

# GORBI: A Web application for the prediction of a protein's functional context

Nives Škunca, *Ruđer Bošković Institute*

## Abstract

The approach of modern molecular biology to understanding complex biological systems is based on the study of many separate components. Numerous complex studies are necessary to identify the function of just one among tens of thousands of proteins in a cell. Even in the well known organisms, for a large fraction of genes, the function is still undetermined. For example, in the budding yeast *Saccharomyces cerevisiae*, about 20% of all proteins are uncharacterised, and in many other studied organisms the percentage is significantly higher. Whole genome sequencing, as one of the high throughput molecular biology techniques, provides insight into the complete genetic material of many organisms, and opens the gate for the introduction of machine learning algorithms in the analysis of biological functions.

The strategy of phylogenetic profiling is based on examining the patterns of gene co-occurrence in different genomes (Pellegrini *et al*, PNAS, 1999). The objective of the search is to recognise gene groups that are kept or lost together during evolution, which would imply their functional connection (method overview: Kenesche *et al*, J. R. Soc. Interface, 2008).

A variant of this approach has been developed in collaboration with our partners from the Jozef Štefan Institute, as a part of the Croatian – Slovenian bilateral project, entitled „Inductive Databases for Genomics and Proteomics“ (prof. Sašo Džeroski – IJS, Dr. Tomislav Šmuc - IRB).

Predictions of gene functions that resulted from this collaboration will be presented in the GORBI web application. We feel that a visualization of the results in a consistent and intuitive way will ease the interpretation and enable the choice of appropriate biological experiments with the goal of confirming the predicted function of newly characterised proteins.