

Lineages report for PHWC

This report gives summaries of UK specific lineages sequenced by PHWC for week 2020-07-03. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-06-28. The analysis (eg time since last sample) is therefore undertaken from this date. 4129 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 172 and the maximum is 1319

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

168 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 12 that remain: 10 are pending extinction, ie last seen three weeks ago. 1 has reactivated. 1 lineage has 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	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK5	1168 (100.0%)	Mar-01, Jun-27	1168	B.1.1.16, B.1.1.2, B.1.1.p16, B.1.1.1, B.1.1, B.1.1.p11	1	0.1011
UK61	419 (100.0%)	Mar-08, May-27	419	B, B.3	32	0.006
UK42	368 (100.0%)	Feb-27, Jun-06	368	B.1.71, B.1.35, B.1, B.1.p11	22	0.0124
UK632	232 (100.0%)	Mar-25, Jun-09	232	B.1.1	19	0.0173
UK3021	225 (100.0%)	Mar-29, Jun-09	225	B.1	19	0.0169
UK495	124 (100.0%)	Apr-01, Jun-03	124	B.1.p11	25	0.0205
UK5741	104 (100.0%)	Mar-17, Jun-02	104	B.1, B.1.44	26	0.0288
UK822	102 (100.0%)	Apr-14, Jun-11	102	B.1	17	0.0338
UK5322	86 (100.0%)	Apr-08, Jun-04	86	B.1.1	24	0.0279
UK605	79 (100.0%)	Mar-17, May-22	79	B.1.1.10, B.1.1	37	0.0229

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.

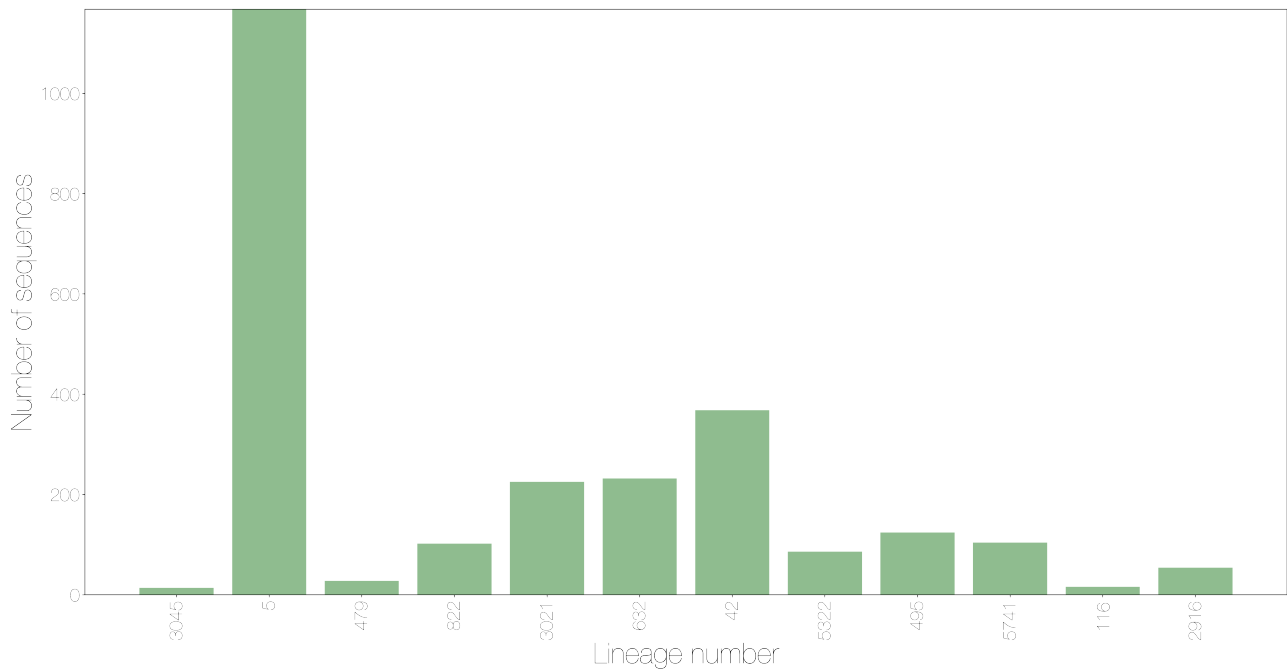


Figure 1: Number of sequences sampled in a lineage by country

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 5 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.

