

Lineages report for PHWC

This report gives summaries of UK specific lineages sequenced by PHWC for week 2020-09-13. There are time lags due to batching, curation and analysis, the most recently sampled sequence is 2020-09-07. The analysis (eg time since last sample) is therefore undertaken from this date. 6475 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 21 and the maximum is 2256

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

361 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 9 that remain: 2 are pending extinction, ie last seen three weeks ago. 3 lineages have gone quiet, ie haven't been seen this week. 2 lineages have reactivated. 2 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 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 2020-07-20

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	Global lineage	Total
UK5	934 (100.0%)	Mar-17, Sep-05	B.1.1.10, B.1.1.30, B.1.1.2, B.1.1.16, B.1.1.21, B.1.1	934 taxa
UK61	553 (100.0%)	Mar-08, Aug-03	B.3, B, B.6	553 taxa
UK175	540 (100.0%)	Feb-27, Jun-27	B.1.71, B.1.35, B.1.5, B.1	540 taxa
UK1091	530 (100.0%)	Mar-20, Jul-15	B.1.1.2, B.1.1	530 taxa
UK1037	381 (100.0%)	Mar-25, Aug-19	B.1.1.30, B.1.1	381 taxa
UK699	309 (100.0%)	Mar-24, Jun-25	B.1.5, B.1, B.1.5.9	309 taxa
UK5741	165 (100.0%)	Mar-17, Jun-19	B.1.44, B.1.35, B.1.122, B.1, B.1.5	165 taxa
UK822	146 (100.0%)	Apr-14, Jun-30	B.1.120, B.1	146 taxa
UK1133	127 (100.0%)	Apr-08, Jun-12	B.1.1, B.1.1.29	127 taxa
UK1683	105 (100.0%)	Mar-12, Jun-01	B.1.1, B.1.1.1	105 taxa

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.

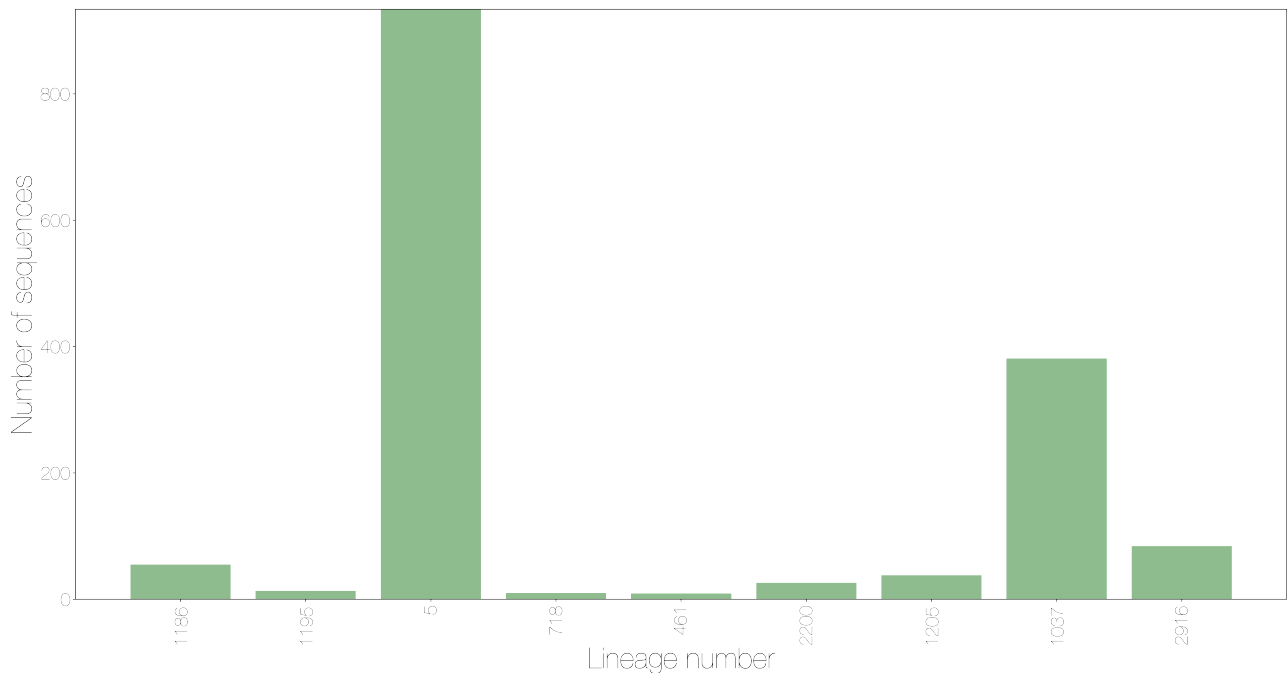


Figure 1: Number of sequences sampled in a lineage by country

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.

```
-----NameError
Traceback (most recent call last)<ipython-input-1-2620455843ef> in
<module>
      2     lag_dict, lags = dp.sequencing_centre_lags(taxa, sc_dict,
current_date, country)
      3 elif sequencing_centre != "":
----> 4     print("The lag for this sequencing centre is " +
str(lags[sequencing_centre]) + " days")
NameError: name 'lags' is not defined
```

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.

