Accuracy Estimation of Protein Multiple Sequence Alignments Using Profile Hidden Markov Models

Liu, Joyce (School: Sentinel High School)

Multiple sequence alignments, alignments of three or more biological sequences play a critical role in phylogenetic tree development and identification of evolutionarily conserved regions. Evaluating the accuracy of multiple sequence alignments is critical for not only selecting the best alignment software and parameters but also vital towards an accurate understanding of evolutionary processes. Profile hidden Markov models are statistical models that determine the probability of a specific amino acid or nucleotide using a position-specific scoring system. In this study, we assess the utility of profile hidden Markov models to evaluate the internal consistency of multiple sequence alignments. We propose that the same measure of internal consistency can be used to evaluate the accuracy of a multiple sequence alignment. Our results demonstrate that profile HMMs independently do not provide an accurate method of accuracy estimate. However, profile HMMs may be useful in comparing sequence alignments if improved upon.