



**HAL**  
open science

## **WHOOOPER : Web application for Hands-On identification of protein co-Occurrence among Phyla, focused on user ERgonomics.**

Sylvain Marthey, Natacha Baffo, María-Natalia Lisa, Gwenaëlle André

### ► **To cite this version:**

Sylvain Marthey, Natacha Baffo, María-Natalia Lisa, Gwenaëlle André. WHOOPER : Web application for Hands-On identification of protein co-Occurrence among Phyla, focused on user ERgonomics.. JOBIM, SFBI, Aug 2025, Bordeaux, France. <hal-05198396>

**HAL Id: hal-05198396**

**<https://hal.inrae.fr/hal-05198396v1>**

Submitted on 4 Aug 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# WHOPPER : Web application for Hands-On identification of protein co-Occurrence among Phyla, focused on user ERgonomics.

Sylvain MARTHEY<sup>1</sup>, Natacha BAFFO<sup>2</sup>, María-Natalia LISA<sup>2,3</sup> and Gwenaëlle ANDRÉ<sup>1</sup>

1 Université Paris-Saclay, INRAE, MaIAGE, Jouy-en-Josas, France

2 Instituto de Biología Molecular y Celular de Rosario (IBR CONICET-UNR), Ocampo y Esmeralda, Rosario S2002LRK, Argentina

3 Plataforma de Biología Estructural y Metabólica (PLABEM), Ocampo y Esmeralda, Rosario S2002LRK, Argentina.

Corresponding Author: sylvain.marthey@inrae.fr

**Tool reference:** <https://whooper.migale.inrae.fr/>

## Keywords

web application, remote homologs identification, evolution, user ergonomics, data visualization.

## Abstract

Thinking the evolution of biological functions across different taxa requires studying the distribution and the prevalence of the functional units involved such as pathways, proteins, domains or folds, their co-occurrence within genomes and their co-evolution. While many tools exist to search for distant protein homologs: BLAST[1], HHblits [2], HMMER[3], Foldseek[4], the web servers handling these tools [5][6][7][4] neither allow the joint analysis of several proteins at the same time, nor an in-depth exploration through dynamic filtering, therefore they are not suitable for an easy, fast and accurate identification of co-occurrences of protein's homologs in genomes. Thus, we created the WHOPPER web application which combines the remote homologs detection power of HHsearch[8] with an interactive, ergonomic and innovative results exploitation environment, and follows the UX design principles of the Visual Information-Seeking Mantra: "Overview first, zoom and filter, then details-on-demand". [9]. Aligned with that vision, WHOPPER allows the user to: (i) align proteins of interest against hundreds of proteomes (ii) jointly visualize the massive results of alignments of these proteins against all proteomes through an innovative graphical view presenting the hits in their phylogenetic context (iii) use dynamic filtering of the results and navigation functionalities using contextual information and metadata of the hits (phylogeny, % identity, size, genomic localization, annotation, etc...) (iv) select and export results of interest in various formats to facilitate post-analysis. The application's code is accessible at: <https://github.com/SmartBioInf/whooper>. A public instance allowing the screening of more than 7500 reference proteomes from eukaryotes, bacteria, and archaea is available at: <https://whooper.migale.inrae.fr/>

## References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990 Oct 5;215(3):403-10.
2. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods.* 2011 Dec 25;9(2):173-5.
3. Eddy SR: Profile hidden Markov models. *Bioinformatics.* 1998, 14: 755-763. [10.1093/bioinformatics/14.9.755](https://doi.org/10.1093/bioinformatics/14.9.755).
4. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, Steinegger M. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol.* 2024 Feb;42(2):243-246
5. Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res.* 2008 Jul 1;36(Web Server issue):W5-9.
6. Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva V. Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinformatics.* 2020 Dec;72(1):e108.
7. Robert D. Finn, Jody Clements, Sean R. Eddy, HMMER web server: interactive sequence similarity searching, *Nucleic Acids Research*, Volume 39, Issue suppl\_2, 1 July 2011, Pages W29–W37.
8. Johannes Söding, Protein homology detection by HMM–HMM comparison, *Bioinformatics*, Volume 21, Issue 7, April 2005, Pages 951–960.
9. B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations," *Proceedings 1996 IEEE Symposium on Visual Languages*, Boulder, CO, USA, 1996.