**Exercise 3: Phylogenetic tree reconstruction**

EMD 531, Spring 2020

The purpose of this exercise is to learn how reconstruct, visualize, and annotate phylogentic trees.

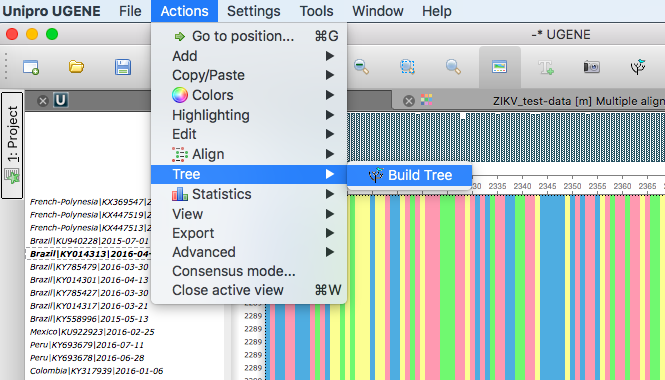
***\*\*Important***: We will practice in class using our test dataset, but you will need to complete the following with your own dataset to complete the exercise.

**This exercise requires UGene:** <http://ugene.net/download.html>

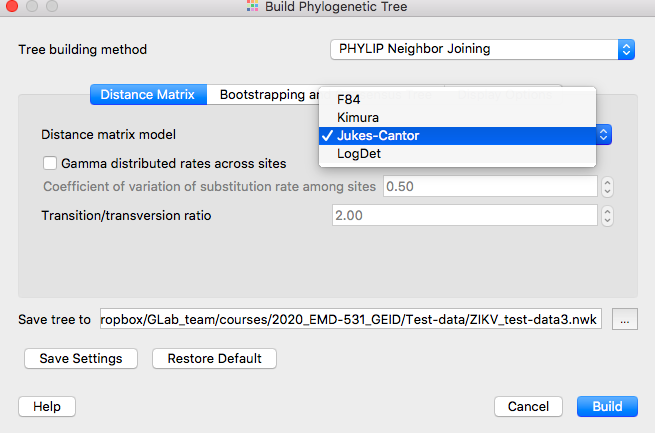
**Overview:**

**Step 1: Neighbor-joining (NJ) tree reconstruction**

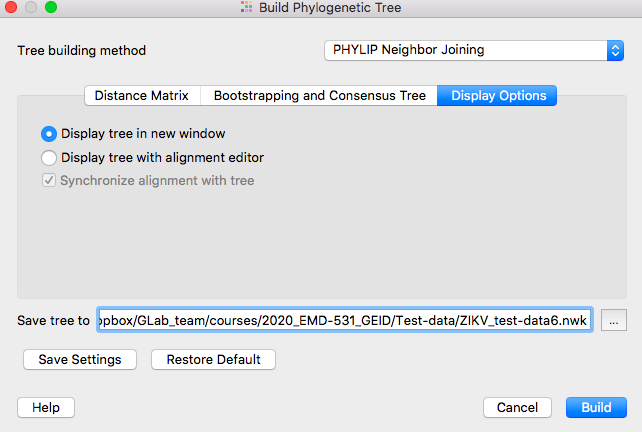
* *The NJ process outlined here is a quick check to examine general relatedness. It shouldn’t be used for final analysis.*
* To build a NJ tree:
  + Open your aligned and trimmed sequences
  + Select:
    - **Actions > Tree > Build Tree** (or select the ‘tree w/ bird icon’ on the top bar)



