Carrying out Phylogenetic Analyses through Computational Model Checking

Ying Xu, University of Georgia

This installment of Computer's *series highlighting the work published in IEEE Computer Society journals comes from* IEEE/ACM Transactions on Computational Biology and Bioinformatics.

Phylogenetic analysis is the main technique used to study evolutionary relationships among a given collection of organisms or homologous biomolecules, such as DNA and proteins encoded or used in these organisms. Researchers use such analyses to classify specified organisms taxonomically, determine the origins of protein-encoding genes, and infer how proteins function based on their conserved sequence motifs.

Typically, phylogenic analysis involves two steps: constructing a phylogenetic tree or network based on the given biosequences and deriving biological information from the tree or network. Many computational methods have been developed to construct phylogenetic trees, but in most published studies the biological information from the phylogenetic trees is derived manually.

In "Temporal Logics for Phylogenetic Analysis via Model Checking" (*IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 10, no. 4, 2013, pp. 1058–1070), Jose Ignacio Requeno and his colleagues present a novel way to derive biological information from a phylogenetic tree or network using model-checking techniques. Specifically, the authors represent a phylogenetic tree (or a network, as they plan to do in the future) as a mathematical model using temporal logic and the hypothesized biological properties of the tree.

The authors propose a computational scheme using tools established in the fields of temporal logic and model checking to computationally determine whether the represented phylogenetic tree has the predicted properties, saving evolutionary biologists the time and effort required to manually examine the hypothesized properties against large quantities of biological data. While the presentation may be somewhat technical for readers who aren't specialists, particularly in biology, the paper clearly outlines the overall logic.

The methodology presented in this paper can potentially be applied to other model-checking problems in data-intensive biological fields, including dynamic property analyses such as flux analyses over provided metabolic networks; dynamic property studies of biomolecular structures such as protein, DNA, or RNA structures; and inferring the structural properties of a collection of given genomes.

This work represents a new and exciting development linking fundamental computer science theories with data-intensive and increasingly model-intensive fields of modern biology. Bringing these two fields together should yield healthy advances in both. ◼

> Ignacio Requeno and his colleagues present a novel way to derive biological information from a phylogenetic tree using model-checking techniques.

YIN XU is a professor in the Department of Biochemistry and Molecular Biology at the University of Georgia. Contact him at xyn@bmb.uga.edu.