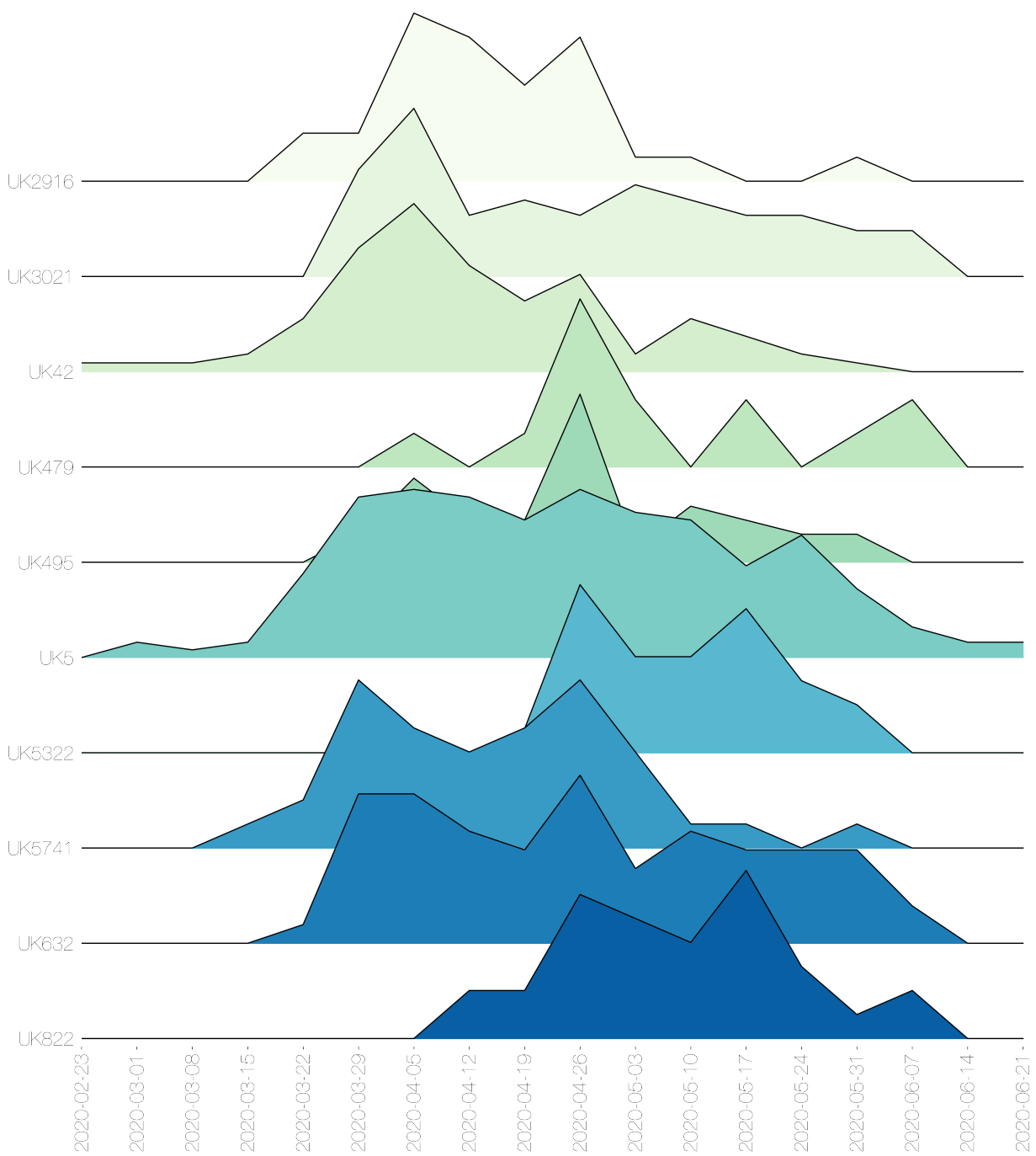


Figure 2: Lineages by number of adm2 regions present by epiweek

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.

