



Tandy Warnow

University of Illinois, College of Engineering

Friday, January 25, 2019

3:00 pm

Luddy Hall, Rm. 1106

Theoretical and Empirical Advances in Species Tree Estimation

Abstract: The estimation of species trees from multi-locus datasets is a basic step in many biological research projects. However, heterogeneity between the loci resulting from processes such as incomplete lineage sorting and horizontal gene transfer make standard approaches (such as concatenation using maximum likelihood) statistically inconsistent. In this talk, I will present very recent work on species tree estimation from multi-locus data sets when gene trees can differ from the species tree due to incomplete lineage sorting. I will cover new methods (ASTRAL, ASTRID, BBKA, and SVDquest) for estimating species trees, and discuss the current understanding about statistical consistency in two contexts: when sequence lengths and number of genes both go to infinity (essentially assuming perfect gene trees) or when the sequence length per gene is bounded but the number of genes goes to infinity. I will also present NJMerge, a new technique for improving the scalability of these methods to large data sets. Much of this talk is based on joint work with Siavash Mirarab (UCSD), Erin Molloy (Illinois), Mike Nute (Illinois), Sébastien Roch (Wisconsin), Pranjal Vachaspati (Illinois), and Théo Zimmermann (Paris).

Biography: Tandy Warnow is the Founder Professor of Computer Science at the University of Illinois at Urbana-Champaign, where she is also an affiliate in Mathematics, Statistics, Bioengineering, Electrical and Computer Engineering, Animal Biology, Entomology, and Plant Biology. Tandy received her PhD in Mathematics at UC Berkeley under the direction of Gene Lawler, and did postdoctoral training with Simon Tavaré and Michael Waterman at USC. Her research combines computer science, statistics, and discrete mathematics, focusing on developing improved models and algorithms for reconstructing complex and large-scale evolutionary histories in biology and historical linguistics. She has published more than 160 papers and one [textbook](#), graduated 11 PhD students, and has 5 current PhD students. She has been a visiting faculty member at many universities, including Princeton University, the University of Maryland, Yale University, Ecole Polytechnique Fédérale de Lausanne (EPFL), and Harvard University. Her awards include the NSF Young Investigator Award (1994), the David and Lucile Packard Foundation Award (1996), a Radcliffe Institute Fellowship (2006), and the John Simon Guggenheim Foundation Fellowship (2011). She was elected a Fellow of the Association for Computing Machinery (ACM) in 2015 and of the International Society for Computational Biology (ISCB) in 2017. Her national service includes being the lead NSF program officer for BigData (2012-2013), chairing the BioData Management and Analysis (BDMA) study section at NIH (2010-2012). Tandy was also a member of the Big Data Senior Steering Group of NITRD subcommittee of the National Technology Council (coordinating federal agencies), 2012-2013.

