Vacharapat Mettanant 2009: A Linear – Time Algorithm for the Multiple Gene Duplication Problem. Master of Engineering (Computer Engineering), Major Field: Computer Engineering, Department of Computer Engineering. Thesis Advisor: Assistant Professor Jittat Fackcheroenphol, Ph.D. 36 pages.

Reconstructing the Tree of Life is one of the most important research goals in evolutionary biology. In this thesis, we consider one of the issues regarding inconsistency among the trees reconstructed using sequences from different genes, called gene trees. This inconsistency is often a result of gene duplication. Given a species tree and a set of gene trees, the Multiple Gene Duplication Problem as formulated by Guigo, Muchnik, and Smith asks for an assignment of gene duplication events on the species tree node to minimize the number of duplication episodes. Recently, Bansal and Eulenstein give the first efficient algorithm for solving that problem. Their algorithm runs in quadratic time. In this thesis we give the first linear-time algorithm for the problem. Our algorithm can be seen as a faster implementation of Bansal and Eulenstein's method. We also present a more practical algorithm that runs in near-linear time. Experiments on synthetic data show a very good improvement on the running time (at least twice faster).

Student's signature Thesis Advisor's signature ___ / ___ / ____