

# Genomics Applications in Grid

The aim of this activity is to use computational Grids to analyse molecular biological data at genomic scale. Using the Grid version of the task system W3H, W3HG, software tools have been integrated in order to perform sequence alignments of cDNA, EST clustering, SNP detection, and gene prediction analysis on large sets of genomic sequences. The major achievements within this activity are:

## Analysis of the W3H task system for Grid (W3HG):

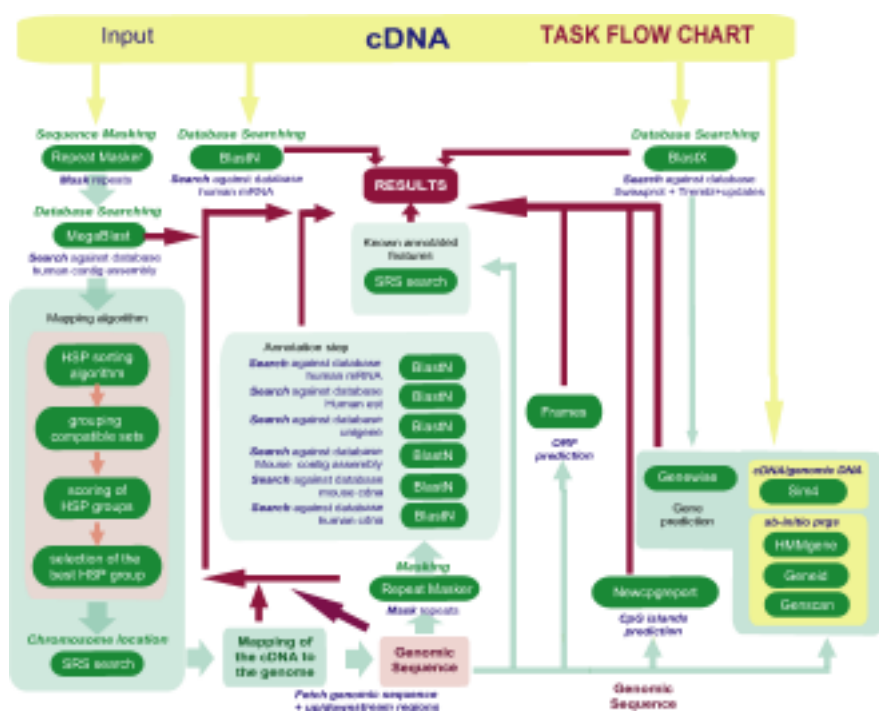
the Grid version of this system provides the basis to unify a larger number of individual bioinformatics tools into a single analytical step and transform them into the Grid environment.

## Grid analysis of cDNA data:

functional annotation of cDNAs involving search for an adequate classification system to transform the raw feature data of a given cDNA into a meaningful score reflecting the quality the given sequence. Analysis of the type of the classification system chosen to maximize sensitivity as well as specificity.

## Grid analysis of the NCBI and Ensembl databases:

integration of precomputed data from the Ensembl project into the NCBI database within the task system cDNA2Genome. Extensions to other organisms as their genomes become available. Pseudogene identification using the tool for cDNA quality. Evaluation of alternative hits in order to identify gene duplications.



Structure of the Grid analysis data flow for cDNAs

## Grid analysis of rule-based multiple alignments:

comparisons on the relative performance and reliability of these methods. Identification of the factors affecting the creation of an alignment. Analysis of the performance of the different alignment methods. Use of BALIBASE, a database of reference alignments. Extraction of general rules that can be applied in the program flow to decide which alignment methods are the most suited for each kind of sequence set.