A Non-Tree Based Approach to Automated Construction of Familial Binding Profiles

Pilib Ó Broin[1,2], Terry Smith[1], Aaron Golden[1,3]

[1]National Centre for Biomedical Engineering Science, NUI, Galway
[2]Department of Medicine (Endocrinology), Albert Einstein College of Medicine
[3]Department of Genetics (Division of Computational Genetics), Albert Einstein College of Medicine

Familial Binding Profiles (FBPs) represent the average binding specificity for a group of structurally related DNA-binding proteins. The construction of such profiles allows the classification of novel motifs based on similarity to known families and can help to reduce redundancy in both motif databases and integration of results from multiple *de novo* motif prediction algorithms.

Many current approaches to automated motif clustering rely on progressive tree based techniques such as hierarchical clustering. These approaches can suffer from so-called poor frozen sub-alignments – situations in which a motif seemingly well-clustered early on in the tree construction process is later found to be more similar to another motif family. In order to avoid this scenario, we examine the application of a Genetic Algorithm (GA) to the clustering problem. In a GA approach, motifs can move freely between clusters at any stage in the clustering process, potentially avoiding the convergence to a local rather than a global maximum. Here we outline our approach, including distance metric, fitness function and evolutionary operators. We also show results obtained with our implementation when clustering a set of motifs from the JASPAR database and make comparisons with other approaches.