

MASTER'S SEMINAR ANNOUNCEMENT
Department of Statistics

A New Amino Acid Distance

Tuesday, August 5, 2003 at 11:00am
Eckhart Hall, Room 110, 5734 S. University Avenue

Hua Tang
Department of Statistics, University of Chicago

ABSTRACT

For 20 amino acids, we reduced the 190 possible pairwise amino acid changes to an elementary set of 75 changes that differ by only one base pair. Based on a Markov Chain Model for substitutions, we developed an Evolutionary Index (EI) which can be computed for each of the 75 elementary changes as the likelihood of substitutions, relative to that of synonymous changes. We use 1306 genes from rodents (mouse vs. rat) and 280 genes from primates (Human vs. Old World Monkey) to estimate the EI's. In either dataset, EI varies more than 12-fold, but the correlation coefficient of EI's from the two datasets is high ($r = 0.885$). This correlation suggests that the amino acid properties have a strong influence on protein evolution when examined at the genomic level over many genes. So there may exist a universal rule which governs the amino acid substitutions. We used a generalized linear model to capture the rule and infer 7 amino acids properties that are significant in determining the amino acid changes. From the model, we propose a new distance matrix which measures the similarity between amino acids. The amino acid distance may lead to an empirical and systematic classification of amino acids, which should refine the analysis of coding region evolution. It will be also implemented into the protein global alignment.