National Institute of Technology Rourkela

Departmental Seminar

Seminar Title	: DNA Sequence Similarity Using Discrete Wavelet Transform
Speaker	: Nikita Banerjee (520cs3011)
Supervisor	: Prof. S. Pyne, PIC, Seminars
Venue	: Convention room CSE Dept.
Date and Time	: 26 Nov 2024 (11:30 AM)

Abstract : The development of next-generation sequencing technologies has resulted in exponential growth in sequence data. Global or local, pairwise or multiple sequence alignment was the foundation of early methods for sequence analysis, which was computationally intensive and time-consuming. In addition, gaps, mismatches, and insertions/deletions might provide unsatisfactory results during the alignment process, particularly for highly divergent sequences. We have used alignment-free sequence analysis to overcome this disadvantage and identify their phylogenetic relationship by using discrete wavelet transformation on the chaos game representation. First, DNA sequences were converted into 2D images to create a feature matrix. The DWT method is then applied to the resulting matrix to extract essential features and scale down their dimensionality; the resulting feature vectors are used for similarity analysis using different distance matrices. Further, we have executed our model on two benchmark datasets, Cichlid fish and Yersinia strains, from the AF project to evaluate performance, and we have achieved top rank with an accuracy of 95% in Cichlid fish and 100% in Yersinia strains.