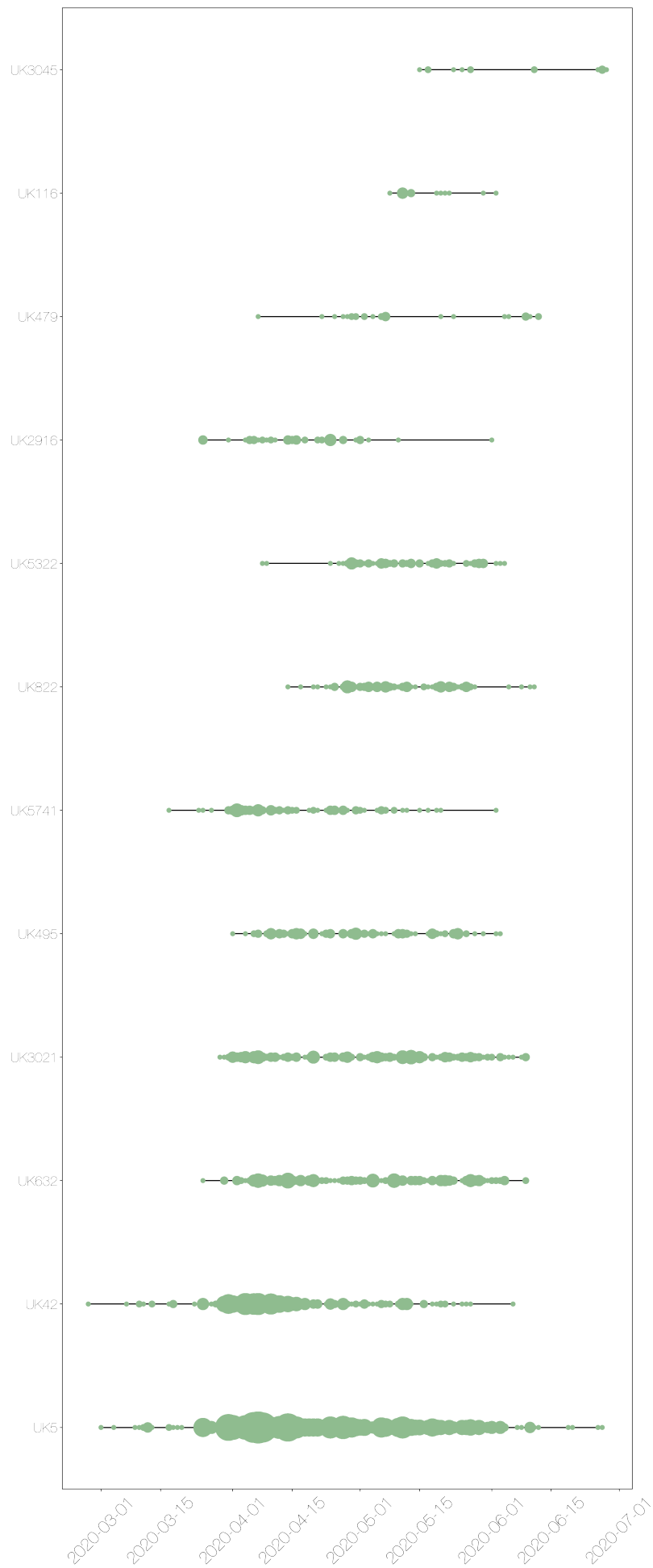


Figure 3: Timeline of lineages, sized by number of sequences from each country.

