

Prokaryote Phylogeny Based on Complete Genomes --- A Compositional Approach

Bailin Hao, Institute of Theoretical Physics and Beijing Genomics Institute, Academia Sinica

The availability of an ever growing number of complete genome sequences of organisms raises the problem of how to infer phylogenetic relationships from these data. As different organisms have different size of genomes, different number and order of genes, the traditional approach of molecular phylogeny based on alignment of conservative genes encounters much difficulty and has led to controversial results. In particular, the location of hyperthermophilic bacteria on the tree of life has triggered a hot debate on whether there has been massive lateral transfer of genes in the evolutionary history. We propose a new way of inferring phylogeny without doing sequence alignments. It uses K-string compositions to construct representative vectors for each species. This approach yields promising results when applied to prokaryote genomes.