76
2020-04-21	107
2020-04-22	107
2020-04-23	81
2020-04-24	91
2020-04-25	70

Day	Scotland
2020-04-26	49
2020-04-27	61
2020-04-28	36
2020-04-29	23
2020-04-30	24
2020-05-01	25
2020-05-02	13
2020-05-03	16
2020-05-04	9
2020-05-05	18
2020-05-06	32
2020-05-07	39
2020-05-08	40
2020-05-09	29
2020-05-10	30
2020-05-11	34
2020-05-12	50
2020-05-13	40
2020-05-14	50
2020-05-15	22
2020-05-16	21
2020-05-17	16
2020-05-18	35
2020-05-19	38
2020-05-20	18
2020-05-21	28
2020-05-22	17
2020-05-23	8
2020-05-24	8
2020-05-25	13
2020-05-26	6
2020-05-27	12
2020-05-28	7
2020-05-29	2
2020-05-30	5
2020-05-31	7
2020-06-01	7
2020-06-02	14
2020-06-03	4
2020-06-04	9
2020-06-05	3
2020-06-06	4
2020-06-07	1
2020-06-08	2
2020-06-09	1
2020-06-10	2
2020-06-11	3
2020-06-12	2
2020-06-13	1
2020-06-14	1
2020-06-15	3
2020-06-16	1
2020-06-17	1
2020-06-18	1
2020-06-22	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
ABERDEEN	Scotland	23	10-50
ABERDEENSHIRE	Scotland	12	10-50
ANGUS	Scotland	69	50-100
ARGYLL AND BUTE	Scotland	14	10-50
CLACKMANNANSHIRE	Scotland	4	1-10
DUMFRIES AND GALLOWAY	Scotland	88	50-100
DUNDEE	Scotland	278	250-300
EAST AYRSHIRE	Scotland	93	50-100
EAST DUNBARTONSHIRE	Scotland	73	50-100
EAST LoTHIAN	Scotland	57	50-100
EAST RENFREWSHIRE	Scotland	40	10-50
EDINBURGH	Scotland	478	400-500
EILEAN SIAR	Scotland	2	1-10
FALKIRK	Scotland	102	100-150
FIFE	Scotland	51	50-100
GLASGOW	Scotland	1246	>500
HIGHLAND	Scotland	10	10-50
INVERCLYDE	Scotland	42	10-50
MIDLoTHIAN	Scotland	146	100-150
MORAY	Scotland	10	10-50
NORTH AYRSHIRE	Scotland	18	10-50
NORTH LANARKSHIRE	Scotland	273	250-300
ORKNEY ISLANDS	Scotland	1	1-10
PERTHSHIRE AND KINROSS	Scotland	118	100-150
RENFREWSHIRE	Scotland	317	300-400
SCOTTISH BORDERS	Scotland	143	100-150
SHETLAND ISLANDS	Scotland	14	10-50
SOUTH AYRSHIRE	Scotland	7	1-10
SOUTH LANARKSHIRE	Scotland	70	50-100
STIRLING	Scotland	18	10-50
WEST DUNBARTONSHIRE	Scotland	49	10-50
WEST LoTHIAN	Scotland	131	100-150