

Distribution of distances between topologies and its effect on detection of phylogenetic recombination

Leonardo de Oliveira Martins · Hirohisa Kishino

Received: 1 September 2008 / Revised: 3 April 2009 / Published online: 28 October 2009
© The Institute of Statistical Mathematics, Tokyo 2009

Abstract Inferences about the evolutionary history of biological sequence data are greatly influenced by the presence of recombination, that tends to disrupt the phylogenetic signal. Current recombination detection procedures focus on the phylogenetic disagreement of the data along the aligned sequences, but only recently the link between the quantification of this disagreement and the strength of the recombination was realised. We previously described a hierarchical Bayesian procedure based on the distance between topologies of neighbouring sites and a Poisson-like prior for these distances. Here, we confirm the improvement provided by this topology distance and its prior over existing methods that neglect this information by analysing datasets simulated under a complex evolutionary model. We also show how to obtain a mosaic structure representative of the posterior sample based on a newly developed centroid method.

Keywords Viral recombination · SPR distance · Markov chain Monte Carlo · Phylogenetics