

Abstract Submitted  
for the MAR06 Meeting of  
The American Physical Society

**Parameters of the proteome evolution from the distribution of sequence identities of paralogous proteins** KOON-KIU YAN, Stony Brook University, JACOB AXELSEN, Compexity Lab, Niels Bohr Institute, Denmark, SERGEI MASLOV, Brookhaven National Laboratory — The evolution of the full repertoire of proteins encoded in a given genome is driven by gene duplications, deletions and modifications of amino-acid sequences of already existing proteins. The information about relative rates and other intrinsic parameters of these three basic processes is contained in the distribution of sequence identities of pairs of paralogous proteins. We introduced a simple mathematical framework that allows one to extract some of this hidden information. It was then applied to the proteome-wide set of paralogous proteins in *H. pylori*, *E. coli*, *S. cerevisiae*, *C. elegans*, *D. melanogaster* and *H. sapiens*. We estimated the stationary per-gene deletion and duplication rates, the distribution of amino-acid substitution rate of these organisms. The validity of our mathematical framework was further confirmed by numerical simulations of a simple evolutionary model of a fixed-size proteome.

Koon-Kiu Yan  
Stony Brook University

Date submitted: 29 Nov 2005

Electronic form version 1.4