# TEMPLE UNIVERSITY
Department of Mathematics

# Applied Mathematics and Scientific Computing Seminar

Wednesday, 30 September 2015, 4:00 p.m.
Room 617 Wachman Hall
(refreshments and social at 3:45 p.m)

## Biological Inference Using Coarse Grained Generative Models for Gene Duplication, Metabolic Pathway Evolution, and Amino Acid Substitution in Proteins

by David Liberles
Department of Biology, Temple University

**Abstract.** My research group has a fundamental interest in understanding the processes driving the lineage-specific functional divergence of genomes. This includes changes in the protein coding content of genomes as well as changes in individual protein sequences, which mediated by protein structure and functional pathway structure can result in changes to pathway function and ultimately organismal phenotype. I will present modeling frameworks for three problems, presented as a mixture of forward and inverse problems.

The first problem is that of gene duplication and loss in genomes, which is one of the dominant processes shaping changes in gene content in eukaryotic genomes. Biologically-motivated duplication and loss models will be described that can be used in both a survival analysis framework and a phylogenetic birth-death process framework. Recent improvements in our loss models include explicit treatment of non-fixed duplicates differently from fixed duplicates and characterization of probabilities of retention under six processes: non-functionalization, and then non-functionalization plus neo-functionalization, sub-functionalization, dosage balance, and lastly, with dosage balance as a transition state to subsequent sub-functionalization and dosage balance to neo-functionalization. Proposed new birth models will also be discussed.

A second problem we will examine is the co-evolution of enzymes in metabolic pathways, where selection acts on pathway function (described by sets of ordinary differential equations in a Michaelis-Menten framework). We describe the important role of mutation-selection balance in describing the evolutionary dynamics and present some statistical approaches that may enable differentiation between inter-molecular compensatory evolution and directional change in pathway function. Lastly, models for the introduction of duplicates with promiscuous functions (the patchwork and retrograde models) are introduced, which act to stochastically change the structure of the differential equations.

A last problem that we will briefly introduce is that of amino acid substitution. We take two approaches to this problem. A first approach is to build linkage and a mixture modeling framework onto a variant of the mutation-selection modeling framework that was introduced by Halpern and Bruno (1998). A second approach is to re-purpose (and re-parameterize) a coarse grained force field designed for molecular dynamics to describe the relative fitness of an amino acid at a site based upon its energetic fit within a protein 3 dimensional structure and known protein-protein interactions. When ultimately structured as inverse problems, these methods are aimed at probabilistically detecting cases where amino acid substitution may have led to altered protein function between closely related genomes.