**Dear all,**

So far this is my current status.

**Simulated Data**

1. **PIP-** data from Hossein. Parameters: Taxa=8, branch length= random, tree rate =8, E= 200, I= 0.5. lambda= 10, mu= 0.05
2. **INDELible**

Data generated by setting same parameter values from PIP. We set same E, lambda values from PIP, we also set lambda=mu.

Taxa= 8, L=200 (E at root is L+num\*), lambda=mu=0.05, INDEL-distribution POW 1.7, same tree from PIP.

**Real Data**

1. **Protein dataset**

We are using data provided by Maria. For Trees, we used the trees generated by PRANK for corresponding sequences. Each dataset consists of 13 taxa.

**Alignment Stage:**

1. **ProPIP-** Requirements (Model parameters (lambda and mu), Tree, Sequence)

**Generating lambda and mu, given TRUE MSA and TREE using ProPIP**

**lambda\_mu\_generator for INDELible🡪** Max provided 3 matlab files to produce lambda and mu values using ProPIP given TRUE MSA and TREE. The ProPIP generated warnings for the MSAs with completely blank columns. However, at the end these warning are discarded, and the corresponding lambda and mu are generated for each input. The model parameters are calculated with the algorithm called Levenberg-marquardt. This method (generate model parameters given tree and MSA) is applied to all the data that I mentioned above.

The corresponding lambda and mu values are then used to generate params files, which we then used as an input to the ProPIP.

Since cluster is running full, currently I am running the alignment process in my system. I tested the 13 sequences from INDELible separately and ProPIP is not able to produce the MSA. Therefore, we exclude these 13 samples from INDELible. Moreover, we have 100 samples from PIP, 4 samples from real data. All alignments are running without error so far.

1. **PRANK**

From INDELible currently we have 87 samples (13 exceptional sequences), 100 samples from PIP, 4 samples from real data. All sequence alignments are completed with this tool without error.

1. **MAFFT**

The ‘ **;** ’at the end of the newick file was the error that we were discussing before and I corrected it and run the aligner successfully. I tested the 13 samples from INDELible separately and MAFFT is able to produce the MSA (tool consider the input sequences as it is). But these samples are not taken for ProPIP and PRANK therefore we do the same for MAFFT. All sequence alignments are completed with this tool without error.

Since INDELible data that we generated is working with all the Aligners, I am moving forward with this tool, if everyone agrees.

I will update everyone again once alignments from ProPIP is done.

Best regards,

Eldhose