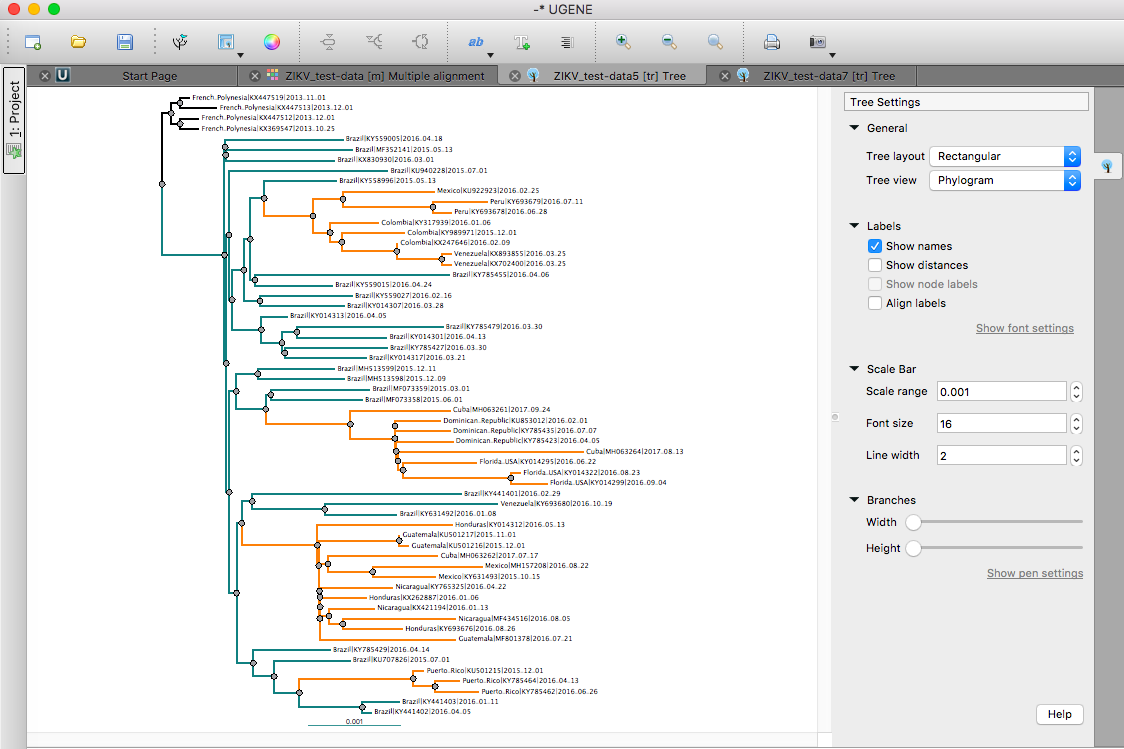
* + For the Tree building method, select **PHYLIP Neighbor Joining**
  + An important consideration for NJ phylogenetic reconstruction is the distance matrix model
    - For this first tree, select the simplest model, **Jukes Cantor**

****

* + Go to the **Display Options** tab and select **Display tree in new window**

****

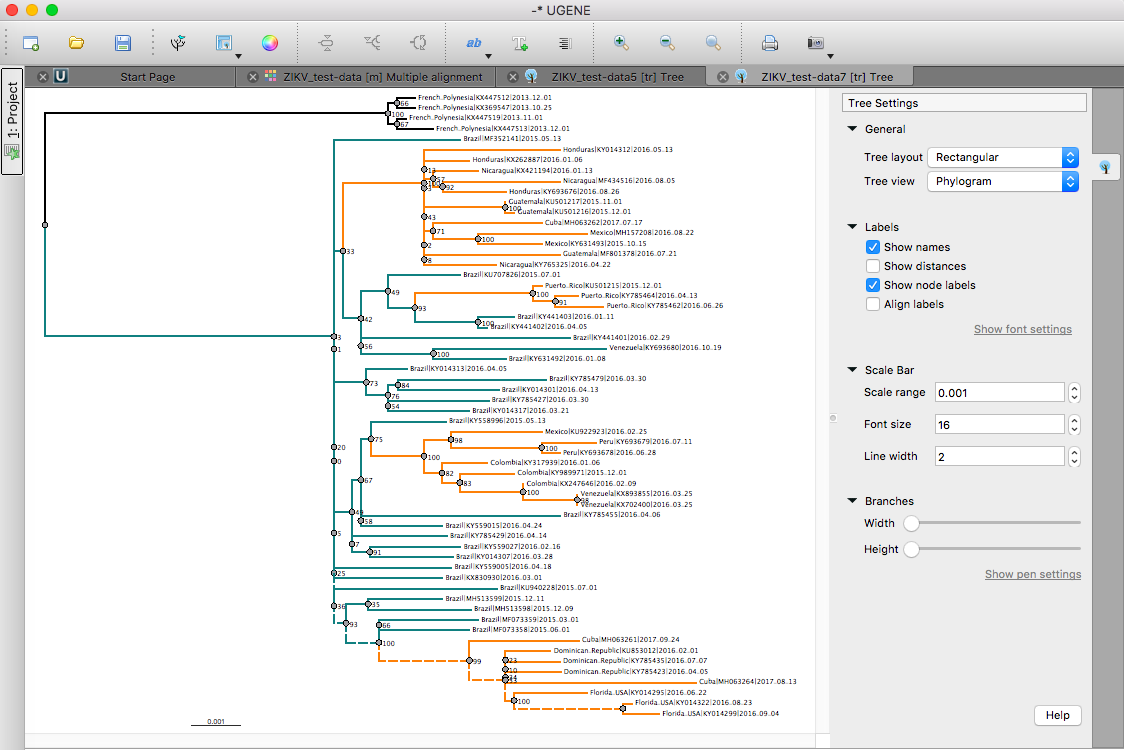
* + Push **Build** without selecting any other options (we won’t do any bootstrap analysis for this tree)
    - This should only take a few seconds to build.
* Use the **Tree Settings** tool bar on the right and annotation toolbar on top to:
  + Change Tree view to **Phylogram**
  + Root the tree by selecting ancestral clade or out-group (click on node) and selecting:
    - **Actions > Reroot tree**
  + Color branches according to your epidemiological question by selecting the node and selecting the ‘color wheel’.
    - In the example below, the branches are colored black for outside of the Americas, teal for Brazil, and orange for the rest of the Americas.
  + Change fonts, line widths, align labels, etc according to preference. Make sure that the tree and labels are readable.

