

MAR05-2004-003092

Abstract for an Invited Paper
for the MAR05 Meeting of
the American Physical Society

Applying Logic Analysis to Genomic Data and Phylogenetic Profiles

TODD YEATES, UCLA-DOE Institute of Genomics and Proteomics

One of the main goals of comparative genomics is to understand how all the various proteins in a cell relate to each other in terms of pathways and interaction networks. Various computational ideas have been explored with this goal in mind. In the original phylogenetic profile method, ‘functional linkages’ were inferred between pairs of proteins when the two proteins, A and B, showed identical (or statistically similar) patterns of presence vs. absence across a set of completely sequenced genomes. Here we describe a new generalization, logic analysis of phylogenetic profiles (LAPP), from which higher order relationships can be identified between three (or more) different proteins. For instance, in one type of triplet logic relation – of which there are eight distinct types – a protein C may be present in a genome iff proteins A and B are both present ($C=A\cap B$). An application of the LAPP method identifies thousands of previously unidentified relationships between protein triplets. These higher order logic relationships offer insights – not available from pairwise approaches – into branching, competition, and alternate routes through cellular pathways and networks. The results also make it possible to assign tentative cellular functions to many novel proteins of unknown function. Co-authors: Peter Bowers, Shawn Cokus, Morgan Beeby, and David Eisenberg