Grand Challenges in Constructing Large Evolutionary Trees

Abstract: Constructing the Tree of Life requires statistical estimation methods that have high accuracy on genome-scale data. These estimations are challenging, since nearly all statistical estimation methods are attempts to solve NP-hard optimization problems, and the heuristics that are used may not have good accuracy on large datasets. Equally troubling, many statistical estimation methods assume that the input data are high quality, and have poor accuracy on low quality data that are typical of many genome-scale evolutionary datasets. In this talk, I will present new computational methods we have developed that enable large-scale phylogeny estimation to be highly accurate and computationally efficient. I will also show how we used these techniques to estimate the avian tree of life (Jarvis et al., Science 2014) and the land plant tree of life (Wickett et al., PNAS 2014).

Biosketch: Tandy Warnow is the Founder Professor of Bioengineering and Computer Science at the University of Illinois at Urbana-Champaign, a member of the Institute for Genomic Biology,
and an affiliate in the departments of Statistics, Mathematics, Animal Biology, and Entomology. Tandy received her PhD in Mathematics in 1991 at the University of California at Berkeley under the direction of Eugene Lawler, and she did postdoctoral training with Simon Tavare and Michael Waterman at the University of Southern California from 1991-1992. Tandy was on the faculty at the University of Pennsylvania from 1994-1999, and then joined the University of Texas, before joining the faculty at the University of Illinois in 2014.

Tandy received the National Science Foundation Young Investigator Award in 1994, the David and Lucile Packard Foundation Award in Science and Engineering in 1996, a Radcliffe Institute Fellowship in 2006, and a Guggenheim Foundation Fellowship for 2011. Her research combines mathematics, computer science, and statistics to develop improved models and algorithms for reconstructing complex
and large-scale evolutionary histories in both biology and historical linguistics. Her current research focuses on phylogeny and alignment estimation for very large datasets (10,000 to 1,000,000 sequences), estimating species trees and phylogenetic networks from collections of gene trees, and metagenomics.