engineDB: a repository of functional analogue gene products

De Sario G[1], Tulipano A[1], Gisel A[1]

[1] Istituto di Tecnologie Biomediche, Sede Bari,CNR, Bari

Motivation

Up to now, more than 4,0 million gene products from more than 150000 different species have been described specifying their function, the processes they are involved in and their cellular localization using a very well defined and structured vocabulary, the Gene Ontology (GO, www.geneontology.org). Finding gene products which have a similar function or are involved in similar biological processes within the same or between different organisms, without relying on the conventional sequence similarity approach, is an approach to find analogous gene products, gene products which have similar functions, but not necessarily similar sequences as homologous gene products. However the comparison of those functionality according to the GO terminology is a very time consuming process.

Methods

engine (gENe analoGue fINdEr) is a tool that parallelizes the search process and distributes the calculation and data over the computational GRID, splitting the process into many sub-processes (Tulipano et al. 2007). We developed a new and more performing version of engine and a process to select the most significant functional analogous gene products calculated by engine. Further, the search results are stored within a relational database (engineDB) hosting the most important information validating the proposed functional analogy between different gene products. A graphical user interface enables the user to visualize the proposed functional analogues for his gene product under investigation ordered by the level of calculated analogy. engineDB visualizes the value of the chi-square test of independence we used for the comparison as a rating for the analogy, the GO terms of both compared gene products and the number of GO terms in common and non in common since those are the terms influencing the analogy calculation and important for the user to understand which functionalities made the gene products in comparison more or less similar.

Results

ENGINE has produced for every gene product listed within the GO database a list of potential functionally analogues within and between species using, in place of the sequence, the gene description of the GO. Those data are publically available either through a search tool as a graphical user interface (http://spank.ba.itb.cnr.it/engine/) to engineDB or as a database dump of the whole data set. The GUI offers to the end user several links to external data such as UniProt, ENSEMBL and RefSeq, download possibilities of various specific data and further information such as sequence similarity and protein domain comparison to complete the comparison data such that the end user has an overview about the proposed functional analogues. The database is updated with every upgrade of the GO database.

Contact : andreas.gisel@ba.itb.cnr.it