

Characterizing Short DNA-binding Motifs in an Evolutionary Context

Haiyan HUANG, *University of California at Berkeley, USA*, E-mail: hhuang@stat.berkeley.edu

Abstract: Identifying regulatory mechanisms is a key step towards the goal of understanding the biology at a molecular network level. The recognition of regulatory elements is challenging due to the short and fuzzy features of binding sequences. A number of statistical and computational methods have been developed for tackling this problem. These efforts fall into two general approaches. The first seeks novel recurrent patterns in a set of DNA sequences which are believed to contain a common mechanism. The other evaluates individual candidate sequences by their similarity to clusters of experimentally determined regulatory sites. In both approaches, the position specific weight matrix (PWM), a mutual independence model, has been commonly used to summarize the sequence patterns. However, with the rapid accumulation of genomic data and the advent of comparative genomics, PWM are becoming less satisfactory. For instance, often the binding sequences are from multiple species at varying evolutionary distances and thus have different degrees of dependency between them. In this talk, I will introduce a tree-based statistical model for characterizing and finding short DNA binding motifs in an evolutionary context. This newly developed motif model has been found to be advantageous in many applications.