

Title:

An Approximation Approach for Multiple Sequence Alignment Using Central Star Method Method with DNA Sequence Datasets.

Name(s):

Gudipati Sai Krishna(M190241CS)

Nandi Dileep Kumar(M190437CS)

Abstract:

A multiple sequence alignment (MSA) is a process of aligning three or more biological sequences, generally protein, DNA, or RNA. MSA's require more sophisticated methodologies than pairwise alignment because they are more computationally complex.

Central Star Method is an approximation solution for aligning multiple sequences. It uses heuristic methods rather than global optimization because identifying the optimal alignment between more than a few sequences of moderate length is computationally expensive and consumes more time.

The brief idea of how the center star method of approximation works as follows:

Let $D(X, Y)$ be the pairwise optimal alignment distance between X, Y . $S_c \in S$ is said to be the center string of S if it maximizes: $\sum D(S_c, S_i)$

Identify the center string S_c of S .

Uses the alignments of S_c with each S_i to create a multiple alignment.

The Motivation of the project is to provide an efficient approximation algorithm for MSA by minimizing time and computational constraints.

The outcomes of the project are:

sequence homology can be inferred.

phylogenetic analysis can be conducted to assess the sequences sharing evolutionary origins.

Problem Statement:

An approximation approach for MSA Using Central Star Method with DNA Sequence Datasets through Python 3.6 to infer sequence homology and also to conduct phylogenetic analysis to assess the sequences sharing evolutionary origins.

References:

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7344909>

<https://www.kaggle.com/thomasonnelson/working-with-dna-sequence-data-for-ml>