TESTING THE PRESENCE OF RECOMBINATION HOT-SPOTS IN THE HIV GENOME AND THEIR ASSOCIATION WITH LOCAL SEQUENCE PROPERTIES

V.N. Minin[†1], M.A. Suchard[1], F. Fang[2], K.S. Dorman[2]

[1] *University of California, Los Angeles, USA*
[2] *Iowa State University, Ames, USA*

[†] E-mail: *vminin@ucla.edu*

We present a Bayesian framework for elucidating spatial preferences of recombination break-points using molecular sequences of multiple HIV recombinants. Recombination events are represented as parameters of a phylogenetic change-point model with a spatially smoothed prior on break-point locations. We use a Gaussian Markov random field (GMFR) prior to achieve smoothing of site-specific recombination log-odds. The posterior distribution of all model parameters is approximated with Markov chain Monte Carlo (MCMC) sampling. The small number of observed recombinant sequences complicates discriminating between preferential and uniform distributions of break-points. We propose to test the recombination hot-spot hypothesis using Bayes factors and show how to calculate them using an MCMC sample from the posterior. We modify our GMRF prior on recombination log-odds in order to model association of recombination frequency and local sequence properties, such as RNA secondary structure and sequence similarity. Such modification not only improves precision of the hot-spot mapping, but also allows us to examine several candidate features of the HIV genome and test their impact on the spatial variation of recombination occurrences. Analysis of recombinant sequences, derived from the HIV *gag* coding region, suggests a hot-spot in this part of the genome, promoted by RNA stem loop elements.