New frontiers in Viral Phylogenetics

Co-PIs: Tanmoy Bhattacharya, tanmoy@lanl.gov, and Sriram Swaminarayan, sriram@lanl.gov

The rapid advances in sequencing methodology now gives us access to unprecedented amounts of data capable of answering qualitatively new kinds of questions. However, the analytical tools are finding it difficult to keep up with the demands. For example, in the field of HIV research, a phylogenetic analysis of a few thousand sequences is the surest way to reveal features of evolution of the virus during acute infection, important for informing a vaccine against it. Similarly, deep sequences from individual patients are now capable of directly tracking the development of drug mutations and of escape from the immune system. The available state-of-the-art maximum likelihood phylogenetic tools, previously developed by us, are able to handle a set of few hundred sequences, which does not provide sufficient statistical power to answer these questions; both extensions in methodology and utilization of novel resources are likely to be needed to address them. This is a proposal to map this maximum likelihood problem on to Roadrunner, optimize the core computational routines, and apply it to data comparing chronic and acute sequences available through our CHAVI collaboration agreement. These core routines will also allow us, in future, to develop a Bayesian framework for phylogenetic analysis which will provide statistically valid credibility intervals on all these results.