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

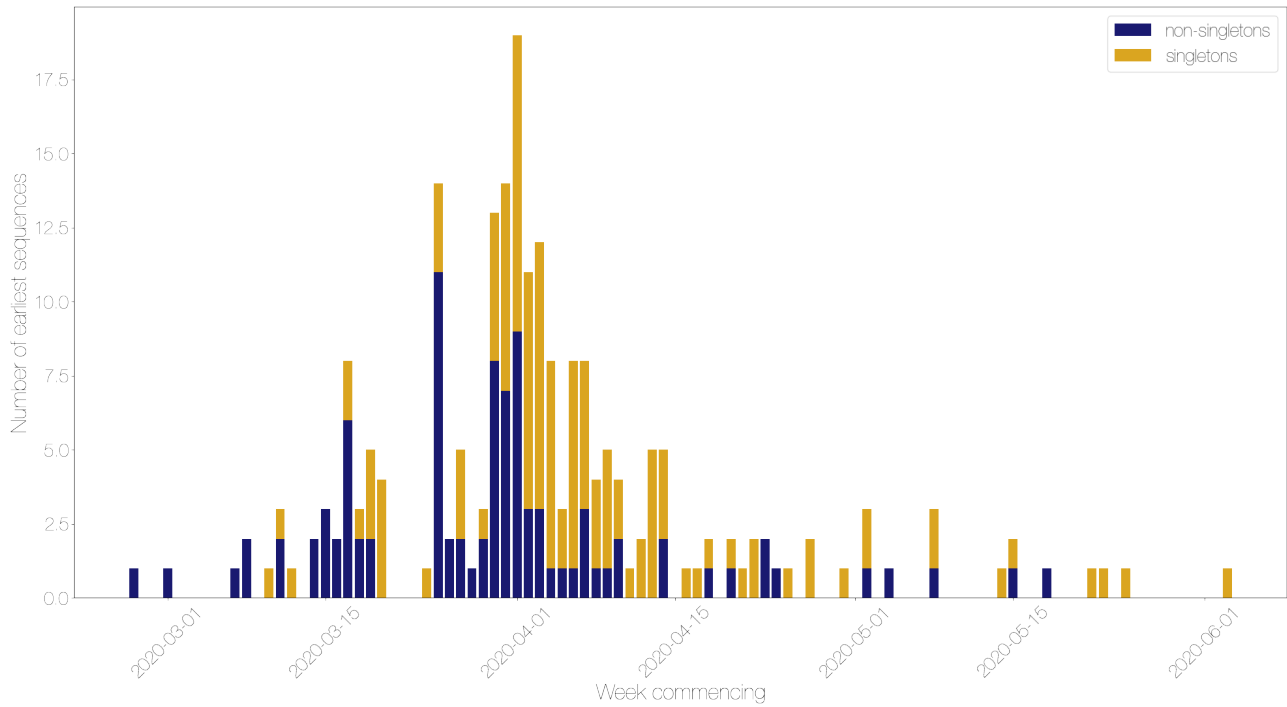


Figure 4: Lineage starts per week, split by singletons and non-singletons

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.

