

Chapter 18

Multiple Sequence Alignment Optimization Using Meta-Heuristic Techniques

Mohamed Issa
Zagazig University, Egypt

Aboul Ella Hassanien
Cairo University, Egypt

ABSTRACT

Sequence alignment is a vital process in many biological applications such as Phylogenetic trees construction, DNA fragment assembly and structure/function prediction. Two kinds of alignment are pairwise alignment which align two sequences and Multiple Sequence alignment (MSA) that align sequences more than two. The accurate method of alignment is based on Dynamic Programming (DP) approach which suffering from increasing time exponentially with increasing the length and the number of the aligned sequences. Stochastic or meta-heuristics techniques speed up alignment algorithm but with near optimal alignment accuracy not as that of DP. Hence, This chapter aims to review the recent development of MSA using meta-heuristics algorithms. In addition, two recent techniques are focused in more deep: the first is Fragmented protein sequence alignment using two-layer particle swarm optimization (FTLPSO). The second is Multiple sequence alignment using multi-objective based bacterial foraging optimization algorithm (MO-BFO).

INTRODUCTION

Bioinformatics is a field that combines computer science and mathematics for analyzing and managing biological data. Developing large databases and complex tools for gene and protein analysis and modeling are the main tasks of bioinformatics besides organization, storing and retrieving biological data (Cohen, 2004). Sequence alignment becomes an essential tool of bioinformatics and it is vital in various tasks such as genomic annotation, protein secondary and tertiary structure prediction, phylogenetic tree construction, modeling binding sites, homology searches, gene regulation networks and functional

DOI: 10.4018/978-1-5225-2229-4.ch018

geneomics (Das, Abraham, & Konar, 2008 ; Durbin, Eddy, & Krogh, 1998). From biological point of view all organisms have a common ancestors and so the similarity between DNA or protein sequences exist. The function of newly known sequences with a known sequence can be known with measuring the similarity (Alberts et al., 2007; Arthur, 2002 ; Zvelebil & Baum, 2008).

Sequence alignment arranges DNA, RNA and protein sequences to locate conserved blocks or region of similarity. It lining up the nucleotides (A,C,G and T) in DNA or amino acids (20 different amino acids) in protein sequences to achieve the maximum possible level of similarity (Song.J, Liu, Song.Y, & Qu, 2007). The function similarity between sequences is predicted corresponding to the regions of similarity. This arrangement needs insertion of gaps in positions that maximize the alignment score and nucleotides/residues matching.

Finding sequence alignment experimentally is sensitive to less accuracy due to experimental errors with much time consuming and cost. Hence, many efforts in the last years to develop software tools that propose efficient model for accurate alignment. Aligning two sequences is called pairwise sequence alignment. While aligning more than two sequences is called multiple sequence alignment (MSA) as shown in Figure 1 (Sievers & Higgins, 2014). MSAs computation is almost computationally expensive and it classified as NP-complete problem. This chapter focus on the MSA techniques.

The MSA's methods are divided into four approaches: Exact, Progressive, Consistency based and iterative approach (Notredame, 2002). In the exact method (DP) was used for pairwise global alignment by computing the alignment over the entire length of the sequences, (Needleman & Wunsch, 1970). In DP a matrix is created and filled with the partial alignment scores of the two sequences. DP tries to find the shortest path with maximum alignment cost between the start and end of the sequences. The main limitations of DP approach are time and space complexities especially for number of sequences more than 2 sequences (Lipman, Altschul, & Kececioglu, 1989; Carrillo & Lipman, 1988)

Progressive approach solve the problems of the exact method by decreasing the time and space complexities (Taylor, 1988 ; Feng & Doolittle, 1987) The idea of using progressive technique is aligning the most related sequences and then incrementally adding the more distant one by one. The common

Figure 1. Example of Aligning 3 DNA sequences (MSA)
Öztürk & Aslan, 2016.

A	—	T	T	—	—	C	T	G	A	—	A	T	—
—	C	T	T	A	C	C	—	G	—	A	A	T	G
A	C	T	A	A	—	C	T	G	—	—	A	T	G

A	—	T	T	T	—	—	C	T	G	—	A	A	—
C	T	—	T	T	A	C	C	—	—	G	A	A	G
A	C	T	A	A	A	—	C	T	G	—	—	A	G

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/multiple-sequence-alignment-optimization-using-meta-heuristic-techniques/180955

Related Content

Optimization of the Wake-Up Scheduling Using a Hybrid of Memetic and Tabu Search Algorithms for 3D-Wireless Sensor Networks

Vrajesh Kumar Chawra and Govind P. Gupta (2022). *International Journal of Software Science and Computational Intelligence* (pp. 1-18).

www.irma-international.org/article/optimization-of-the-wake-up-scheduling-using-a-hybrid-of-memetic-and-tabu-search-algorithms-for-3d-wireless-sensor-networks/300359

Potentials of Quadratic Neural Unit for Applications

Ricardo Rodriguez, Ivo Bukovsky and Noriyasu Homma (2011). *International Journal of Software Science and Computational Intelligence* (pp. 1-12).

www.irma-international.org/article/potentials-quadratic-neural-unit-applications/60745

On Localities of Knowledge Inconsistency

Du Zhang (2011). *International Journal of Software Science and Computational Intelligence* (pp. 61-77).

www.irma-international.org/article/localities-knowledge-inconsistency/53163

Techniques for the Automated Segmentation of Lung in Thoracic Computed Tomography Scans

William F. Sensakovic and Samuel G. Armato (2012). *Machine Learning in Computer-Aided Diagnosis: Medical Imaging Intelligence and Analysis* (pp. 145-158).

www.irma-international.org/chapter/techniques-automated-segmentation-lung-thoracic/62228

Learning Algorithms for RBF Functions and Subspace Based Functions

Lei Xu (2012). *Machine Learning: Concepts, Methodologies, Tools and Applications* (pp. 1034-1065).

www.irma-international.org/chapter/learning-algorithms-rbf-functions-subspace/56186