

ABSTRACT

Phyletic patterns denote the presence and absence of orthologous genes in completely sequenced genomes and are used to infer functional links between genes, on the assumption that genes involved in the same pathway or functional system are co-inherited by the same set of genomes. Phyletic patterns as a representation of the distribution of COGs (clusters of orthologous groups) across genomes, useful for tracking the evolutionary.

To derive clusters of related patterns from their pairwise distances, we can use hierarchical clustering techniques. The hierarchical clustering process in phyletic pattern then used to detect the evolutionary.

On this final project, it will developed an implementation of detection of evolutionary using hierarchical clustering technique based on agglomerative types. Implementation use HTML programming language, PHP script, and MySQL DBMS. The system then compares the clustering results from three agglomerative hierarchical clustering techniques : single linkage, complete linkage, and group average.

The evaluation results on all three techniques of agglomerative hierarchical clustering shows that single linkage construct better clusters. Beside that, there are some factor that effect the accuracy of clustering : number of data and phyletic pattern attribute values from dataset.

Keywords : phyletic pattern, COG, evolutionary, hierarchical clustering