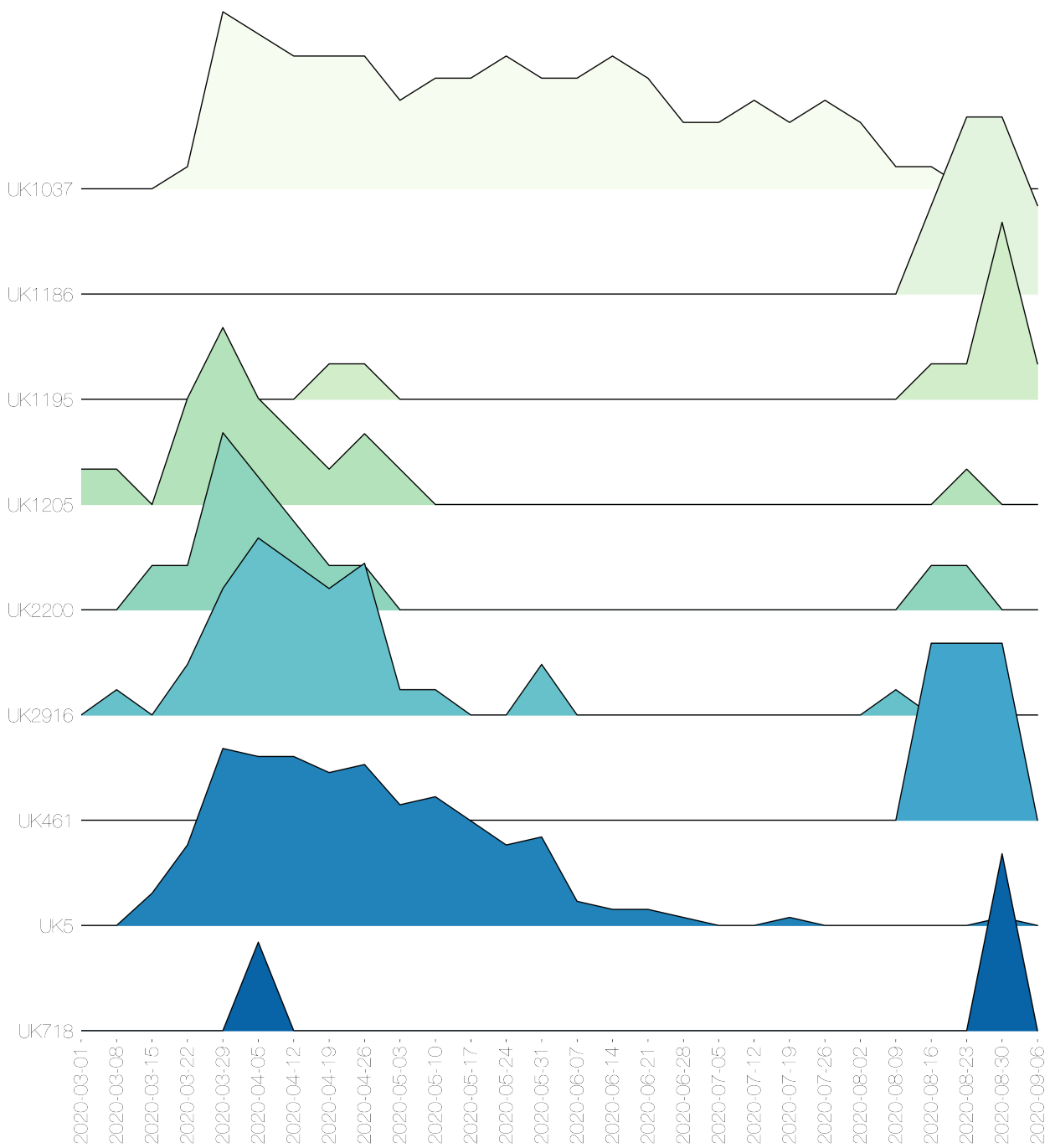


Figure 2: Lineages by number of adm2 regions present by epiweek

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.

