

Postgraduate Seminar Series

Venue: Graduate Seminar Room

Date & Time: June 1, 2024 at 2:30 PM

Speaker Information

Farha Akhter Munmun (Std No. 0419052092) is a part time M.Sc. student in the department of CSE, BUET. She completed her undergraduate studies from Rajshahi University of Engineering and Technology (RUET) in December, 2017. Her research interest lies in the field(s) of Bioinformatics, Machine Learning, and Network Security. She is currently doing her postgraduate thesis under the supervision of Dr. Md. Shamsuzzoha Bayzid. She will be speaking about her ongoing research in this talk.



Improving phylogenomic analysis using naïve binning with quartetbased super tree estimation

Estimating a species tree from biomolecular sequences is extremely difficult, especially when confronted with gene tree heterogeneity resulting from incomplete lineage sorting (ILS). Two of the most popular techniques for estimating species tree are: combined analysis (CA), which concatenates multiple sequence alignments of different genes into a single supergene alignment and then estimates a tree from this alignment, and another one is summary methods, which compute gene trees from different loci and then combine the inferred gene trees into a species tree. CA could be highly accurate in many cases as the combined gene alignments offer a high level of phylogenetic signals. However, it is agnostic about gene tree discordance (i.e., different genes having different evolutionary histories), leading to statistical inconsistency. On the other hand, summary methods can explicitly account for gene tree discordance and the underlying biological reasons, and thus could be statistically consistent. But they do not perform well when the number of genes is limited and the gene trees are not well estimated (i.e., gene tree estimation errors are prevalent). In this study, we have introduced a hybrid pipeline for species tree estimation that combines the strengths of both the combined analysis method and summary methods. Specifically, we have updated the process flow of a widely used quartetbased summary method called SVDquartets by combining SVDquartets with an existing technique called “binning” and a highly accurate quartet amalgamation technique called wQFM. We assessed the performance of our proposed hybrid model on a collection of simulated and real biological datasets that cover a wide range of challenging model conditions with varying numbers of genes, and varying amounts of gene tree estimation errors, and gene tree discordance. We found that our proposed method exhibited mixed performance: it outperformed the original SVDquartets in certain model conditions, while underperforming in others. However, the results on both simulated and real biological datasets suggest that this hybrid model could be a promising approach for estimating species trees, especially when we have a sufficiently large number of genes.