****

* Paste a screenshot to answer question **E3-1** below.

**Step 2: Maximum-likelihood tree reconstruction**

* The ML process here will provide an accurate view the relatedness among your sequences, but will take a lot longer to build than the NJ tree.
* To build a ML tree:
  + Open your aligned and trimmed sequences
  + Select:
    - **Actions > Tree > Build Tree** (or select the ‘tree w/ bird icon’ on the top bar)
  + Tree building method: select **PhyML Maximum Likelihood**
  + Substitution model: select **GTR**
    - *This is an important consideration. GTR is the most complex model, and also takes the most time to run.*
  + Branch support: select **perform bootstrap** and set to **100** replicates
  + Display options: select **Display tree in new window**
  + Press **Build**
  + *This may take several minutes to hours… do not turn your computer off.*
* Use the **Tree Settings** tool bar on the right and annotation toolbar on top to:
  + Change Tree view to **Phylogram**
  + Under Labels, make sure the **Show node labels** is selected.
  + Root the tree by selecting ancestral clade or out-group (click on node) and selecting:
    - **Actions > Reroot tree**
  + Color branches according to your epidemiological question by selecting the node and selecting the ‘color wheel’.
  + Change fonts, line widths, align labels, etc according to preference. Make sure that the tree and labels are readable.

****

* Paste a screenshot to answer question **E3-1** below.
* A version of the tree above can be visualized by opening the

**Questions:**

*Combine answers with all exercises,* ***due before class 16 on March 5th****. Be prepared to discuss your answers during class 16.*

**Rerun the steps about using your data, not the ZIKV test. Then answer the following questions.**

**E3-1**: Paste images of your simple NJ and ML (with bootstrap support) trees made from your data (i.e. not the ZIKV test data). Label each as NJ or ML, and make sure that the branches are colored according to address your epidemiological question (color each in the same format). Explain how you colored the branches. (9 points)

**E2-2:** How did you decide to root your trees? Briefly explain what aspects of your NJ and ML trees are similar and different. (Max 200 words; 3 points)

**E3-3**: Do you find any ‘polytomies’, nodes with more than two descendent lineages (often with low bootstrap support), creating a "pitchfork"? Several polytomies can be seen in the ML tree with the ZIKV test data. How may these impact your findings and what could you do to resolve the issue? [*Note: if you do not have any polytomies in your trees, answer based on the test data tree*] (Max 200 words; 3 points)

**E3-4**: While these phylogenetic reconstructions may not yet completely answer your epidemiological question, briefly explain what they have revealed. What additional information do you need to answer your question in full? (Max 300 words; 5 points)