

Algorithms for Statistical Multiple Alignment

Jotun Hein

Department of Statistics

University College

1 South Parks Road

Oxford OX1 3TG

UNITED KINGDOM

Abstract

Sequence analysis has during the last decade experience an increased emphasis on statistics. Unfortunately, the alignment problem has lagged seriously behind, but this is rapidly changing. In this talk algorithms are presented that allows the calculation of the probability of a set of sequences related by a binary tree that have evolved according to the Thorne-Kishino-Felsenstein model for a fixed set of parameters. The recursions are based on a Markov chain generating sequences and their alignment at nodes in a tree. Dependent on whether the complete realization of this Markov chain is decomposed into the first transition and the rest of the realization or the last transition and the first part of the realization, two kinds of recursions are obtained that are computationally similar, but probabilistically different. The running time of the algorithms are L to the k 'th power, where L is the length of the observed sequences and k is the number of sequences - leaves at the binary tree. If time permits a Gibbs sampling approach and extension of the basic model will be discussed. Further developments of this alignment approach could allow for better homology testing, a combination with hmm-based genome annotation approaches.