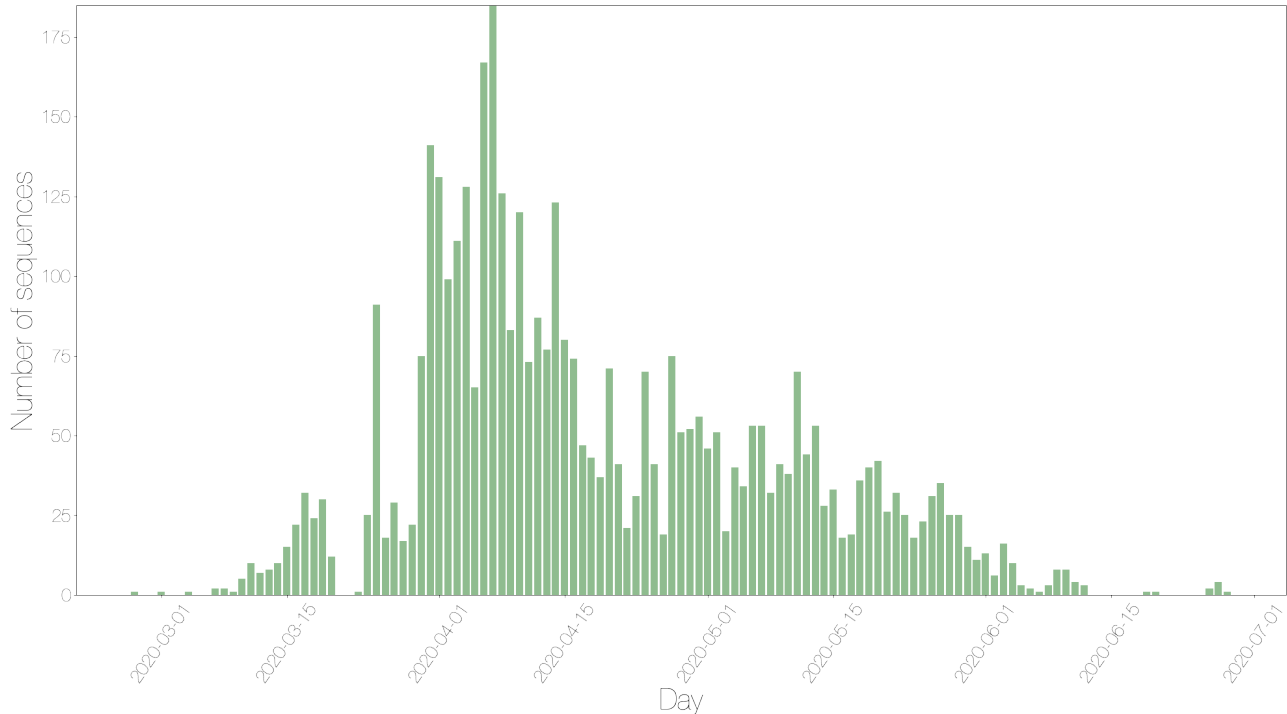


Figure 5: Sequences taken on each day by country

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 585 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
UK5	1168 (100.0%)	Mar-01, Jun-27	1168	B.1.1.16, B.1.1.2, B.1.1.p16, B.1.1.1, B.1.1, B.1.1.p11	1	0.1011
UK61	419 (100.0%)	Mar-08, May-27	419	B, B.3	32	0.006
UK42	368 (100.0%)	Feb-27, Jun-06	368	B.1.71, B.1.35, B.1, B.1.p11	22	0.0124
UK632	232 (100.0%)	Mar-25, Jun-09	232	B.1.1	19	0.0173
UK3021	225 (100.0%)	Mar-29, Jun-09	225	B.1	19	0.0169
UK495	124 (100.0%)	Apr-01, Jun-03	124	B.1.p11	25	0.0205
UK5741	104 (100.0%)	Mar-17, Jun-02	104	B.1, B.1.44	26	0.0288
UK822	102 (100.0%)	Apr-14, Jun-11	102	B.1	17	0.0338
UK5322	86 (100.0%)	Apr-08, Jun-04	86	B.1.1	24	0.0279
UK605	79 (100.0%)	Mar-17, May-22	79	B.1.1.10, B.1.1	37	0.0229
UK2464	78 (100.0%)	Mar-26, May-11	78	B.1.p11	48	0.0124
UK2735	76 (100.0%)	Mar-27, May-31	76	B.1.1	28	0.031
UK86	61 (100.0%)	Mar-30, May-30	61	B.1	29	0.0351
UK107	61 (100.0%)	Mar-14, Apr-23	61	B.2.1, B	66	0.0101
UK199	55 (100.0%)	Mar-18, May-14	55	B.1.5, B.1	45	0.0235
UK2916	54 (100.0%)	Mar-25, Jun-01	54	B.1	27	0.0475
UK5676	54 (100.0%)	Mar-15, May-01	54	B.2	58	0.0153
UK370	50 (100.0%)	Mar-19, Apr-27	50	B.1.1.10	62	0.0128
UK2200	35 (100.0%)	Mar-15, Apr-30	35	B.1.5, B.1.5.6	59	0.0229
UK109	35 (100.0%)	Mar-15, May-19	35	B.1.5	40	0.0478
UK187	29 (100.0%)	Mar-30, Apr-30	29	B.1	59	0.0188
UK479	28 (100.0%)	Apr-07, Jun-12	28	B.1.1	16	0.1528
UK5561	23 (100.0%)	Mar-18, May-24	23	B.2.2	35	0.087
UK600	22 (100.0%)	Apr-01, May-26	22	B.1.1	33	0.0794
UK167	21 (100.0%)	Mar-25, May-19	21	B.1	40	0.0688
UK567	20 (100.0%)	Mar-30, May-15	20	B.2.2	44	0.055

