



**L.H. BAKER CENTER FOR BIOINFORMATICS AND
BIOLOGICAL STATISTICS AND PLANT SCIENCES INSTITUTE
SEMINAR SERIES**

MICHAEL SANDERSON, Ph.D.
Department of Ecology and Evolutionary Biology,
University of Arizona

**Building the tree of life accurately: are genomes
enough?**

Ten percent of the two million described biological species have been sequenced for at least one gene. In the next decade it will be possible to obtain complete genome sequences for many, if not most, of these species. In this talk I will discuss the prospects for accurate reconstruction of a phylogenetic "tree of life" based on such data. Distinct problems of scaling emerge in the directions of exploiting entire genomes on the one hand and large numbers of species on the other. Using examples from plant comparative genomics and plant molecular phylogenetics, I will discuss several computationally challenging problems that must be solved to leverage these extraordinarily large data sets. First, large scale phylogenetic data sets are dominated by missing entries, which arise by diverse biological processes and can impede inference. Second, our conventional phylogenetic paradigm focuses on single copy genes, but as data sets scale up, these become increasingly rare. Plant genomes in particular are end-products of whole genome duplications and frequent individual gene duplication and loss. Finally, rounding out the tree at local scales among closely related species must overcome obstacles both biological and technical: evolutionary processes that are most evident at this scale (lineage sorting, introgression), and the technical challenges of using rapidly evolving noncoding regions of the genome. Multiple sequence alignment in this context performs relatively poorly, exerting a downstream impact on tree quality. After describing these problems and some avenues for solution, I will discuss some of the known theoretical results on accuracy of phylogenetic inference and try to suggest ways they should be re-cast in light of genomic data.

Date: Wednesday, April 4

Time: 4:00-5:00 PM

Room: 1414 Molecular Biology