



THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

TANDY WARNOW

Departments of Computer Science and Bioengineering
University of Illinois at Urbana-Champaign

Grand Challenges in Phylogenomics

MONDAY, October 31, 2016, at 12:15 PM
Stevanovich Center, MS 112, 5727 S. University Avenue

ABSTRACT

Estimating the Tree of Life will likely involve a two-step procedure, where in the first step trees are estimated on many genes, and then the gene trees are combined into a tree on all the taxa. However, the true gene trees may not agree with the species tree due to biological processes such as deep coalescence, gene duplication and loss, and horizontal gene transfer. Statistically consistent methods based on the multi-species coalescent model have been developed to estimate species trees in the presence of incomplete lineage sorting; however, the relative accuracy of these methods compared to the usual "concatenation" approach is a matter of substantial debate within the research community. In this talk I will present new state of the art methods we have developed for estimating species trees in the presence of incomplete lineage sorting (ILS), and show how they can be used to estimate species trees from genome-scale datasets. I will also discuss tradeoffs between data quantity and quality, and the implications for big data genomic analysis.