

Composition Vector Approach to Phylogeny: Success and Foundation

Bailin Hao

T-Life Research Center and Department of Physics, Fudan University,
Shanghai 200433, China
Institute of Theoretical Physics, Academia Sinica, Beijing 100190, China
Santa Fe Institute, Santa Fe, NM 87501, USA

The composition vector approach (CVTree) is an alignment-free and parameter-free method to infer phylogeny from whole-genome data. Since it was proposed in 2003 the method has been successfully applied to viruses, chloroplasts, prokaryotes, and fungi. In particular, the phylogenetic trees obtained by CVTree method agree with latest taxonomies in major classification and fine branchings, while the disagreements often reveal problems debated by biologists for years. In this talk, after introducing the new CVTree update (to appear in the 2009 Web Server Issue of Nucleic Acids Research) we will touch on a few issues related to the foundation of the CVTree approach, such as relation of branch lengths to genetic distance, optimal choice of the peptide size (K) for constructing the composition vectors, test of the trees by statistical resampling (bootstrapping and jackknifing), etc.