Hi, I'm Muthumeenakshi. I'll be guiding you how to construct a phylogenetic tree today. In this video, you'll be looking into the aspects of how to build a phylogenetic tree using the software IQ-TREE.

Now, in order to build a phylogenetic tree, we should have a multiple sequence alignment file. Now, this is basically done using SARS-CoV-2 genomes. We have already provided you the metadata and the SARS-CoV-2 genomes in a Fasta format. So using the SARS-CoV-2 genomes in the Fasta format, we'll take that as an input to first construct multiple sequence alignment.

And multiple sequence alignment today I'll be showing you using the MAFFT tool. So click on Google, and type MAFFT, M-A-F-F-T, which is a multiple sequence alignment programme. You click on the first link that opens, and there's a place for input where you have to choose the file.

So as I already mentioned, go and select the sars_cov2_genome.fasta. So this is the input file for multiple sequence alignment. Click on Open. So now you have given the input data for running the multiple sequence alignment. Keep all the other parameters default, and then scroll down, and give the option of Submit.

This will take a few minutes. You can see the command over here, like how the multiple sequence alignment has been running. Please wait for a while until your multiple sequence aligned file is projected in the window.

Now that we have got the results of multiple sequence alignment of our input data, so here we can see, with respect to the reference, the other sequences have also been aligned. You can view your multiple sequence aligned file by scrolling down. You can also download the same file where you have different formats for downloading.

One is a Clustal format. Another is a Fasta format. So you can click on the Fasta format, which downloads the multiple sequence aligned file dot fasta, and this will be your input file for your IQ-TREE builder.

We will start building the phylogenetic tree using IQ-TREE. Click on IQ-TREE in Google, and select the second option, which is IQ-TREE web server, fast, accurate phylogenetic trees. Here you have an option called alignment file. Browse. Go to the Downloads where we have the aligned file, multiple sequence aligned file.

Click on Open. Scroll down. Keep all the options defaulting. So here we are giving Submit Job. The job is being submitted. You can click on OK. Here on the left-hand side panel you can see the status of your job that has been submitted. You can see whether it is waiting or whether it is progressing. Just wait for a while until your job is completely run.

So the status-- we saw it was waiting, and you can also reload. You can refresh the web page, and then it will come as Success here. So here, the job which we have given for constructing the phylogenetic tree is completed. You can go to the option called Download Selected Jobs, where your present job will be downloaded. And this gets downloaded in a zip format.

So if you open this, you will see this folder containing four different files. One is a contree. The second as IQ-TREE. The third one is fasta.log, and fourth one is .treefile. You're going to rename this a.treefile into Newick format, tree.nwk. And this tree.nwk will be the input for Microreact. Nwk is the obligation for Newick.

So now that we have got the phylogenetic tree that has been constructed, we are going to view the same phylogenetic tree along with the metadata in microreact.org. This microreact.org is an online platform in order to view our phylogenetic tree and the metadata. The metadata looks something like this. It will always be in .csv format, preferably in Excel.

So here you can see there are so many columns and rows. The first-- is always the ID, which gives the sample number, and this is the same as that that was-- these samples are the same that was given as an input in multiple sequence alignment file. You can also view the collection date, the year, the month, day, the location, and the country, the state also from where the sample has been procured.

In order to give a geographical location to the particular data, we have also provided the latitude and longitude for the same set of samples. Coming to the patient characteristics, we see here that the samples are being differentiated with respect to the gender. They have also given the patient age.

Along with that we have included something called as a Nextclade and Lineage. You can also provide other inputs like how there are different substitutions for a particular sample based on your genomic analysis as well in your metadata, which will again give a different projection to your particular data in Microreact, so moving on to microreact.org.

If you already have an account in Microreact, you can upload your files through your account, or else you can create an account. That can also be done. Else, there is always an option called here as an upload. You can drag and drop your files from your system, or you can add using this plus sign. So here, in plus sign, go to Browse Files, and as I've already mentioned before, we are going to upload the metadata.csv as well as the phylogenetic tree that has been constructed.

Select the files, and click on Open. You can see here metadata is being uploaded in .CSV format, and the phylogenetic tree that was constructed using IQ-TREE have been uploaded in Newick format. Click on Continue. And here it asks you to specify the ID column, so again click on Continue. See, you just have to specify which column your ID belongs to. It is the first column, so we are again keeping it as default. We are clicking on Continue.

So we can view here three different panels. The left side panel gives the geographical location of the data that has been distributed, and the right side is a phylogenetic tree. And down we have the timeline of the data that has been distributed from different time points.

You can see the eye tool here. Just click on that. And you can see the different labels columns, coloured columns. So based on the colour column, you can select whichever you want to-- projecting in different colours. To start with, we are going to select the Nextclade.

So this gives you two different colours. One is in green. Another is an yellow. Go to the legend option here on the right side. So the green colour indicates it belongs to the variant called 21K Omicron, whereas yellow belongs to 21L.

The same legend which has been projected towards the left side on the geographical side where you can see the green colour is coming from different community sectors that has been distributed in New Delhi. Where there's the yellow that has been coloured has come from a single family where there was a major outbreak that was observed.

Further ahead, if you want to add on more to your particular data, you click on this particular option, and you can add on the metadata blocks. So the search columns-- whichever you want to select you can select. And here I'm going to select the state and gender. I'm going to select the patient age as well.

So if you can see on here, the same has been selected. Go on to the legend, and again see which colour belongs to what. So colours by state-- it is Delhi. Colours by gender-- female is green, and male is yellow. And we also see the Colours by patient age.

So if you click on to one particular colour here, we can see the Colours by Nextclade, which is 21L Omicron, which belongs to the state Delhi. And this has been from a male gender, and the approximate age is 39. So based on this, you can, again, play around, or you can add how many ever more data you want to.