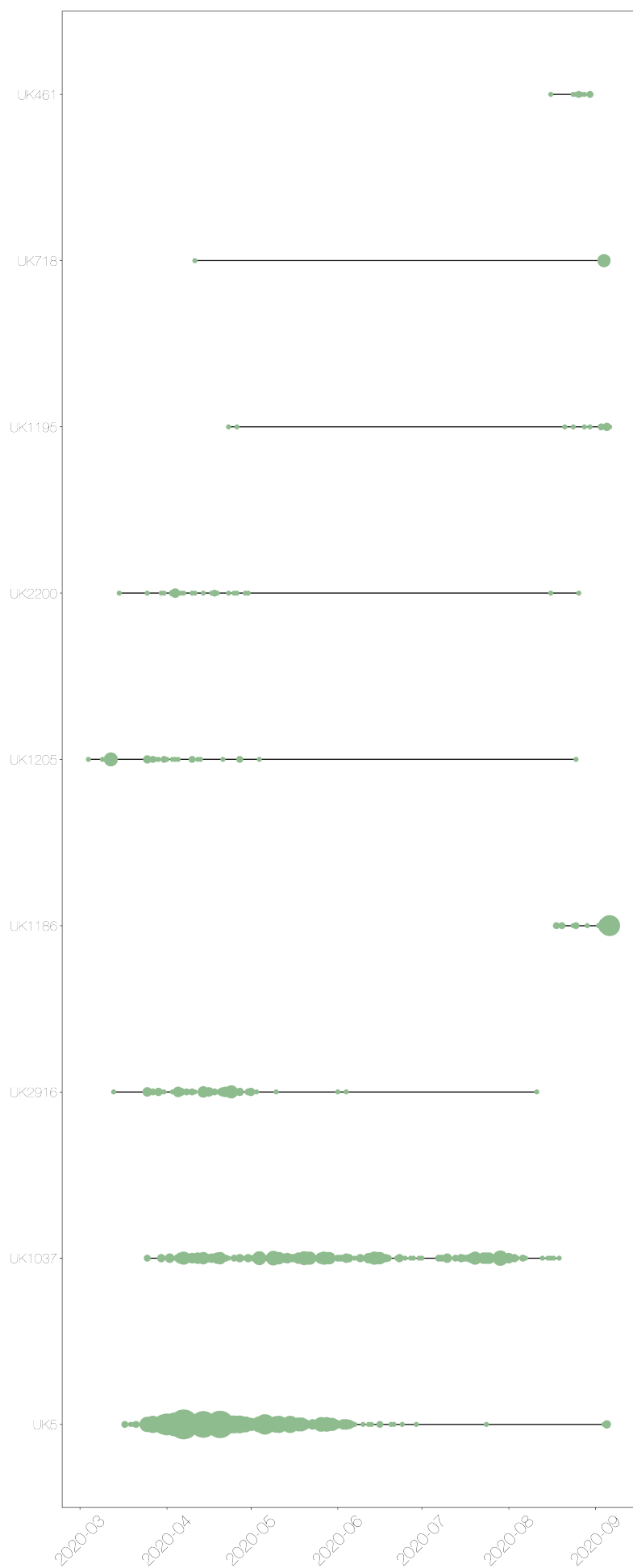


Figure 3: Timeline of lineages, sized by number of sequences from each country.

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

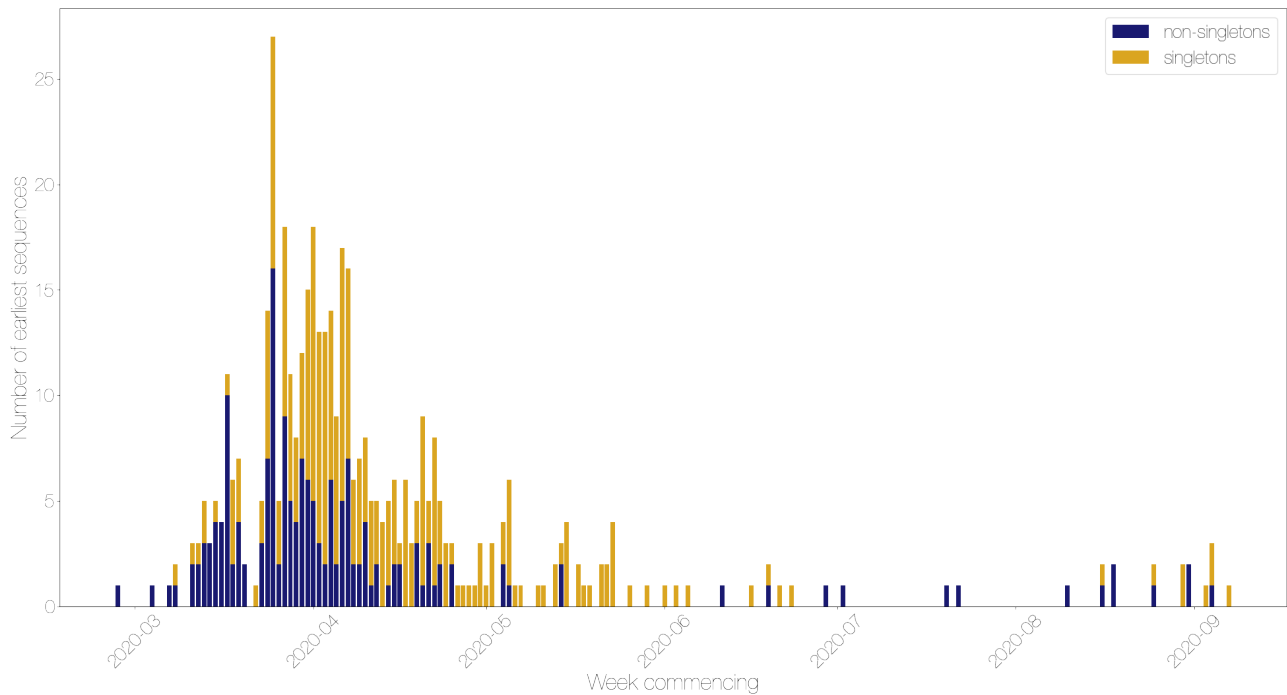


Figure 4: Lineage starts per week, split by singletons and non-singletons

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.

