Gregory Nuel MAP5 (CNRS – Université Paris-Descartes)

:ppALIGN: posterior probabilities for score based alignments

Résumé :

Score-based pairwise alignments are widely used in bioinformatics in particular in molecular database search tools, like the BLAST family. Such algorithms are usually fast due to sophisticated heuristics, but unfortunately the underlying scoring model lacks in a statistical description of the reliability of the reported alignments. In particular close to gaps, in low-score regions or in low-complexity regions (exhibiting repetitive pattern) of the alignment a huge number alternative alignments arise which leads the certainty about the optimal alignment to decrease.

ppALIGN is a software package that uses hidden Markov model techniques to compute position-wise reliability of score based pairwise alignments of DNA or protein sequences. The design of the model allows for a direct connection between the scoring function and parameters of the probabilistic model. For this reason it is suitable to directly analyze the outcome of popular score based aligners and search tools without the need to choose a complicated set of parameters. It only requires the classical score parameters (the score matrix and gap costs). The package comes along with a library written in C++, a standalone program for a single user defined alignment (ppALIGN) and another one (ppBLAST) that treats a complete result set of BLAST. The main algorithms essentially exhibit a linear time complexity (in the alignment lengths), and they are hence suitable for on-line computation. We have included alternative decoding algorithms to provide alternative alignments.

ppALIGN is a fast program/library that helps to detect and quantify questionable regions in pairwise alignments. Due to its structure, the input and output interface it is suitable to connect with other post-processing tools. Empirically, we illustrate its usefulness in terms of correctly predicted reliable regions for sequences generated using the ROSE model for sequence evolution.