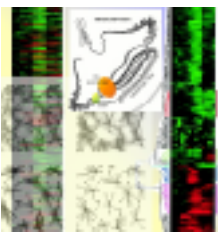


Transcriptomics and Phylogenetics Applications in Grid

Transcriptomics

Transcriptomics encapsulates the analysis of global gene expression patterns and is a key area due to the fact that the adaptation to variable conditions of a cell or an organism as well as its development and differentiation is largely determined by the profile of gene expression.

DNA array technology, allowing simultaneous large-scale expression measurements on great number of genes using minute samples, is increasingly becoming the method of choice for gene expression profiling. The applications of microarrays are deployed in research fields such as pharmacology, diagnostics and environmental monitoring.



Transcriptomics

We implemented clustering as method of gene expression data analysis and have embedded Hierarchical, Kmeans, SOM, PCA and MCL clustering techniques in the EGEE Grid infrastructure. This package is complimented by the development of a procedure to calculate the optimal amount of clusters. Further extensions include a method of robustness testing and the means to calculate a FOM to indicate if the current dataset is well suited for a specific algorithm.

Phylogenetics

The reconstruction of phylogenetic relationships between gene sequences is a crucial step towards the understanding of their evolution and is done by building phylogenetic trees that encapsulate the evolutionary information within its branches which provides a meaningful way to order large amounts of data.

In our objective to Validate an infrastructure to perform Application of Phylogenetics we have embedded the parallel version of MrBayes and a selection from the well known Phylip package into the EGEE Grid framework. We analysed a dataset newly assembled by the Department of Evolutionary Biology of the University of Siena. This dataset is the largest available so far for Pancrustacea, consisting of 100 complete (or almost complete) sequences of mitochondrial genomes.



Phylogenetics

In the frame of the project we performed phylogenetic analysis on these amino acid sequences employing parallel MrBayes taken from the BioinfoGRID portal. Our phylogenetic analysis of the mitochondrial protein coding genes confirmed the reciprocal paraphyly of Hexapoda and Crustacea. A preliminary version of the MrBayes interface has been established and is accessible through the BioinfoGRID portal.