MADDOG results

MADDOG can take a while to run. Instead, we'll stop the run by pressing ctrl + z and use the output from an earlier run I completed on this dataset!

Task 1

Go to the `RAGE-workshop-2024/5-phylogenetics_and_maddog/C-tutorial_maddog` folder. You'll see a folder called `Nigeria`. Open this.

You'll see all the new folders and files produced by the MADDOG run!

Start by looking in `Report/report.html`

This summarises all the MADDOG outputs in a html file.

Read through the file and answer the following questions: 1. How many lineages are there? How does this compare to our answer with assignment? Why is it different? 2. Which lineage was first detected from this dataset? 3. How have the lineage dynamics changed over time? 4. What are the 2 most prevalent lineages in the dataset? How many sequences do they each have?

Task 2

We'll now look at Figure 2.

This is called a sunburst plot. It's a way of showing how lineages are related, and how many sequences are in each lineage. We start in the middle with the broadest ancestor of all the lineages, and each circle moving outwards represents a lineage descended from the one inside it.

E.g. Lineage Africa_A1.1.1 is descended from Africa_A1.1, which is descended from Africa_A1 and so on.

The plot is interactive! You can hover your mouse over a section to reveal the number of sequences, and click on a section to zoom in on it. To zoom back out, keep clicking on the innermost circle.

E.g. If you click on Africa_A1.1, it will zoom in to only include this lineage and those descended from it. Click the Africa_A1.1 circle to zoom out, and then keep clicking the inner circle until Africa is the inner circle again.

By exploring the sunburst plot, try to answer the following questions:

1. Which lineage is Africa-2_A1 descended from?
2. How many lineages are *direct* descendants of Africa_A1 (e.g. only 'children', not further descendants)
3. How many sequences from this dataset are from lineage Africa-2_A1.1?
4. How many sequences from this dataset are from lineage Africa-2_A1? Why is it included?
5. How many 'levels' of lineages from Africa to Africa-2_A1.1?
6. If this dataset represents everything we know about lineages in a region, what might some discussion points be about lineage diversity in the region?

Task 3

Although the summary is very useful, MADDOG also produces other outputs. Some of which we can use for further analsyis.

The most important of these is `/Outputs/sequence_data.csv`

Open it in excel so we can sort and view the data more easily.

Have a look at this file and answer the following questions:

1. What lineage is sequence 4 from?

2. Which sequences are most closely related to sequence 8? Could we tell this from GLUE? Could we tell this from our previous assignment output? Why/why not?

3. Given all of the outputs you've looked at here - what further analysis would you do? What further figures would be useful to produce?

4. How do you think detailed lineage information could help with your own research? Try to formulate some research questions that could be addressed using your own current/planned sequencing activities and lineage designation!