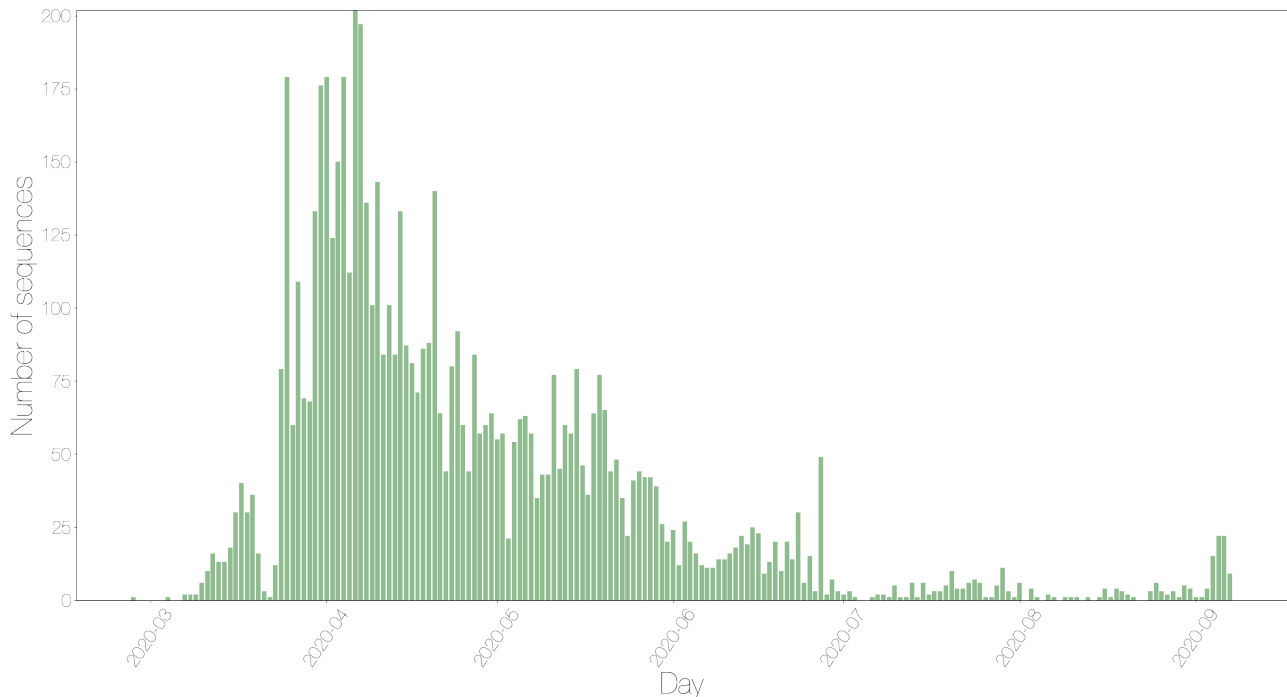


Figure 5: Sequences taken on each day by country

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 1090 sequences without enough geographical information to map from this centre.

