Analysis of prevalence of epistasis on the basis of huge phylogenies

Galya V Klink, galkaklink@gmail.com

IITP RAS, Moscow,127051,Russian Federation

Georgii A Bazykin, yegor.bazykin@gmail.com

IITP RAS, Moscow,127051,Russian Federation

MSU, Leninskie Gory, Moscow, 119991, Russian Federation

Epistatic interactions between amino acid sites shape the site-specific fitness landscapes, affecting the site-specific probabilities of fixations of different amino acids. There is abundant evidence that epistasis has a major role in shaping the evolution of protein sequences; however, it is hard to quantify its contribution.

Here, we reconstruct the phylogeny of several mitochondrial proteins from ~3,000 metazoan species, and use this data to obtain high-resolution site-specific distributions of times between points of occurrence of every amino acid observed at each site. We show that substitutions to the same amino acid are clustered on the phylogenetic tree, and that the extent of clustering is higher in conservative sites. Furthermore, substitutions giving rise to amino acids that segregate as minor frequency alleles in the human population are also phylogenetically clustered near the human branch, showing that much of this polymorphism would be deleterious in other species.