

CODON-BASED DETECTION OF POSITIVE SELECTION CAN BE BIASED BY HETEROGENEOUS DISTRIBUTION OF POLAR AMINO ACIDS ALONG PROTEIN SEQUENCES

Xuhua Xia

Department of Biology, University of Ottawa 30 Marie Curie, P.O. Box 450, Station A,
Ottawa, Ontario, Canada, K1N 6N.

Sudhir Kumar

Center for Evolutionary Functional Genomics
The Biodesign Institute and The School of Life Sciences, Arizona State University
Tempe AZ, 85287-5301 USA

The ratio of the number of nonsynonymous substitutions per site (K_a) over the number of synonymous substitutions per site (K_s) has often been used to detect positive selection. Investigators now commonly generate K_a/K_s ratio profiles in a sliding window to look for peaks and valleys in order to identify regions under positive selection. Here we show that the interpretation of peaks in the K_a/K_s profile as evidence for positive selection can be misleading. Genic regions with $K_a/K_s > 1$ in the MRG gene family, previously claimed to be under positive selection, are associated with a high frequency of polar amino acids with a high mutability. This association between an increased K_a and a high proportion of polar amino acids appears general and not limited to the MRG gene family or the sliding-window approach. For example, the sites detected to be under positive selection in the HIV1 protein-coding genes with a high posterior probability turn out to be mostly occupied by polar amino acids. These findings caution against invoking positive selection from K_a/K_s ratios and highlight the need for considering biochemical properties of the protein domains showing high K_a/K_s ratios. In short, a high K_a/K_s ratio may arise from the intrinsic properties of amino acids instead of from extrinsic positive selection.