

Abstract Submitted
for the MAR05 Meeting of
The American Physical Society

An entropic tool for genome analysis CHIH-YUAN TSENG, Computational Biology and Bio-informatics Lab, Physics Dept., National Central Univ. — Shannon information (SI) defines the difference of Shannon entropy and its global maximum value. It was found SI in a genome tends to be much larger than that in its random match for all extant prokaryotic and eukaryotic complete genomes in Chang et al's work. Thus a better sense of the magnitude of the SI in a sequence is obtained by measuring it relative to the SI in the random match, the reduced SI. They observed a linear relation between reduced SI and sequence length L , which implies a k -dependent but genome-independent constant. This forms a universality class that indicates that reduced SI is a signature of complete genomes undiminished by the enormous diversity in growth and evolution experienced by individual genomes. Although their studies revealed intriguing results, the mechanism was not clear. Our main goal here is to investigate it through the method of maximum entropy (ME). The rationale hinges on the use of relative entropy. ME indicates preferred probability distribution of frequency- occurrence of k -string in real genome sequences updated from random sequences is the one that maximizes relative entropy of genomes and random sequences subject to certain constraints. Our result shows the existence of universality classes to be simply a trivial consequence if frequencies-occurrence of k -string is chosen as the relevant variable. However, the use of this result is far from being exhausted, which may provide a track to develop a genomic growth model.

Chih-Yuan Tseng
Computational Biology and Bio-informatics Lab, Physics dept., National Central Univ.

Date submitted: 05 Dec 2004

Electronic form version 1.4