Analysis of prevalence of epistasis on the basis of huge phylogenies

Galya Klink, Georgii A Bazykin

MSU, Leninskie Gory, Moscow, 119991, Russian Federation, galkaklink@gmail.com

Epistatic interactions between amino acid sites shape the local fitness landscapes, affecting the site-specific probabilities of fixations of different amino acids. Comparisons of prevalences of amino acids at individual sites with site-specific dn/ds values suggest that the vast majority of amino acid fixations that occur at a given instant would be deleterious at another instant (Breen et al. 2012). However, this analysis can be confounded by the differences in fitness between different fixed variants (McCandlish et al. 2012). Indeed, a slightly deleterious allele can remain fixed for a while, inflating the number of amino acids observed at a site, but not increasing the dn/ds value significantly. This effect can be particularly pronounced in huge datasets, in which even rare slightly deleterious mutations are likely to be observed. The data on instantaneous selection coefficients associated with allele replacements is implicit in phylogenies. Here, we reconstruct the phylogeny of 8,000 cytochrome B proteins from 8,000 metazoan species, and use this data to obtain high-resolution site-specific distributions of survival times of all the amino acids observed at a site. An average amino acid site is occupied by ~10 amino acids at different species, consistent with the previous estimates (Breen et al. 2012). However, the sums of branch lengths occupied by individual amino acids differed greatly, with most of the amino acids occurring only at small near-terminal clades. In other words, the molecular clock is strongly overdispersed, consistent with selection disfavoring rare amino acids; and much of the incongruence between site-specific amino acid prevalences and dn/ds values can be explained without invoking epistasis. Still, the data cannot be fully explained under the assumption of invariant fitness landscape, and thus some changes between relative site-specific fitnessed of different amino acids, probably associated with epistasis, occur.

1. M.S. Breen et al. (2012) Epistasis as the primary factor in molecular evolution, *Nature*, **490:**535-538.

2. D.M.McCandlish et al. (2012) Epistasis not needed to explain low dN/dS, *Cornell University Library*, arXiv:1212.5239