

MAYANK KALE

PARALLELIZATION OF MAXIMUM LIKELIHOOD

MOTIVATION

- ▶ To analyze large amount of data using computationally intensive, model-based optimality criterion such as maximum likelihood (ML), there must be development in parallel and distributed algorithms for ML analysis.

MAXIMUM LIKELIHOOD

- ▶ Maximum likelihood is a general statistical method for estimating unknown parameters of a probability model.
- ▶ In phylogenetics there are many parameters, including rates, differential transformation costs, and, most important, the tree itself.
- ▶ Likelihood is defined to be a quantity proportional to the probability of observing the data given the model, $P(D|M)$.

ALGORITHMS

- ▶ GAML Genetic Algorithm
- ▶ GARLI Genetic Algorithm
- ▶ fastDNAmL

BIBLIOGRAPHY

- ▶ Matthew J. Brauer, Mark T. Holder, Laurie A. Dries, Derrick J. Zwickl, Paul O. Lewis, David M. Hillis; Genetic Algorithms and Parallel Processing in Maximum-Likelihood Phylogeny Inference. *Mol Biol Evol* 2002; 19 (10): 1717-1726. doi: 10.1093/oxfordjournals.molbev.a003994
- ▶ Zwickl, D. J. (1970, January 01). Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Retrieved April 18, 2017, from <http://hdl.handle.net/2152/2666>
- ▶ Gary J. Olsen, Hideo Matsuda, Ray Hagstrom, Ross Overbeek; fastDNAm1: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Bioinformatics* 1994; 10 (1): 41-48. doi: 10.1093/bioinformatics/10.1.41