Abstract

This thesis uses mathematical models of evolutionary processes to make inferences from DNA sequences about variation in the processes of molecular evolution. Questions addressed include the differences in evolutionary processes observed at each codon position, including in overlapping reading frames; how this variation may be used to locate protein-coding open reading frames in genomic alignments; differences in evolutionary signal either side of an origin of replication; differences in evolutionary processes between regions of a chromosome and how to combine large datasets of separate genes to produce a single phylogeny and the biological factors that are important to account for when combining such data.

Chapter 1 discusses simple models of genetic change along evolutionary lineages and introduces many of the methodological principles upon which the research in this thesis is carried out. Chapter 2 uses evolutionary models to analyse the differences in evolutionary processes at each of the three codon positions of protein-coding DNA and then proceeds to study the same processes in overlapping reading frames, using the hepatitis B virus as a model organism. Chapter 3 uses the variation in the evolutionary processes at the different sites in a codon to formulate a powerful search tool for protein-coding sequences in windows across large-scale genomic alignments, which are largely non-protein-coding. Chapter 4 introduces a model of evolution that tests for the signal we might expect to observe at a eukaryotic origin of DNA replication. This is then developed into a sliding window approach, where we simultaneously move two spatially distinct windows of DNA across an alignment and test for a signal between them. Chapter 5 generalises such two-window