Lineages report for GSTT

———————————————————————————KeyError Traceback (most recent call last) in 2 3 if sequencing_centre != “”: —-> 4 country = sc_dict[sequencing_centre] 5 else: 6 country = adm1 KeyError: ‘GSTT’

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.

———————————————————————————NameError Traceback (most recent call last) in 2 print(“The minimum number of introductions is” + str(len(specific_min)) + " and the maximum is " + str(len(specific_max))) 3 else: —-> 4 print(“The minimum number of introductions is” + str(len(min_intros)) + " and the maximum is " + str(len(max_intros))) NameError: name ‘min_intros’ is not defined

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

———————————————————————————NameError Traceback (most recent call last) in 2 print(str(len(specific_smalls)) + " are lineages which were sampled less than five times in " + country + “, and so have been left out of visualisation in the interests of clarity”) 3 else: —-> 4 print(str(smalls_count) + " are lineages which only contained five sequences or fewer, and so have been left out of visualisation in the interests of clarity“) NameError: name ‘smalls_count’ is not defined

Furthermore, those sequences which haven’t been sampled in the last month are not shown.

———————————————————————————NameError Traceback (most recent call last) in 2 status_counts, reactivated_lineages, continuining_lineages = lin_exp.describe_lineages(specific_bigs) 3 else: —-> 4 status_counts, reactivated_lineages, continuining_lineages = lin_exp.describe_lineages(intro_bigs) 5 6 reactivateds = status_counts[“Reactivated”] NameError: name ‘intro_bigs’ is not defined

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.

———————————————————————————NameError Traceback (most recent call last) in —-> 1 print(“The global lineages are correct as of the data release on” + lineage_version) NameError: name ‘lineage_version’ is not defined

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”

———————————————————————————NameError Traceback (most recent call last) in 4 5 else: —-> 6 intro_country_counts, intro_country_percentages, intro_country_together = descrip.prep_dicts(intro_countries) 7 dataframe, tree_order = descrip.make_dataframe(intro_country_together, intro_object_dict) 8 NameError: name ‘intro_countries’ is not defined

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.

```
---------------------------------------------------------------------------NameError
Traceback (most recent call last)<ipython-input-1-bf373ae9d8a3> in
<module>
      1 #df_counts, df_thinned, df_acctrans_counts =
dp.make_plotting_dfs(intro_country_counts, intro_object_dict) #these
are never actually used any more
      2
```

```
----> 3 dp.plot_bars(intro_bigs, country, sequencing_centre)
NameError: name 'intro_bigs' is not defined
```

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.

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.

———————————————————————————NameError Traceback (most recent call last) in —-> 1 dp.make_timeline(intro_bigs, sequencing_centre, country) 2 timeline_df = dp.raw_data_timeline(intro_bigs) NameError: name 'intro_bigs' is not defined

The date of first sequence in the cluster sampled by GSTT is shown in figure five for every cluster with date information. The date of first sequence in the cluster sampled by a pillar 2 lab is shown in figure five for every cluster with date information.

```
---------------------------------------------------------------------------NameError
Traceback (most recent call last)<ipython-input-1-081063bf309a> in
<module>
      3     starts_raw = dp.raw_data_starts(single, multi)
      4 else:
----> 5     multi, single = dp.plot_starts(intro_alls)
      6     starts_raw = dp.raw_data_starts(single, multi)
NameError: name 'intro_alls' is not defined
```

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.

```
---------------------------------------------------------------------------NameError
Traceback (most recent call last)<ipython-input-1-68a455547428> in
<module>
      3     raw_seqs_over_time =
dp.raw_data_seqs_over_time(date_counts)
      4 else:
----> 5     date_counts = dp.plot_sequences_over_time(taxa, country,
sequencing_centre)
      6     raw_seqs_over_time =
dp.raw_data_seqs_over_time(date_counts)
NameError: name 'taxa' is not defined
```

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.

```
---------------------------------------------------------------------------NameError
Traceback (most recent call last)<ipython-input-1-6eef44d7511d> in
<module>
----> 1 map_output = map.make_map(input_geojsons, adm2_cleaning_file,
metadata_file, summary_output, week, sequencing_centre, country,
pillar2)
NameError: name 'country' is not defined
```

```
---------------------------------------------------------------------------NameError
Traceback (most recent call last)<ipython-input-1-9abc83bdc0dd> in
<module>
----> 1 if type(map_output) != bool:
      2     new_uncleans, mapping_data = map_output
      3     no_seqs = False
      4 else:
      5     no_seqs = map_output
NameError: name 'map_output' is not defined
```

——————————————————————————NameError Traceback (most recent call last) in —-> 1 if not no_seqs: 2 if new_uncleans: 3 print("There are some sequences with locations that are not matched to real Admin2 regions, some manual curation required.") NameError: name 'no_seqs' is not defined

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.

——————————————————————————NameError Traceback (most recent call last) in —-> 1 print(dataframe.to_markdown(tablefmt='grid')) NameError: name 'dataframe' is not defined

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

Table S2 is not appropriate for this report and so has been omitted.

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

Table S3 is not appropriate for this report and so has been omitted.

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

——————————————————————————NameError Traceback (most recent call last) in —-> 1 print(starts_raw.to_markdown()) NameError: name 'starts_raw' is not defined

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

——————————————————————————NameError Traceback (most recent call last) in —-> 1 print(raw_seqs_over_time.to_markdown()) NameError: name 'raw_seqs_over_time' is not defined

Table S7 Raw data for the figure seven with the number of sequences assigned to each admin2 region.

——————————————————————————NameError Traceback (most recent call last) in —-> 1 if not no_seqs: 2 print(mapping_data.to_markdown()) NameError: name 'no_seqs' is not defined

```
---------------------------------------------------------------------------NameError
Traceback (most recent call last)<ipython-input-1-c2b516fe2325> in
<module>
----> 1 writing.write_summary_files(summary_output, dataframe,
omitted, week, intro_alls, timeline_df)
NameError: name 'dataframe' is not defined
```