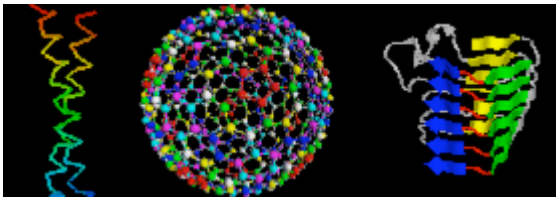
MIT
Department of Mathematics
& The Theory of
Computation Group
At CSAIL



Bioinformatics Seminar

Speaker: Serafim Batzoglou, Stanford
Title: Models and algorithms for genomic sequences, proteins, and networks of protein interactions.
Date: Monday, 1 May 2006
Time & Location:
Refreshments: 11 am in the Theory of Computation Lab at MIT's Building 32, Stata Center Room G-575
Talk: 11:30 am the Theory of Computation Lab at MIT's Building 32, Stata Center, Room G-575
URL: http://www-math.mit.edu/compbiosem/

Abstract:

This talk has two parts: the first part is on new ways to model and analyze biological sequences, which are the most abundant kinds of genomic data; the second part describes methods for constructing and comparing interaction networks, which are emerging as canonical data sets of the post-genomic era.

Algorithms for biological sequence analysis. One of the most fruitful developments in bioinformatics in the past decade was the wide adoption of Hidden Markov Models (HMMs) and related graphical models to an array of applications such as gene finding, sequence alignment, and non-coding RNA folding. Conditional Random Fields (CRFs) are a recent alternative to HMMs, and provide two main advantages: (1) they enable more elaborate modeling of biosequences by allowing us to conveniently describe and select rich feature sets. For example, when comparing two residues during protein alignment, using a CRF allows leveraging in a principled manner the chemical properties of the neighborhood of those residues. (2) CRFs allow training of parameters in a way that is more effective for making predictions on new input sequences. I will describe three practical CRF-based tools that improve upon state-of-the-art methods in terms of accuracy: CONTRAlign, a protein aligner; CONTRAST, a gene finder; and CONTRAfold, a method for predicting the secondary structure of non-coding RNAs. Our tools are available at http://contra.stanford.edu.

Networks of protein interactions. Graphs that summarize pairwise interactions between all proteins of an organism have emerged as canonical data sets that can be constructed using multiple sources of functional genomic data. We construct protein interaction networks for all sequenced microbes by rigorously integrating information extracted from genomic sequences as well as microarrays and other predictors of pairwise interactions. We then align these networks in multiple species using Graemlin, a tool that we developed for that purpose, and search for modules (subgraphs) of proteins that exhibit homology as well as conservation of pairwise interactions among many organisms. Graemlin provides substantial speed and sensitivity gains compared to previous network alignment methods; it can be used to compare microbial networks at http://graemlin.stanford.edu.

The seminar is co-hosted by Professor Peter Clote of Boston College's Biology and Computer Science Departments and MIT Professor of Applied Math Bonnie Berger. Professor Berger is also affiliated with CSAIL & HST.

Massachusetts Institute
of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

For General Questions, please contact kvdickey@mit.edu