**Motivation**

- Phylogenetic software is a vital tool that helps biologists analyze evolutionary history.
- Underlying Markov Chain Monte Carlo (MCMC) process is treated like a black box.
- Popular software like MrBayes are known to converge slowly and even inaccurately.

**Research Question**

- How well does MrBayes perform under noisy input?
- What are the characteristics of tree topology and sequence generation that cause poor convergence?

**Methodology**

1. Generate ground truth tree topology and DNA sequences with standard graph traversals.
2. Input leaf data into MrBayes and run until convergence.
3. Obtain the number of iterations to converge and the tree distance between the most probable output tree and the ground truth.

**Background**

- **Phylogeny**: Studying evolutionary relationships of species.
- **Markov Chain**: Graph of states with probabilistic edges.
- **MCMC**: Performs a random walk until converging to the stationary distribution.
- **Stationary Distribution**: Huge space of all phylogenetic trees weighted by their likelihood.
- **Metropolis Coupled MCMC (MC³)**: Optimization on the normal MCMC algorithm for faster convergence.

**MrBayes**

- Input aligned DNA sequence for each species and outputs a sample from the posterior distribution (proportional to the likelihood).
- Default convergence heuristic is the average standard deviation of split frequency (ASDSF).
- Parameters include sequence length, number of species, branch lengths, and substitution model.

**Conclusion**

- Converges accurately even with noisy data.
- Exponential convergence on some mixed tree topologies, which real world data may resemble.
- Exponential convergence requires specific conditions which are uncommon.
- MC³ converges in fewer iterations and should generally be used over single chain MCMC.

**Citations**


**Acknowledgements**

We would like to thank Professor Eric Vigoda, Professor Diba Mirza, Dr. Soojin Yi, Dr. Todd Oakley, and Chinmay Sonar for their support and guidance.