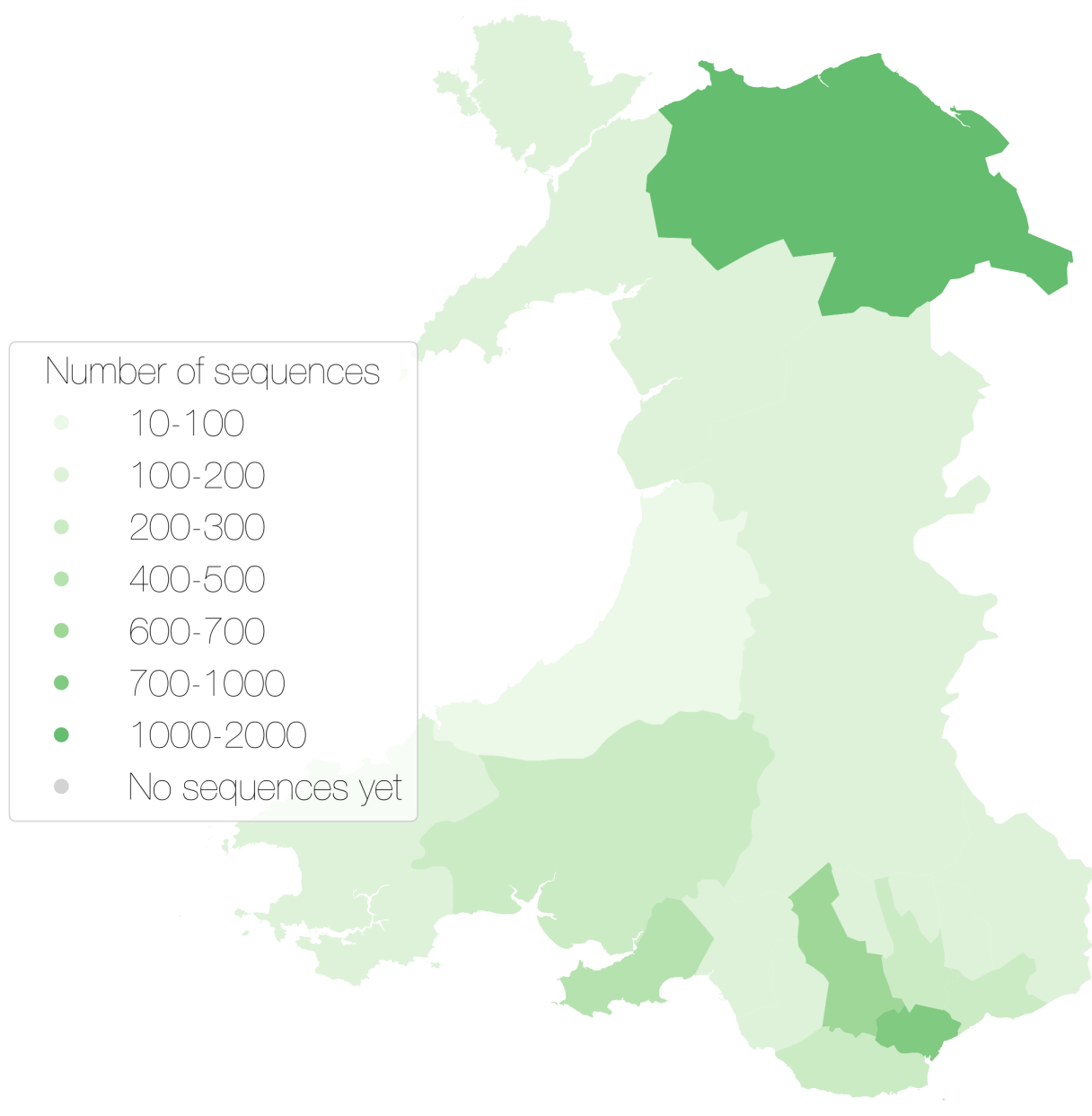


Figure 6: Map showing the number of sequences sampled by adm2 region

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	Global lineage	Total
UK5	934 (100.0%)	Mar-17, Sep-05	B.1.1.10, B.1.1.30, B.1.1.2, B.1.1.16, B.1.1.21, B.1.1	934 taxa

Lineage name	Wales	Date range	Global lineage	Total
UK61	553 (100.0%)	Mar-08, Aug-03	B.3, B, B.6	553 taxa
UK175	540 (100.0%)	Feb-27, Jun-27	B.1.71, B.1.35, B.1.5, B.1	540 taxa
UK1091	530 (100.0%)	Mar-20, Jul-15	B.1.1.2, B.1.1	530 taxa
UK1037	381 (100.0%)	Mar-25, Aug-19	B.1.1.30, B.1.1	381 taxa
UK699	309 (100.0%)	Mar-24, Jun-25	B.1.5, B.1, B.1.5.9	309 taxa
UK5741	165 (100.0%)	Mar-17, Jun-19	B.1.44, B.1.35, B.1.122, B.1, B.1.5	165 taxa
UK822	146 (100.0%)	Apr-14, Jun-30	B.1.120, B.1	146 taxa
UK1133	127 (100.0%)	Apr-08, Jun-12	B.1.1, B.1.1.29	127 taxa
UK1683	105 (100.0%)	Mar-12, Jun-01	B.1.1, B.1.1.1	105 taxa
UK992	103 (100.0%)	Mar-28, Jun-18	B.1.11, B.1	103 taxa
UK2464	102 (100.0%)	Mar-17, May-11	B.1	102 taxa
UK1076	95 (100.0%)	Mar-17, Jul-28	B.1.1	95 taxa
UK1119	86 (100.0%)	Apr-07, Jun-30	B.1.1	86 taxa
UK2916	84 (100.0%)	Mar-13, Aug-11	B.1, B.1.98	84 taxa
UK1951	81 (100.0%)	Mar-18, Jun-19	B.1.1, B.1.1.1	81 taxa
UK107	75 (100.0%)	Mar-17, May-28	B.2.1, B.2	75 taxa
UK5676	72 (100.0%)	Mar-15, May-06	B.2	72 taxa
UK805	68 (100.0%)	May-14, Jul-21	B.1.1	68 taxa
UK1065	66 (100.0%)	Apr-04, Jul-02	B.1.1	66 taxa
UK2067	65 (100.0%)	Mar-23, May-31	B.1.1	65 taxa
UK199	65 (100.0%)	Mar-17, May-21	B.1.5, B.1	65 taxa
UK356	58 (100.0%)	Apr-07, Jul-13	B.1, B.1.11	58 taxa
UK1186	55 (100.0%)	Aug-18, Sep-07	B.1.79, B.1	55 taxa
UK692	48 (100.0%)	Apr-23, Jun-08	B.1.35, B.1.5, B.1	48 taxa
UK719	45 (100.0%)	Mar-25, May-29	B.1.1, B.1.1.10	45 taxa
UK109	42 (100.0%)	Mar-15, May-19	B.1.99, B.1	42 taxa
UK86	39 (100.0%)	Mar-24, May-28	B.1	39 taxa
UK1205	38 (100.0%)	Mar-04, Aug-25	B.1.1, B.1.1.1	38 taxa
UK945	38 (100.0%)	Mar-23, Jun-15	B.1.1	38 taxa

