

Poster C-21
C.A.S.S.I.O.P.E.



Authors:

lopez rascol (EA 3781)
gouret (EA 3781)
grusea (EA 3781)
danchin (AFMB)
pontarotti (EA 3781)

Short Abstract: We present here CASSIOPE, an intelligent multi-agents system which integrates concept, process and automation of genome comparison based on evolution concepts and specific mathematical models, i.e. a systematic phylogenetic analysis to determine orthologous and paralogous relationship and specific statistical tests to evaluate the significant of a region conservation.

Long Abstract:

We present, here, a process and a software which allow automated genomes comparison at the level of genes' organization along chromosomes and being based on high concepts of evolution. We are interested in the evolutionary history of vertebrates genomes to highlight different chromosomal events such as duplications, polyploidisation (whole genome duplication), translocation, inversion ... with an aim of reconstructing the whole genome of the last common ancestor of vertebrates in silico. This work is splitted in two parts: multiple genomes comparison and ancestral genome reconstruction which are very strongly dependent.

The first step consists in multiple genomes comparison to find orthologous (which indicates speciation events) or paralogous (which indicates duplication events) genomic regions which present significant conservation (i.e. an homology between two regions indicating that they arose from a common ancestral region) between all vertebrates genomes. Such analyses have already been proposed but no other approach uses a systematic phylogenetic analysis to determine orthologous and paralogous relationship and specific statistical tests to evaluate the significant of a region conservation. Moreover, because we want an as robust as possible analysis, we need to take the maximum number of available genomic information on data bases and by the way we need to automate the conserved genomic regions process. So we developed CASSIOPE, an intelligent multi-agents system which integrates concepts, process and automation. This software searches all conserved regions (both orthologous and paralogous) between vertebrate genomes (fully sequenced and assembled genomes available in Ensembl data base), it: starts with a genomic region in a specific taxa for each gene presents in this region searches all orthologous genes by a phylogenetic analysis (which gives an « history evaluation »). localizes the orthologous genes on the chromosomes of the target species defines regions of putative conservation: a statistical test is applied to evaluate the significance of the putative conservation If conservation is significant, a process is launched starting from the new significant conserved region, that allows to confirm the symmetry of the conservation and some times to widen initial regions. The system stops

when any other new region is found. This first step have been tested and validated with a specific genomic region: MHC region.

Currently, CASSIOPE is defining all vertebrates conserved regions to obtain a maximum of significantly conserved regions to begin ancestral reconstruction analysis. This will be the second step, for which concepts, mathematical models and automation will be established and integrated in CASSIOPE.