

STATISTICS COLLOQUIUM

MONDAY, DECEMBER 03, 2007

TALK: 4:00 PM — SCIENCE CENTER RM. 309

RECEPTION: 5:15 PM — SCIENCE CENTER, 7TH FLOOR

“Evolutionary and functional classification of protein sequences”

Liisa Holm

University of Helsinki

And

Radcliffe Institute for Advanced Study

ABSTRACT

Proteins are major constituents of all cells and central to our understanding of molecular biology. The past and present genome projects have provided us with an exponentially growing wealth of protein sequences from many diverse organisms. However, this wealth creates a new problem. The majority of protein sequences have not been studied experimentally in the laboratory and our first-hand knowledge about them is minimal. The biological function of a novel protein sequence can often be inferred by studying similar protein sequences in other organisms. Just as whole organisms are related to each other by evolutionary descent, so are the proteins between organisms related by inheritance and mutation. Models of sequence evolution have improved on moving from uniform substitution matrices to position-specific profiles and pairwise profile comparisons. We have developed a method which exploits information from multiple profile comparisons and which has superior performance in the detection of remote homologs. Our method also generates a feature space which is useful for extracting functionally divergent protein subgroups and their function specific residues.

References:

1. Heger A, Mallick S, Wilton C, Holm L (2007). The global trace graph, a novel paradigm for searching protein sequence databases. *Bioinformatics* 23, 2361-2367.
2. Marttinen P, Corander J, Törönen P, Holm L (2006). Bayesian search of functionally divergent protein subgroups and their function specific residues. *Bioinformatics* 22, 2466-2474.