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

Copyright © 2020, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

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).

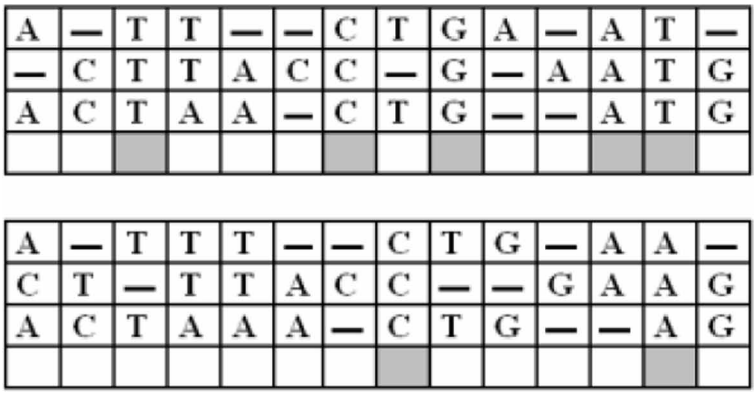
Sequence alignment arranges DNA, RNA and protein sequences to locate conserved blocks or region of similarity. It lining up the nucleotides (A,C,G and T) in DNA or amino acids (20 different amino acids) in protein sequences to achieve the maximum possible level of similarity (Song.J, Liu, Song.Y, & Qu, 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

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



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

Analysis of Heart Disease Using Parallel and Sequential Ensemble Methods With Feature Selection Techniques: Heart Disease Prediction
Dhyan Chandra Yadavand Saurabh Pal (2021). *International Journal of Big Data and Analytics in Healthcare (pp. 40-56).*
www.irma-international.org/article/analysis-of-heart-disease-using-parallel-and-sequential-ensemble-methods-with-feature-selection-techniques/268417

Introduction to Programming R and Python Languages
(2017). *Comparative Approaches to Using R and Python for Statistical Data Analysis (pp. 32-77).*
www.irma-international.org/chapter/introduction-to-programming-r-and-python-languages/175144

User-Independent Detection for Freezing of Gait in Parkinson's Disease Using Random Forest Classification
Amruta Meshramand Bharatendra Rai (2019). *International Journal of Big Data and Analytics in Healthcare (pp. 57-72).*
www.irma-international.org/article/user-independent-detection-for-freezing-of-gait-in-parkinsons-disease-using-random-forest-classification/232336

ICTs and Domestic Violence (DV): Exploring Intimate Partner Violence (IPV)
Bolanle A. Olaniran (2021). *International Journal of Big Data and Analytics in Healthcare (pp. 31-44).*
www.irma-international.org/article/icts-and-domestic-violence-dv/277646

Fuzzy Logic-Based Predictive Model for the Risk of Sexually Transmitted Diseases (STD) in Nigeria
Jeremiah A. Balogun, Florence Alaba Oladeji, Olajide Blessing Olajide, Adanze O. Asinobi, Olayinka Olufunmilayo Olusanyaand Peter Adebayo Idowu (2020). *International Journal of Big Data and Analytics in Healthcare (pp. 38-57).*
www.irma-international.org/article/fuzzy-logic-based-predictive-model-for-the-risk-of-sexually-transmitted-diseases-std-in-nigeria/259987