THE UNIVERSITY OF CHICAGO

COMPUTATIONAL AND APPLIED MATHEMATICS
STUDENT SEMINAR

ROSALIND PAN
Department of Biochemistry and Molecular Biology
University of Chicago

Alignment-free evolutionary analysis of unalignable protein regions

Tuesday, February 23, 2021, 3:00 PM (Chicago Time)
Zoom Meeting ID: 972 4989 6395
Password: 673056

ABSTRACT

Understanding how functional information is encoded in the amino-acid sequences of proteins is a key challenge in modern biology. A common method to tackle this challenge is to identify conserved sequence features by examining the sequence alignment of related proteins. However, traditional conservation analysis becomes unreliable when the sequences of the target region are poorly aligned. To address this fundamental challenge, here we build a model that summarizes the sequence features of protein regions as joint probability distributions, which circumvent the need for residue-level homology in sequence comparisons, thereby permitting alignment-free evolutionary analysis. By targeting poly(A)-binding proteins (PABP), a family of proteins that modulate the stress response of budding yeast, we demonstrate that the analyses routinely carried out on aligned sequences can be satisfactorily performed on unalignable sequences using our model. Importantly, the Jensen-Shannon distances between sequence-based probability distributions can estimate the phylogenetic distances between the unalignable regions of PABP with high reliability. Moreover, by treating the target sequences as Markov chains, we show that our model captures the majority of the sequence information in the unalignable regions of PABP. We also demonstrate our ability to identify evolutionarily conserved sequence features in the target sequences despite poor residue-level conservation.