HONG GU, Dalhousie University

Estimation of Darwinian Positive Selection Using Generalized Codon-based Models

The generalized codon-based model for protein-coding DNA sequences offers great flexibility in the choice of rate matrix and significantly improves the goodness-of-fit. We further develop the mixed model approach for this type of model by allowing random effects on one or more parameters across the amino acid sites. By allowing random effects on the parameter that estimates the nonsynonymous change rate on different sites, generalized codon-based models include as special cases the M-series models in Yang et al (2000). We demonstrate the more accurate estimation of heterogeneous selection pressure at amino acid sites, comparing to the estimation by M-series models.