Interpreting genomic polymorphism data: what history has to tell us

Simon Tavaré

Departments of Biological Sciences, Mathematics and Preventive Medicine

University of Southern California

Technological advances in molecular biology have made it possible to survey genome-wide DNA sequence variation in natural populations. These data include restriction fragment length polymorphisms, microsatellite repeats, single nucleotide polymorphisms (SNPs) and complete DNA sequences of particular loci. The analysis and interpretation of the patterns of variation seen in such data is complicated by the fact that the sampled chromosomes share a common ancestry, thus making the data highly dependent. To make matters worse, the nature of this common ancestry is not known precisely and therefore needs to be modeled. Since the early 80s, population geneticists have used the *coalescent* as a stochastic description of the ancestry of a sample of chromosomes, and there is now an extensive literature on inference and estimation for such processes. In this talk I will give an overview of coalescent methods, touching on a number of applications including inference about the age of mutations and the hunt for disease genes using linkage disequilibrium mapping.