

# Dynamic Programming Algorithms for Biobjective Sequence Alignment

Maryam Abbasi<sup>1</sup>, Luís Paquete<sup>2</sup>, Miguel Pinheiro<sup>3</sup>

<sup>1</sup> CISUC, Department of Informatics Engineering, University of Coimbra, Portugal, [maryam@dei.uc.pt](mailto:maryam@dei.uc.pt);

<sup>2</sup> CISUC, Department of Informatics Engineering, University of Coimbra, Portugal, [paquete@dei.uc.pt](mailto:paquete@dei.uc.pt);

<sup>3</sup> Biocant, Biotechnology Innovation Center, Cantanhede, Portugal, [Monsanto@biocant.pt](mailto:Monsanto@biocant.pt);

## Abstract

In this work, the multiobjective formulation of the pairwise sequence alignment problem is considered, where a score vector function takes into account simultaneously two components, similarity score and indels or gaps. Similarity score for amino acids and nucleotide sequences is obtained by a substitution matrix, such as BLOSUM or PAM, and by the subtraction of the number of matches and mismatches, respectively. This formulation is of interest to the practitioner since it allows to get rid of parameters in the usual single weighted-sum objective function and to explore a tractable set of alignments that are not reachable by any other methods [1]. In this work, explicit recurrence equations for the dynamic programming algorithms are given as well as a new pruning technique that improves the run-time. Furthermore, the algorithm is applied to 7 PAP genes of *Candida* clade [2] and the results are compared with those of a distance tree, found by simulating the divergence of these sequences from a common ancestor. Similar conclusions were achieved by the two methods. Extensions for multiple sequence alignment are also discussed.

## References:

1. M. A. Roytberg, M. N. Semionenkov, and O. Yu. Tabolina. Pareto-optimal alignment of biological sequences. *Biophysics*, 44(4):565-577, (1999).
2. G. Butler et al. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* 459, 657–662 (2009).