Lineage name	Wales	Date range	Total sequences	Global lineage	Time since last sample (days)	Activity score
UK179	20 (100.0%)	Mar-17, May-07	20	B.1.1.p11	52	0.0516
UK206	19 (100.0%)	Apr-02, May-20	19	B.1	39	0.0684
UK2913	18 (100.0%)	Mar-16, May-24	18	B.1, B.1.p11	35	0.116
UK116	16 (100.0%)	May-08, Jun-02	16	B.1	26	0.0641
UK695	16 (100.0%)	Mar-25, Apr-12	16	B.1.67	77	0.0156
UK72	15 (100.0%)	Mar-11, Apr-17	15	B	72	0.0367
UK202	14 (100.0%)	Apr-24, May-19	14	B.1.1	40	0.0481
UK425	14 (100.0%)	Mar-28, May-05	14	B.1.1	54	0.0541
UK3045	14 (100.0%)	May-15, Jun-28	14	B.1.1, B.1.1.p11	0	active today
UK607	12 (100.0%)	Mar-11, Apr-24	12	B	65	0.0615
UK89	12 (100.0%)	Apr-10, May-28	12	B.1.1.9	31	0.1408
UK317	12 (100.0%)	Mar-19, Apr-20	12	B.3	69	0.0422
UK64	12 (100.0%)	Mar-25, May-05	12	B.1	54	0.069
UK15	11 (100.0%)	Mar-17, Apr-13	11	B.1.1	76	0.0355
UK327	10 (100.0%)	Apr-05, May-05	10	B.1	54	0.0617
UK275	8 (100.0%)	Mar-31, Apr-18	8	B.1.13	71	0.0362
UK633	8 (100.0%)	Apr-03, Apr-28	8	B.1.1.p16, B.1.1.16	61	0.0585
UK696	8 (100.0%)	Apr-10, May-01	8	B.1.5, B.1	58	0.0517
UK5498	7 (100.0%)	Apr-01, Apr-14	7	B.2	75	0.0289
UK462	7 (100.0%)	Apr-01, Apr-20	7	B.1	69	0.0459
UK119	7 (100.0%)	Mar-30, Apr-14	7	B.2.5	75	0.0333
UK451	6 (100.0%)	Mar-25, Apr-05	6	B.2.1	84	0.0262

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	5

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

Week commencing	UK5	UK42	UK632	UK3021	UK495	UK5741	UK822	UK5322	UK2916	UK479
2020-02-23	0	1	0	0	0	0	0	0	0	0
2020-03-01	2	1	0	0	0	0	0	0	0	0
2020-03-08	1	1	0	0	0	0	0	0	0	0
2020-03-15	2	2	0	0	0	1	0	0	0	0
2020-03-22	11	6	1	0	0	2	0	0	2	0
2020-03-29	21	14	8	7	2	7	0	0	2	0
2020-04-05	22	19	8	11	6	5	0	1	7	1
2020-04-12	21	12	6	4	3	4	2	0	6	0
2020-04-19	18	8	5	5	3	5	2	1	4	1
2020-04-26	22	11	9	4	12	7	6	7	6	5
2020-05-03	19	2	4	6	1	4	5	4	1	2
2020-05-10	18	6	6	5	4	1	4	4	1	0
2020-05-17	12	4	5	4	3	1	7	6	0	2
2020-05-24	16	2	5	4	2	0	3	3	0	0
2020-05-31	9	1	5	3	2	1	1	2	1	1
2020-06-07	4	0	2	3	0	0	2	0	0	2
2020-06-14	2	0	0	0	0	0	0	0	0	0
2020-06-21	2	0	0	0	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
2020-02-27	0	1	1
2020-03-01	0	1	1
2020-03-07	0	1	1
2020-03-08	0	2	2
2020-03-10	1	0	1
2020-03-11	1	2	3
2020-03-12	1	0	1
2020-03-14	0	2	2
2020-03-15	0	3	3
2020-03-16	0	2	2
2020-03-17	2	6	8
2020-03-18	1	2	3
2020-03-19	3	2	5
2020-03-20	4	0	4
2020-03-24	1	0	1
2020-03-25	3	11	14
2020-03-26	0	2	2
2020-03-27	3	2	5
2020-03-28	0	1	1
2020-03-29	1	2	3
2020-03-30	5	8	13
2020-03-31	7	7	14
2020-04-01	10	9	19
2020-04-02	8	3	11
2020-04-03	9	3	12
2020-04-04	7	1	8
2020-04-05	2	1	3
2020-04-06	7	1	8
2020-04-07	5	3	8
2020-04-08	3	1	4
2020-04-09	4	1	5
2020-04-10	2	2	4
2020-04-11	1	0	1
2020-04-12	2	0	2
2020-04-13	5	0	5
2020-04-14	3	2	5
2020-04-16	1	0	1
2020-04-17	1	0	1
2020-04-18	1	1	2
2020-04-20	1	1	2
2020-04-21	1	0	1
2020-04-22	2	0	2
2020-04-23	0	2	2
2020-04-24	0	1	1
2020-04-25	1	0	1
2020-04-27	2	0	2
2020-04-30	1	0	1
2020-05-02	2	1	3
2020-05-04	0	1	1
2020-05-08	2	1	3
2020-05-14	1	0	1
2020-05-15	1	1	2
2020-05-18	0	1	1
2020-05-22	1	0	1
2020-05-23	1	0	1
2020-05-25	1	0	1
2020-06-03	1	0	1

