# Chapter 31 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

DOI: 10.4018/978-1-7998-1204-3.ch031

#### Multiple Sequence Alignment Optimization Using Meta-Heuristic Techniques

tree construction, modeling binding sites, homology searches, gene regulation networks and functional 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 complexties (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 MSA techniques that based on the progressive approach are CLUSTALW (Thompson, Higgins, &Gibson, 1994), MUSCLE (Edgar, 2004), CLUSTAL OMEGA (Sievers & Higgins, 2014) and Multi-Align

Α	—	Т	Τ	-	-	С	Τ	G	Α	—	Α	Т	—
_	С	Т	Т	Α	С	С	-	G	-	Α	Α	Т	G
Α	С	Т	Α	Α	Ι	С	Т	G	-	_	Α	Т	G

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

Α	-	Т	Т	Т	-	_	С	Т	G	-	Α	Α	_
С	Т	Ι	Т	Т	$\mathbf{A}$	С	С	Ι	Ι	G	Α	Α	G
$\mathbf{A}$	С	Т	Α	Α	$\mathbf{A}$	-	С	Т	G	Ι	Ι	Α	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/243133

## **Related Content**

#### A Predictive Analytics Framework for Blood Donor Classification

Kavita Pabrejaand Akanksha Bhasin (2021). International Journal of Big Data and Analytics in Healthcare (pp. 1-14).

www.irma-international.org/article/a-predictive-analytics-framework-for-blood-donor-classification/277644

## Data Stream Mining of Event and Complex Event Streams: A Survey of Existing and Future Technologies and Applications in Big Data

Chris Wrench, Frederic Stahl, Giuseppe Di Fatta, Vidhyalakshmi Karthikeyanand Detlef D. Nauck (2016). *Enterprise Big Data Engineering, Analytics, and Management (pp. 24-47).* www.irma-international.org/chapter/data-stream-mining-of-event-and-complex-event-streams/154552

#### Remote Patient Monitoring for Healthcare: A Big Challenge for Big Data

Andrew Stranieriand Venki Balasubramanian (2019). *Managerial Perspectives on Intelligent Big Data Analytics (pp. 163-179).* 

www.irma-international.org/chapter/remote-patient-monitoring-for-healthcare/224338

## Integrated Approach for Automatic Crackle Detection Based on Fractal Dimension and Box Filtering

Cátia Pinho, Ana Oliveira, Cristina Jácome, João Manuel Rodriguesand Alda Marques (2020). *Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications (pp. 815-832).* www.irma-international.org/chapter/integrated-approach-for-automatic-crackle-detection-based-on-fractal-dimensionand-box-filtering/243145

#### Commercial Banks' Digital Paradigm and Customers Responses in the UAE

Muhammad Jumaa (2020). International Journal of Data Analytics (pp. 68-79). www.irma-international.org/article/commercial-banks-digital-paradigm-and-customers-responses-in-the-uae/244170