**REAL DATA ANALYSIS**

**To Do:**

1. 4 Data sets (13 taxa on each data set)
2. Generate Guide Tree in PRANK
3. Use tree from PRANK in ProPIP to align the MSA (use Un-gapped sequence)
4. Align using PRANK (use Un-gapped sequence)
5. Align using MAFFT
6. Compare Inferred MSA from ProPIP with TRUE MSA from PIP ???
7. Compare Inferred MSA from PRANK with TRUE MSA from PIP ???
8. Compare Inferred MSA from MAFFT with TRUE MSA from PIP ???
9. Compare Inferred MSA from ProPIP with TRUE MSA from INDELible ???
10. Compare Inferred MSA from PRANK with TRUE MSA from INDELible ???
11. Compare Inferred MSA from MAFFT with TRUE MSA from INDELible ???

**Dear all,**

So far this is my current status.

**Simulated Data**

1. **PIP-** As per Hossein’s parameters, I received 100 replicates with TRUE MSA and Tree for each replicate.
2. **INDELible-** Simulated 100 replicates with same mu rate as PIP (indel distribution🡪 POW 1.7), with same tree provided by PIP. For simulating I wrote a code in R Language, which can control all the parameters and inputs. This R file will produce a control file (control.txt). Which is then used as an input to INDELible binary file.

**Real Data**

1. **Protein Alignment sequences**