A fundamental problem in evolutionary biology is to determine relationships between species. Are humans most closely related to chimpanzees or to gorillas? To answer such questions, some genetic data like DNA sequences might be collected and aligned, and then a Maximum Likelihood or Bayesian analysis undertaken to give the tree or distributions on trees best depicting the evolutionary relationships for these data.

If, however, the DNA sequences are from genes, then this paradigm constructs a gene tree, and it is well known that a gene tree may disagree with the true species tree. Kingman introduced the coalescent model to probabilistically model the phenomenon of gene lineages coalescing within a species tree.

In this talk, we consider the information content of gene tree distributions, and how we can summarize gene tree distributions while still retaining enough information to identify the underlying species tree. More formally, we discuss the identifiability of the species tree parameter under the coalescent model.

This talk will be non-technical and accessible. (Received August 18, 2011)