

Improving the supertree approach by analyzing protein clusters with paralogs and including distance data

Sakamoto, T.¹, Ortega, J.M.¹

¹*Laboratório de Biodados, Departamento de Bioquímica e de Imunologia, ICB, Universidade Federal de Minas Gerais (UFMG).*

In phylogenomics, we can estimate a species tree using the supertree approach that consists in reconciling several gene trees to construct one concise tree. Currently, there are several methods that combine the information of thousand trees to construct a species tree, but some common restrictions and limitations are imposed by them. One of them is the non-acceptance of input trees containing paralogs in the analysis. This restriction is becoming more significant as the number of organisms with genomic data increases since this consequently increases the number of protein clusters with paralogy relationship and reduces the number of gene clusters acceptable for this analysis. Another key challenge for supertree approach is the estimation of evolutionary distance to allow users to address questions about the evolutionary rates and dates in the final tree. Here we present HyperTriplets, a phylogenomics tool that uses a triplet-based approach for supertree reconstruction. HyperTriplets takes as input a set of rooted phylogenetic trees in Newick format and decomposes them in triplets. The algorithm counts the occurrence of each type of triplets and uses these data to reconstruct the final tree. HyperTriplets deals with trees containing paralogs by classifying all internal nodes between speciation and duplication nodes. This allows the algorithm to extract during the tree decomposition only those triplets without duplication nodes. In the same time, HyperTriplets extracts the phylogenetic distance between pair of samples in the tree. These values are used by the algorithm to estimate the branch lengths of the final tree. HyperTriplets was tested using phylogenetic trees deposited in PhylomeDB database. The source code of HyperTriplets can be accessed at biodados.icb.ufmg.br/hypertriplets.

Financial Support: FAPEMIG, CAPES.