# AMD Midwest NEWSLETTER

March 2020

IN THIS ISSUE – We've established we're related…but how related?

This month's issue is going to expand on the basics of phylogenetics we established in January's issue (find it here if you missed it: https://staph-b.github.io/midwest-region/newsletters/2020-01_newsletter.pdf) and examine why a phylogenetic tree appears to change through the addition of more isolates.

Evolving phylogenetic trees

Have you ever looked at a phylogenetic tree at the start of an outbreak and seen two isolates that look like they might be related, but by the end of an outbreak they are on two different branches and now appear unrelated? Which tree is correct? The short answer is both trees are correct! Using the figure below, we'll explain why.
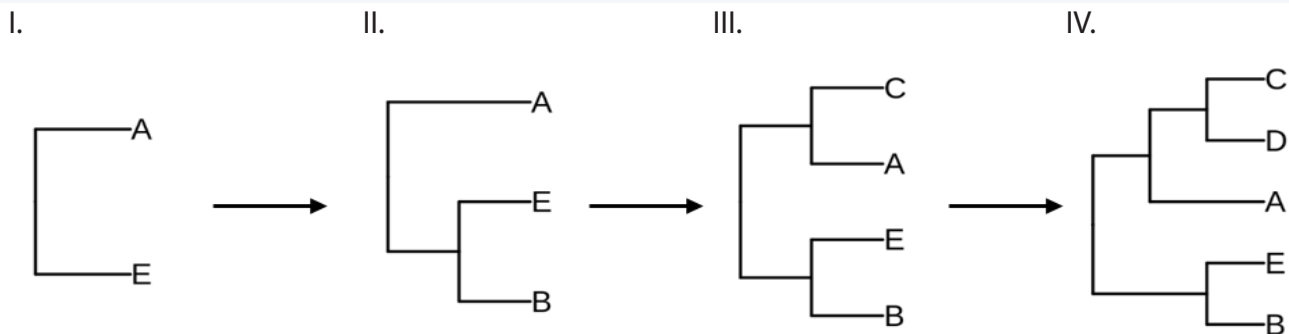
I. II. III. IV.

Future Events

Let's imagine the two samples above (A and E) are from an ongoing outbreak.

I. If we have only these two samples to examine, they are going to appear directly related. Additionally, if we have only these two samples, along with a dozen unrelated isolates of the same species, then A and E will be most closely related to each other and found clustered together on the tree.

II. If we add in a third sample (B), which is more closely related to E, the tree has now split into two different clades with our original samples no longer clustered together.

III & IV. As we add in the last two samples (C and D), the final tree has now grown and changed to include other isolates that are more closely related to one another than our original isolates.

In summary: The addition of isolates that are related will cause new clades to appear and develop. This does not mean that the original isolates were not related or that the original tree clustered the isolates incorrectly, but that there are other isolates more closely related that were added to the analysis.

Bootstrapping time - Let's add statistics to our phylogenetic trees

When looking at a phylogenetic tree, how can we be confident in a cluster that is present? Through the use of bootstrap statistics, we can calculate a statistical percentage of how well the placement of each branch on the tree is supported. Bootstrapping is performed by random resampling of aligned sequence data. The percentage is then calculated to reflect the percentage of times the resampled alignment cluster the isolates together on the tree. A boostrap value is calculated for every node on the tree (pictured below).
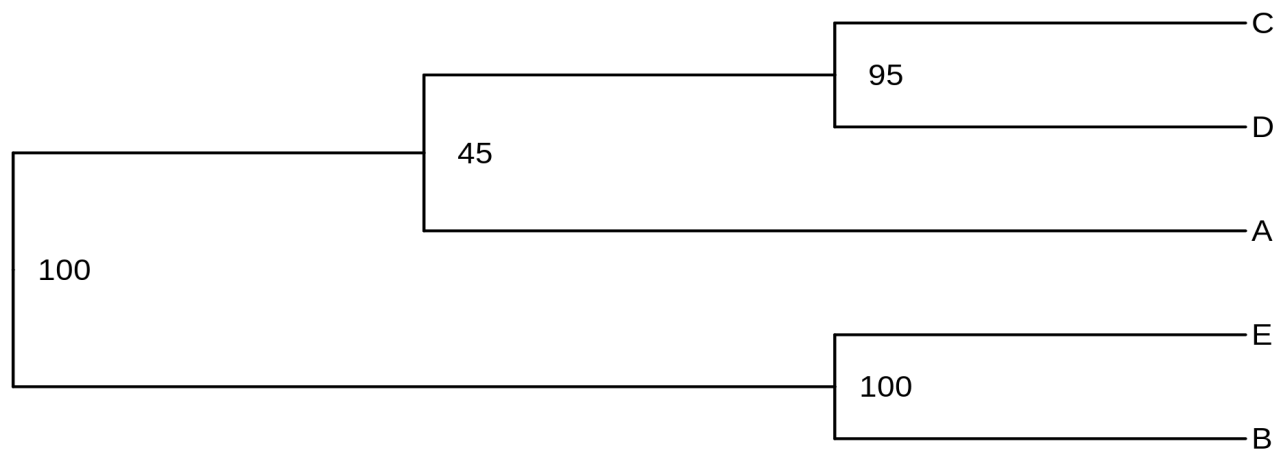
In the example above, samples E and B cluster together 100% of the time the alignment was resampled, so we can be very confident those two isolates should be placed next to one another on the tree. Conversely, sample A was found clustered with samples C and D in only 45% of resamplings, so we are less confident about its placement on the tree. A bootsrap value of 100 is ideal, but there is no hard fast rule for determining a good bootstrap value. Research and academia lower the value as low as 50-70% for identification of clusters to perform molecular epidemiology studies.