

# GORBI: WEB APPLICATION FOR THE PREDICTION OF A PROTEIN'S FUNCTIONAL CONTEXT



Fran Supek<sup>1</sup>, Tomislav Šmuc<sup>1</sup>, Jelena Repar<sup>2</sup>, Matko Bošnjak<sup>1</sup>, Nives Škunca<sup>1</sup>

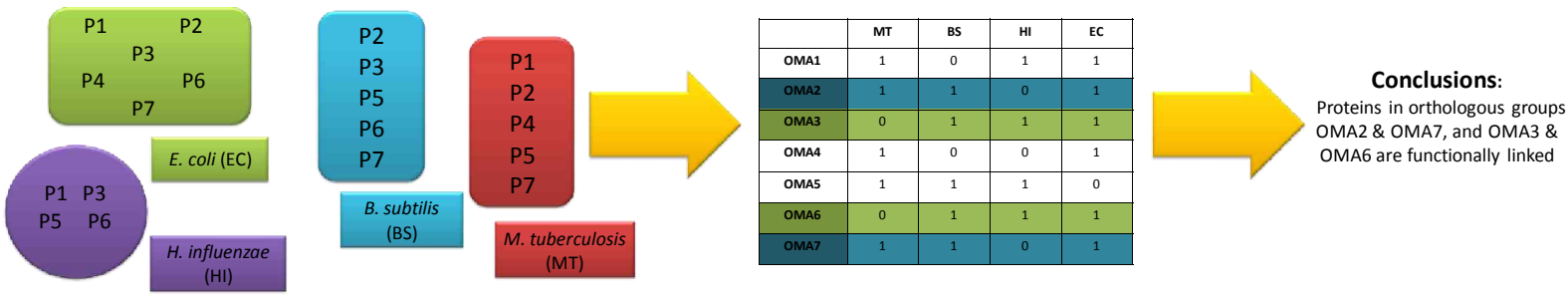
<sup>1</sup>Division of Electronics, Ruđer Bošković Institute, Bijenička cesta 54, Zagreb, Croatia

<sup>2</sup>Division of Molecular Biology, Ruđer Bošković Institute, Bijenička cesta 54, Zagreb, Croatia

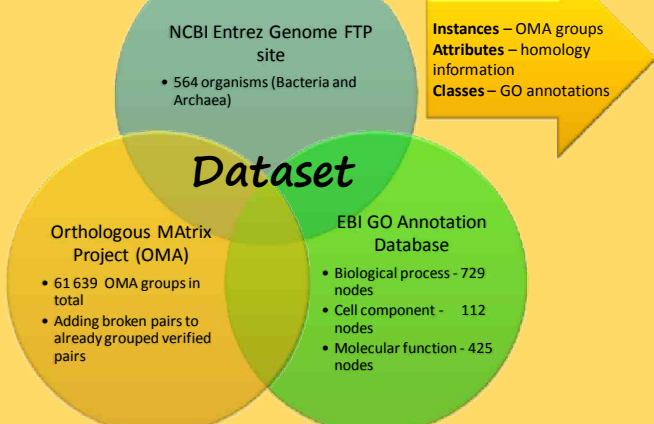
Contact: nskunca@irb.hr

## 1. Phylogenetic profiling

A method introduced by Pellegrini *et al.* is illustrated here. A modification of the original principle is used to present the dataset to the HMC learner. Each profile (row in the table) denotes a particular OMA's protein pattern of presence (1) or absence (0) in the organism (column). Two OMA groups showing similar profiles (signaling for coordinated evolutionary gain and loss) are expected to be functionally related.



## 2. Workflow



**Classifier**  
HMC – Hierarchical Multi-label Classification (Vens *et al.* 2008) - a machine learning algorithm based on decision trees that takes into account the DAG layout of GO and considerably improves computational efficiency by learning to predict all classes at once.

Visualizing the results

**GORBI** beta  
Gene Ontology @ Ruđer Bošković Institute

SEARCH METHOD ACKNOWLEDGMENT ABOUT

GORBI predicts gene function in prokaryotic genomes, currently in *Escherichia coli* K12 and *Bacillus subtilis*. GORBI relies on phylogenetic profiles of genes and hierarchical multi-label classification. For more information please [click here](#).

Get the GO ID you are interested in from <http://www.geneontology.org/>

Find your favorite Gene ID at <http://www.ncbi.nlm.nih.gov/sites/entrez>

Query Gene Ontology: example #1, #2, #3

Query protein: example #1, #2, #3

GO ID (Find GO ID from category name)

Entrez gene ID (Find Entrez gene ID from gene name)

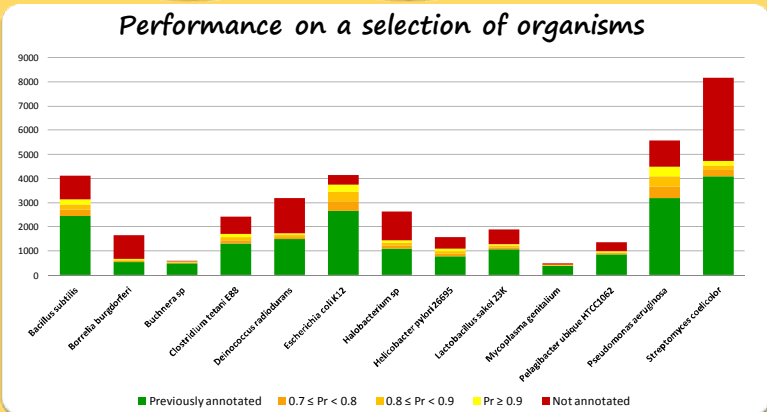
Query organism: *Escherichia coli* K12

If not selected, only GORBI predictions will be shown

Show previously known annotations  
Only the most specific GO terms will be shown. For more details, please consult AmiGO

Query! Reset

<http://gorbi.irb.hr/>



## 3. Conclusions and future perspectives

- ❖ Predicting GO assignments for genes in OMA groups, using the HMC algorithm in combination with homologous pairs produced by the OMA algorithm shows convincing predictive power.
- ❖ The model built from this data can be used as:
  1. A powerful computation tool in narrowing the search space of potential gene candidates for a particular function
  2. Annotation tool for the genes in newly sequenced genomes.
- ❖ GORBI, a user friendly search interface is being developed, providing a listing of function predictions, covering two angles in wet lab research: focusing on 1) *function* (needing prediction for genes having that function in any organism), and 2) *organism* (needing prediction of any function in this particular organism).

## References

Altenhoff, A. M. and C. Dessimoz (2009). "Phylogenetic and functional assessment of orthologs inference projects and methods." *PLoS Comput Biol* 5(1): e1000262.

Ashburner, M., C. A. Ball, *et al.* (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." *Nat Genet* 25(1): 25-9.

Kensche, P. R., V. van Noort, *et al.* (2008). "Practical and theoretical advances in predicting the function of a protein by its phylogenetic distribution." *Journal of the Royal Society Interface* 5(19): 151-170.

Roth, A. C., G. H. Gonnet, *et al.* (2008). "Algorithm of OMA for large-scale orthology inference." *BMC Bioinformatics* 9: 518-528.

Vens, C., J. Struyf, *et al.* (2008). "Decision trees for hierarchical multi-label classification." *Machine Learning* 73(2): 185-214.

## Acknowledgements

The work presented here is supported by the Croatian Ministry of Science, Education and Sports iProjekti scheme. The authors acknowledge the funding provided by **The 3rd Adriatic Meeting On Computational Solutions in the Life Sciences** organizers for the attendance of the conference.