Ori Sargsyan
Theoretical Biology and Biophysics and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, 87545, USA

Analytical framework for identifying and differentiating recent hitchhiking and severe bottleneck effects from multi-locus DNA sequence data

Hitchhiking and severe bottleneck effects have impact on the dynamics of genetic diversity of a population by inducing homogenization at a single locus and at genome-wide scale, respectively. As a result, identification and differentiation of the signatures of such events from DNA sequence data at a single locus is challenging.

In this talk I will present an analytical framework for identifying and differentiating recent homogenization events at multiple neutral loci in low recombination regions. The dynamics of genetic diversity at a locus after a recent homogenization event is modeled according to the infinite-sites mutation model and the Wright-Fisher model of reproduction with constant population size. In this setting, I derived analytical expressions for the distribution, mean, and variance of the number of polymorphic sites in a random sample of DNA sequences from a locus affected by a recent homogenization event. Based on this framework, three likelihood-ratio based tests are developed for identifying and differentiating recent homogenization events at multiple loci.

Lastly, I apply the framework to two data sets. First, I consider human DNA sequences from four non-coding loci on different chromosomes for inferring evolutionary history of modern human populations. The results suggest, in particular, that recent homogenization events at the loci are identifiable when the effective human population size is 50000 or greater in contrast to 10000, and the estimates of the recent homogenization events are agree with the “Out of Africa” hypothesis. Second, I use HIV DNA sequences from HIV-1-infected patients to infer the times of HIV seroconversions. The estimates are contrasted with other estimates derived as the mid-time point between the last HIV-negative and first HIV-positive screening tests. The results show that significant discrepancies can exist between the estimates.