Lineage name	Wales	Date range	Global lineage	Total
UK1060	36 (100.0%)	Mar-12, Aug-09	B.1.1	36 taxa
UK1913	35 (100.0%)	Apr-03, May-14	B.1.1	35 taxa
UK187	34 (100.0%)	Mar-30, Apr-30	B.1	34 taxa
UK597	34 (100.0%)	Mar-27, Jun-25	B.1.35, B.1.5, B.1	34 taxa
UK167	32 (100.0%)	Mar-24, Jun-04	B.1	32 taxa
UK336	30 (100.0%)	Mar-29, Jun-02	B.1	30 taxa
UK387	29 (100.0%)	Mar-15, May-29	B.1	29 taxa
UK345	28 (100.0%)	Mar-30, May-15	B.2.2	28 taxa
UK2913	28 (100.0%)	Mar-16, May-24	B.1, B.1.11	28 taxa
UK1487	27 (100.0%)	Mar-24, May-07	B.1.5, B.1	27 taxa
UK315	27 (100.0%)	Mar-16, Apr-30	B.2.2, B.2	27 taxa
UK2200	26 (100.0%)	Mar-15, Aug-26	B.1.5.6, B.1.5, B.1	26 taxa
UK59	26 (100.0%)	Mar-14, Apr-14	B.3	26 taxa
UK607	25 (100.0%)	Mar-11, Apr-24	B	25 taxa
UK695	22 (100.0%)	Mar-25, Apr-12	B.1.67, B.1	22 taxa
UK1026	21 (100.0%)	Mar-28, May-14	B.1.1	21 taxa
UK508	21 (100.0%)	Mar-19, Apr-19	B.1.35, B.1, B.1.5.6	21 taxa
UK72	21 (100.0%)	Mar-11, Apr-17	B.2.1, B	21 taxa
UK1235	19 (100.0%)	Mar-30, May-07	B.1.1	19 taxa
UK1044	17 (100.0%)	Mar-24, May-01	B.1.5, B.1	17 taxa
UK151	17 (100.0%)	Mar-19, Apr-20	B.3	17 taxa
UK1323	16 (100.0%)	Apr-10, Jun-20	B.1.1	16 taxa
UK402	16 (100.0%)	Apr-10, Jun-01	B.1	16 taxa
UK600	15 (100.0%)	Mar-17, Apr-13	B.1, B.1.1	15 taxa
UK893	14 (100.0%)	Mar-25, May-05	B.1	14 taxa
UK1684	13 (100.0%)	Apr-08, May-06	B.1.1.1	13 taxa
UK1939	13 (100.0%)	Mar-25, Apr-20	B.1.1	13 taxa
UK1195	13 (100.0%)	Apr-23, Sep-06	B.1.1, B.1.1.25	13 taxa
UK1242	12 (100.0%)	Apr-22, Jun-03	B.1.1	12 taxa

Lineage name	Wales	Date range	Global lineage	Total
UK718	10 (100.0%)	Apr-11, Sep-05	B.1.1	10 taxa
UK1535	10 (100.0%)	Mar-27, May-22	B.1.1	10 taxa
UK347	9 (100.0%)	Apr-10, May-01	B.1	9 taxa
UK501	9 (100.0%)	Apr-01, May-16	B.1	9 taxa
UK275	9 (100.0%)	Mar-31, Apr-20	B.1, B.1.13	9 taxa
UK461	9 (100.0%)	Aug-16, Aug-30	B.1	9 taxa
UK4	8 (100.0%)	Mar-17, Apr-11	B	8 taxa
UK119	8 (100.0%)	Mar-30, Apr-22	B.2	8 taxa
UK449	8 (100.0%)	Mar-27, Apr-26	B.1.5, B.1.120, B.1	8 taxa
UK5498	8 (100.0%)	Mar-27, Apr-14	B.2	8 taxa
UK1157	8 (100.0%)	Mar-17, Jun-01	B.1.1.7, B.1.1	8 taxa
UK1987	7 (100.0%)	Mar-20, May-01	B.1.1	7 taxa
UK462	7 (100.0%)	Apr-01, Apr-20	B.1	7 taxa
UK139	7 (100.0%)	Mar-14, Mar-26	B.2.1	7 taxa
UK1352	7 (100.0%)	Mar-25, Apr-05	B.2.1	7 taxa
UK527	7 (100.0%)	Mar-24, Apr-10	B.1	7 taxa
UK1636	7 (100.0%)	Mar-27, Apr-07	B.1.1	7 taxa
UK698	7 (100.0%)	Apr-01, Aug-01	B.1.1	7 taxa
UK993	7 (100.0%)	Mar-27, Apr-14	B.1.1	7 taxa
UK1155	7 (100.0%)	Mar-07, Jun-03	B.1.1	7 taxa
UK184	6 (100.0%)	Mar-25, Apr-01	B.2.1	6 taxa
UK1347	6 (100.0%)	May-05, May-11	B.1.1	6 taxa
UK1126	6 (100.0%)	Mar-19, Apr-24	B.1.1	6 taxa

Table S2 Raw data for figure two showing lags between the most recent sequence and current date for each sequencing centre