Table S6 Raw data for figure six showing the number of sequences taken over time.

Day	Wales
2020-02-27	1
2020-03-01	1
2020-03-04	1
2020-03-07	2
2020-03-08	2
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	24
2020-03-19	30
2020-03-20	12
2020-03-23	1
2020-03-24	25
2020-03-25	91
2020-03-26	18
2020-03-27	29
2020-03-28	17
2020-03-29	22
2020-03-30	75
2020-03-31	141
2020-04-01	131
2020-04-02	99
2020-04-03	111
2020-04-04	128
2020-04-05	65
2020-04-06	167
2020-04-07	185
2020-04-08	126
2020-04-09	83
2020-04-10	120
2020-04-11	73
2020-04-12	87
2020-04-13	77
2020-04-14	123
2020-04-15	80
2020-04-16	74
2020-04-17	47
2020-04-18	43
2020-04-19	37
2020-04-20	71
2020-04-21	41
2020-04-22	21
2020-04-23	31
2020-04-24	70
2020-04-25	41
2020-04-26	19
2020-04-27	75
2020-04-28	51
2020-04-29	52
2020-04-30	56
2020-05-01	46

Day	Wales
2020-05-02	51
2020-05-03	20
2020-05-04	40
2020-05-05	34
2020-05-06	53
2020-05-07	53
2020-05-08	32
2020-05-09	41
2020-05-10	38
2020-05-11	70
2020-05-12	44
2020-05-13	53
2020-05-14	28
2020-05-15	33
2020-05-16	18
2020-05-17	19
2020-05-18	36
2020-05-19	40
2020-05-20	42
2020-05-21	26
2020-05-22	32
2020-05-23	25
2020-05-24	18
2020-05-25	23
2020-05-26	31
2020-05-27	35
2020-05-28	25
2020-05-29	25
2020-05-30	15
2020-05-31	11
2020-06-01	13
2020-06-02	6
2020-06-03	16
2020-06-04	10
2020-06-05	3
2020-06-06	2
2020-06-07	1
2020-06-08	3
2020-06-09	8
2020-06-10	8
2020-06-11	4
2020-06-12	3
2020-06-19	1
2020-06-20	1
2020-06-26	2
2020-06-27	4
2020-06-28	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
ANGLESEY	Wales	80	50-100
BLAENAU GWENT	Wales	59	50-100
BRIDGEND	Wales	114	100-150
CAERPHILLY	Wales	142	100-150
CARDIFF	Wales	585	>500
CARMARTHENSHIRE	Wales	147	100-150
CEREDIGION	Wales	16	10-50
CONWY	Wales	162	150-200
DENBIGHSHIRE	Wales	194	150-200
FLINTSHIRE	Wales	131	100-150
GWYNEDD	Wales	123	100-150
MERTHYR TYDFIL	Wales	103	100-150
MONMOUTHSHIRE	Wales	85	50-100
NEATH PORT TALBOT	Wales	119	100-150
NEWPORT	Wales	165	150-200
PEMBROKESHIRE	Wales	73	50-100
POWYS	Wales	77	50-100
SWANSEA	Wales	276	250-300
TORFAEN	Wales	91	50-100
VALE OF GLAMORGAN	Wales	191	150-200
WREXHAM	Wales	166	150-200