CSIC-ELIXIR portfolio: Bioinformatic software from CSIC in European infrastructure

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Fabián Robledo-Yagüe[1,2], Mónica Chagoyen[1,3], Jose María Carazo[1,3], Bruno Contreras-Moreira[1,4] Jaime Huerta-Cepa[1,5], Javier de las Rivas[1,6], Ana Conesa[1,2]

[1] Consejo Superior de Investigaciones Científicas (CSIC)
[2] Instituto de Biología Integrativa de Sistemas (I2SysBio), Valencia, España
[3] Centro Nacional de Biotecnología (CNB), Madrid, España
[4] Estación Experimental Aula Dei (EEAD), Zaragoza, España
[5] Centro de Biotecnología y Genómica de Plantas (CBGP), Madrid, España
[6] Instituto de Biología Molecular y Celular del Cáncer (IBMCC), Salamanca, España

iNB
Spanish National Bioinformatics Institute

eliXir
SPAIN

What is ELIXIR?

ELIXIR EU Research Infrastructure (ESFRI) brings together life science informatics resources from across Europe, to maintain key molecular databases, analysis tools and foster collaborations. CSIC contributes with 6 groups and 10 resources.

INFRAS2024 to boost CSIC participation in ELIXIR

INFRAS2024 project aims to increase the overall quality of the CSIC-ELIXIR portfolio by improving the tools' robustness and user experience. We will support other CSIC tools to become part of the ELIXIR portfolio. Please use this QR code to submit the Expression of Interest for your tool!

Genomics of Gene Expression Lab - I2SysBio

SQANTI3 is designed to enable quality control and filtering of long read-defined transcriptomes, often rich in artifacts and false-positive isoforms. A good curation of the transcriptome is indispensable to proceed with FIT analysis and produce valid, biologically sound conclusions and hypothesis.

INFRAS2024 focused on increasing final user experience, by containerizing SQANTI3, increasing the range of input formats, facilitating tool usage, benchmarking the tool time and memory usage on HPCs and general bug fixing.

SQANTIQ3

Genomics of Gene Expression Lab

Computational Systems Biology Group - CNB

MBROLE3 performs functional enrichment analysis for a set of chemical compounds. MBROLE3 integrates a large amount of biological and chemical information related to chemical compounds extracted from multiple databases.

INFRAS2024 updated the databases and improved the conversor to include new sources of IDs to improve the compatibility with different databases and increase the number of compounds by supporting standard InChiKeys

MBROLE 3
Metabolites Biological Role

CSBG

Computational and Structural Biology Group - EEAD

Regulatory Sequence analysis Tool (RSAT) is a web-platform hosting organisms genomes and tools for sequence analysis, motif discovery and analysis, comparative genomics and more. The Computational and Structural Biology Group hosts the RSAT plants server, optimized for these organisms, and also provides other tools to the portfolio, such as GET_HOMOLOGUES, for pan-genome sequence analysis

INFRAS2024 improved and updated the installation methods available for RSAT and added several enhancements for the admins of the tools, like deployment in Apptainer, scripts, documentation and automation.

RSAT Plants

Comparative Genomics and Metagenomics Group - CBGP

EggNOG, a database of orthology relationships, functional annotation, and gene evolutionary histories, holds data for 12 000 organisms with 17 million orthologous groups. Together with EggNOG-mapper, a tool for fast functional annotation of novel sequences, allowing for powerful ontology analysis.

The Comparative Genomics and Metagenomics Group also provides a tools for the study of phylogenies, **ETE toolkit**, which automates manipulation, analysis and visualization of phylogenetic trees.

NOG EggNOG 6.0.0

Biocomputing Unit - CNB

3DBIONOTES-WS automatically annotates biomedical and biochemical information in structural models, like post-transcriptional modifications, genomic variations associated with diseases, linear motifs, antigenic epitopes, disordered regions and domain families.

The Unit also develops tools for 3D microscope image processing, **Scipion** and **XMIPP**, for 3D models of macromolecules and single-particle 3D electron microscopy respectively.
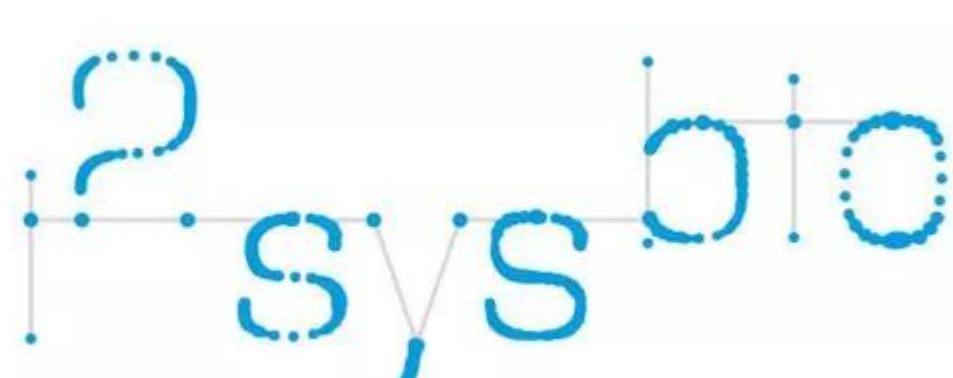
3DBIONOTES-WS

BIO COMPUTING UNIT

Bioinformatics and Functional Genomics Group - IBMCC

The **APID** server host a wide database of interacting proteins for several organisms. APID can search interacting proteins according to different level of experimental evidences for hundreds of organisms with data from primary databases, with a compendium of 90 000 proteins and 678 000 interactions.

The platform also provides complete interactomes of multiple species, including information on the experimental evidences, methods, and publications that validate each protein-protein interaction; as well as a complementary tool to build protein interaction networks obtained for a list of query proteins

Agile **P**rotein **I**nteractomes **D**ataServer

i2sysbio

CNB
CENTRO NACIONAL DE BIOTECNOLOGIA

EEAD
Estación Experimental de Aula Dei - CSIC

CBGP
CENTRO DE BIOTECNOLOGÍA Y GENÓMICA DE PLANTAS
UPM-INIA

25 anniversary
CIC Salamanca