-----NameError Traceback (most recent call last) in 1 if not
pillar2: ---> 2 lag_df = pd.DataFrame(lag_dict) 3 print(lag_df.to_markdown()) 4 else: 5 print("Table S2 is
not appropriate for this report and so has been omitted.") NameError: name 'lag_dict' is not defined

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

Week commencing	UK5	UK1037	UK2916	UK1186	UK1205	UK2200	UK1195	UK718	UK461
2020-03-01	0	0	0	0	1	0	0	0	0

Week commencing	UK5	UK1037	UK2916	UK1186	UK1205	UK2200	UK1195	UK718	UK461
2020-03-08	0	0	1	0	1	0	0	0	0
2020-03-15	4	0	0	0	0	1	0	0	0
2020-03-22	10	1	2	0	3	1	0	0	0
2020-03-29	22	8	5	0	5	4	0	0	0
2020-04-05	21	7	7	0	3	3	0	1	0
2020-04-12	21	6	6	0	2	2	0	0	0
2020-04-19	19	6	5	0	1	1	1	0	0
2020-04-26	20	6	6	0	2	1	1	0	0
2020-05-03	15	4	1	0	1	0	0	0	0
2020-05-10	16	5	1	0	0	0	0	0	0
2020-05-17	13	5	0	0	0	0	0	0	0
2020-05-24	10	6	0	0	0	0	0	0	0
2020-05-31	11	5	2	0	0	0	0	0	0
2020-06-07	3	5	0	0	0	0	0	0	0
2020-06-14	2	6	0	0	0	0	0	0	0
2020-06-21	2	5	0	0	0	0	0	0	0
2020-06-28	1	3	0	0	0	0	0	0	0
2020-07-05	0	3	0	0	0	0	0	0	0
2020-07-12	0	4	0	0	0	0	0	0	0
2020-07-19	1	3	0	0	0	0	0	0	0
2020-07-26	0	4	0	0	0	0	0	0	0
2020-08-02	0	3	0	0	0	0	0	0	0
2020-08-09	0	1	1	0	0	0	0	0	0
2020-08-16	0	1	0	2	0	1	1	0	1
2020-08-23	0	0	0	4	1	1	1	0	1
2020-08-30	1	0	0	4	0	0	5	2	1
2020-09-06	0	0	0	2	0	0	1	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
2020-03-04	0	1	1
2020-03-07	0	1	1
2020-03-08	1	1	2
2020-03-11	1	2	3
2020-03-12	1	2	3
2020-03-13	2	3	5
2020-03-14	0	3	3
2020-03-15	1	4	5
2020-03-16	0	4	4
2020-03-17	1	10	11
2020-03-18	4	2	6
2020-03-19	3	4	7
2020-03-20	0	2	2
2020-03-22	1	0	1
2020-03-23	2	3	5
2020-03-24	7	7	14
2020-03-25	11	16	27
2020-03-26	3	2	5
2020-03-27	9	9	18
2020-03-28	6	5	11
2020-03-29	4	4	8
2020-03-30	5	7	12
2020-03-31	9	6	15
2020-04-01	13	5	18

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-04-02	10	3	13
2020-04-03	11	2	13
2020-04-04	8	6	14
2020-04-05	7	2	9
2020-04-06	12	5	17
2020-04-07	9	7	16
2020-04-08	4	2	6
2020-04-09	5	2	7
2020-04-10	4	4	8
2020-04-11	4	1	5
2020-04-12	3	2	5
2020-04-13	4	0	4
2020-04-14	4	1	5
2020-04-15	4	2	6
2020-04-16	1	2	3
2020-04-17	6	0	6
2020-04-18	3	0	3
2020-04-19	2	3	5
2020-04-20	8	1	9
2020-04-21	2	3	5
2020-04-22	7	1	8
2020-04-23	3	2	5
2020-04-24	3	0	3
2020-04-25	1	2	3
2020-04-26	1	0	1
2020-04-27	1	0	1
2020-04-28	1	0	1
2020-04-29	1	0	1
2020-04-30	3	0	3
2020-05-01	1	0	1
2020-05-02	3	0	3
2020-05-04	2	2	4
2020-05-05	5	1	6
2020-05-06	1	0	1
2020-05-07	1	0	1
2020-05-10	1	0	1
2020-05-11	1	0	1
2020-05-13	2	0	2
2020-05-14	1	2	3
2020-05-15	4	0	4
2020-05-17	2	0	2
2020-05-18	1	0	1
2020-05-19	1	0	1
2020-05-21	2	0	2
2020-05-22	2	0	2
2020-05-23	4	0	4
2020-05-26	1	0	1
2020-05-29	1	0	1
2020-06-01	1	0	1
2020-06-03	1	0	1
2020-06-05	1	0	1
2020-06-11	0	1	1
2020-06-16	1	0	1
2020-06-19	1	1	2
2020-06-21	1	0	1
2020-06-23	1	0	1
2020-06-29	0	1	1
2020-07-02	0	1	1
2020-07-20	0	1	1

Day	Number of singleton starts	Number of non-singleton starts	Total
2020-07-22	0	1	1
2020-08-10	0	1	1
2020-08-16	1	1	2
2020-08-18	0	2	2
2020-08-25	1	1	2
2020-08-30	2	0	2
2020-08-31	0	2	2
2020-09-03	1	0	1
2020-09-04	2	1	3
2020-09-07	1	0	1

