University of New Mexico
UNM Digital Repository

Shared Knowledge Conference   2018 Conference

Nov 7th, 3:00 PM - 4:00 PM

Analysis of ranked gene tree probability distributions under the coalescent process for detecting anomaly zones

Anastasiia Kim
anastasiiakim@unm.edu

Follow this and additional works at: https://digitalrepository.unm.edu/skc

Part of the Applied Statistics Commons, Biostatistics Commons, and the Probability Commons

Analysis of ranked gene tree probability distributions under the coalescent process for detecting anomaly zones

Anastasiia Kim

Department of Mathematics and Statistics
University of New Mexico

In phylogenetic studies, gene trees are used to reconstruct species tree. Under the multispecies coalescent model, gene trees topologies may differ from that of species trees. The incorrect gene tree topology (one that does not match the species tree) that is more probable than the correct one is termed anomalous gene tree (AGT). Species trees that can generate such AGTs are said to be in the anomaly zone (AZ). In this region, the method of choosing the most common gene tree as the estimate of the species tree will be inconsistent and will converge to an incorrect species tree when the number of loci increases.

In this work, we focus on unranked and ranked trees. The difference between these two is that the ranked gene tree not only depicts the topological relationship among gene lineages but also the sequence in which the lineages coalesce (join). In our project, software which allows computing probabilities of ranked gene trees given a species tree under coalescent process was developed to study how the parameters of the species tree simulated under a birth-death process can affect the AZ. Since some combinations of topologies and branch lengths in a species tree can produce AGTs, we compute the probabilities of ranked and unranked gene trees for the entire distribution of 5-8 taxon species trees to find a set of branch length space in which a species tree has either unranked AGTs, ranked AGTs, or both. Because the number of all possible tree topologies grows exponentially with the number of species, we propose some heuristic approaches for inferring large trees. Studying the properties of AGTs, as well as a connection between ranked and unranked anomaly zones, will help to find strategies for solving the problem of AGTs during phylogenetic inference.