Tree Annotation - Standard

4.2.1b.6: Extracting metadata

We can add additional information to our tree. This requires a well formatted metadata file.

We'll be using the file `nig-af2-seqs-metadata.csv`. You can find this in `RAGE-workshop-2024/5-phylogenetics_and_madd` Where you found the sequences we used to build the tree.

We'll need to edit this a bit first to make it a format that FigTree understands! We might want to assess the year, state or host of the sequences.

Task 1

Open your annotations file in excel. Just do this by clicking on the file in your finder window (NOT THE COMMAND LINE). You'll see we have 6 columns. Start by deleting the `state_code` and `sample_collection_date` columns - we won't be using those here.

The `sample_collection_decimal_date` column is really useful for some analysis, but for now we just want the exact year (with no decimals!).

Can you figure out how to change this in excel? Give the column a more appropriate heading now we've changed the data!

FigTree only accepts annotations in a .txt file, so you'll need to save this edited metadata file as `nig-af2-seqs-metadata.txt`.

Task 2

Import the annotations file in to FigTree. We do this by going to File->Import annotations and selecting the txt file we just looked at.

We can now colour our tips by state! In the side tab, go to Tip labels and colour by state.

Task 3

You can also colour branches according to your annotations. Go to Tree->Annotate Nodes from Tips and select `state` as the annotation.

Now, if you go to the Appearance tab, you can colour by state.

Task 4

Again, play around until you have a tree you're happy with.

When you've finalised your tree, you can export it by File->Export X where X is the file type.

Task 4

Now try to annotate your tree in FigTree with tips coloured by host.

Task 5

Try to annotate your tree to have tip labels displaying the host type, coloured by year, and branches coloured by state. Add a legend to show the states.

From this tree - where is the human case from? Where did the cluster of sequences that the human case belongs to originate from? _____

Task 6

From this tree - where does the outbreak appear to have started from?

Does this change if the tree is midpoint rooted or not? Why? What are the implications of this?

What could we do to overcome this issue and identify where the outbreak started?