Using genome-wide data for phylogenetic inference and analysis has become very common in the post-genomic era, giving rise to the field of phylogenomics (which I use here to mean phylogenetic inference from genome-wide data). The multispecies coalescent (MSC) model has emerged as the main stochastic process that helps capture the intricate relationship between species trees and gene trees. Combined with models of sequence evolution, the MSC can be viewed as a generative model of genomic sequence data in the context of a (species) phylogenetic tree. In particular, the MSC naturally explains and allows for quantifying the phenomenon of incomplete lineage sorting (ILS).

From a biological perspective, a significant outcome of the use of genome-wide data has been the increasing evidence, or hypotheses, of reticulation during the evolution of various groups of eukaryotic species (reticulation has long been acknowledged as a major evolutionary process in prokaryotes, but not so in eukaryotes!). Therefore, developing models and methods that allow for reticulation in evolutionary analyses is very important.

Reticulate evolutionary histories are best represented as phylogenetic networks, which extend the tree model to allow for admixtures of genetic material. In this talk, I will describe the multispecies network coalescent (MSNC) model, which extends the MSC model so that it operates within the branches of a phylogenetic network. This extended model naturally allows for modeling vertical and horizontal evolutionary processes acting within and across species boundaries. In particular, it simultaneously accounts for gene tree incongruence across loci due to both hybridization and incomplete lineage sorting. I will then describe a likelihood function for this model, as well as a method for Bayesian sampling of phylogenetic networks and their parameters using reversible-jump Markov chain Monte Carlo (RJMCMC). I will also briefly describe how the MSNC model could be combined with hidden Markov models to simultaneously capture spatial and temporal dependencies in a group of genomes in the presence of introgression. I will demonstrate the utility of this framework via a reanalysis of a mosquito genomic data set, and discuss the method's performance on synthetic data as well. All the methods I describe have been implemented in our open-source software package, PhyloNet, which is publicly available at http://bioinfo.cs.rice.edu/phylonet.