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

Day	Wales
2020-02-27	1
2020-03-04	1
2020-03-07	2
2020-03-08	2
2020-03-09	2
2020-03-10	6
2020-03-11	10
2020-03-12	16
2020-03-13	13
2020-03-14	13
2020-03-15	18
2020-03-16	30
2020-03-17	40
2020-03-18	30
2020-03-19	36
2020-03-20	16
2020-03-21	3
2020-03-22	1
2020-03-23	12
2020-03-24	79
2020-03-25	179
2020-03-26	60
2020-03-27	109
2020-03-28	69
2020-03-29	68
2020-03-30	133
2020-03-31	176
2020-04-01	179
2020-04-02	124
2020-04-03	150
2020-04-04	179
2020-04-05	112
2020-04-06	202
2020-04-07	197
2020-04-08	136
2020-04-09	101
2020-04-10	143
2020-04-11	84
2020-04-12	101
2020-04-13	84
2020-04-14	133
2020-04-15	87
2020-04-16	81
2020-04-17	71

Day	Wales
2020-04-18	86
2020-04-19	88
2020-04-20	140
2020-04-21	64
2020-04-22	44
2020-04-23	80
2020-04-24	92
2020-04-25	60
2020-04-26	44
2020-04-27	84
2020-04-28	57
2020-04-29	60
2020-04-30	64
2020-05-01	55
2020-05-02	57
2020-05-03	21
2020-05-04	54
2020-05-05	62
2020-05-06	63
2020-05-07	57
2020-05-08	35
2020-05-09	43
2020-05-10	43
2020-05-11	77
2020-05-12	45
2020-05-13	60
2020-05-14	57
2020-05-15	79
2020-05-16	46
2020-05-17	36
2020-05-18	64
2020-05-19	77
2020-05-20	65
2020-05-21	44
2020-05-22	48
2020-05-23	35
2020-05-24	22
2020-05-25	41
2020-05-26	44
2020-05-27	42
2020-05-28	42
2020-05-29	39
2020-05-30	26
2020-05-31	20
2020-06-01	24
2020-06-02	12
2020-06-03	27
2020-06-04	20
2020-06-05	16
2020-06-06	12
2020-06-07	11
2020-06-08	11
2020-06-09	14
2020-06-10	14
2020-06-11	16
2020-06-12	18
2020-06-13	22
2020-06-14	19
2020-06-15	25

Day	Wales
2020-06-16	23
2020-06-17	9
2020-06-18	13
2020-06-19	20
2020-06-20	10
2020-06-21	20
2020-06-22	14
2020-06-23	30
2020-06-24	6
2020-06-25	15
2020-06-26	3
2020-06-27	49
2020-06-28	2
2020-06-29	7
2020-06-30	3
2020-07-01	2
2020-07-02	3
2020-07-03	1
2020-07-06	1
2020-07-07	2
2020-07-08	2
2020-07-09	1
2020-07-10	5
2020-07-11	1
2020-07-12	1
2020-07-13	6
2020-07-14	1
2020-07-15	6
2020-07-16	2
2020-07-17	3
2020-07-18	3
2020-07-19	5
2020-07-20	10
2020-07-21	4
2020-07-22	4
2020-07-23	6
2020-07-24	7
2020-07-25	6
2020-07-26	1
2020-07-27	1
2020-07-28	5
2020-07-29	11
2020-07-30	3
2020-07-31	1
2020-08-01	6
2020-08-03	4
2020-08-04	1
2020-08-06	2
2020-08-07	1
2020-08-09	1
2020-08-10	1
2020-08-11	1
2020-08-13	1
2020-08-15	1
2020-08-16	4
2020-08-17	1
2020-08-18	4
2020-08-19	3
2020-08-20	2

Day	Wales
2020-08-21	1
2020-08-24	3
2020-08-25	6
2020-08-26	3
2020-08-27	2
2020-08-28	3
2020-08-29	1
2020-08-30	5
2020-08-31	4
2020-09-01	1
2020-09-02	1
2020-09-03	4
2020-09-04	15
2020-09-05	22
2020-09-06	22
2020-09-07	9

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	148	100-200
BLAENAU GWENT	Wales	106	100-200
BRIDGEND	Wales	178	100-200
CAERPHILLY	Wales	244	200-300
CARDIFF	Wales	822	700-1000
CARMARTHENSHIRE	Wales	221	200-300
CEREDIGION	Wales	21	10-100
CLWYD	Wales	1112	1000-2000
GWYNEDD	Wales	176	100-200
MERTHYR TYDFIL	Wales	190	100-200
MONMOUTHSHIRE	Wales	104	100-200
NEATH PORT TALBOT	Wales	186	100-200
NEWPORT	Wales	218	200-300
PEMBROKESHIRE	Wales	100	100-200
POWYS	Wales	114	100-200
RHONDDA, CYNON, TAFF	Wales	620	600-700
SWANSEA	Wales	426	400-500
TORFAEN	Wales	113	100-200
VALE OF GLAMORGAN	Wales	286	200-300