OSCARS
Open Science Clusters' Action
for Research & Society

Funded Project

FAIR AI models for functional annotation of biodiversity genomics resources

Presenter: ANA M. ROJAS, CABD/CSIC, 0000-0003-0750-9099

Implemented by

Centro
Andaluz
de Biología
del Desarrollo

Metazoa
Phylogenomics
Lab

ibmcp

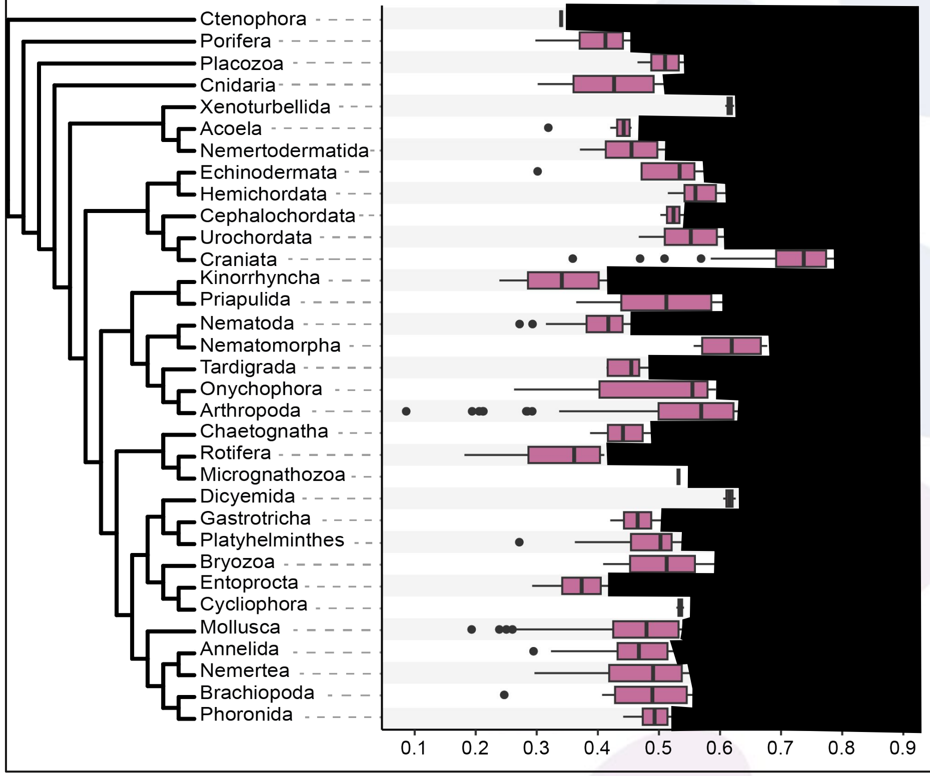
CSIC
Consejo Superior de Investigaciones Científicas

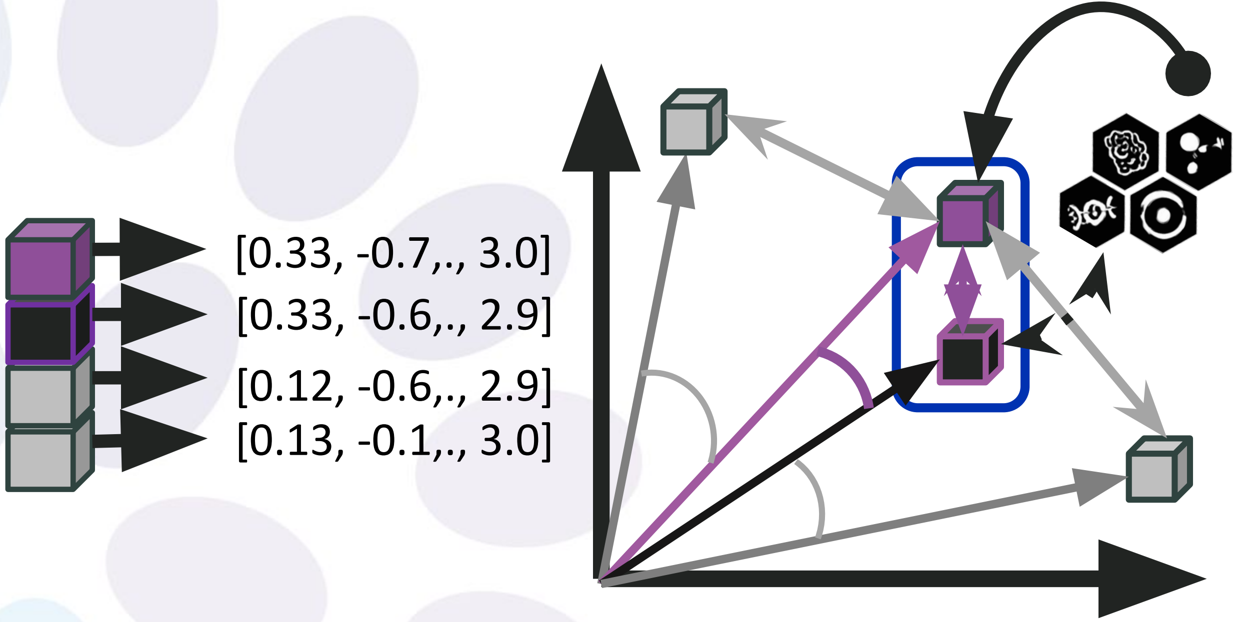
eosc

Funded by
the European Union

CHALLENGE ADDRESSED

- **AI-driven annotation** of the **uncharacterized coding fraction** of genomes from biodiversity projects using protein **embeddings**.





CHALLENGE ADDRESSED

OSCARS

- AI-driven annotation of the uncharacterized coding fraction of genomes from biodiversity projects using protein embeddings.

- Identification of **conserved lineage-specific** features

- Develop a **FAIR and scalable** sequence-based **functional prediction method**

Fantasia

SOLUTION

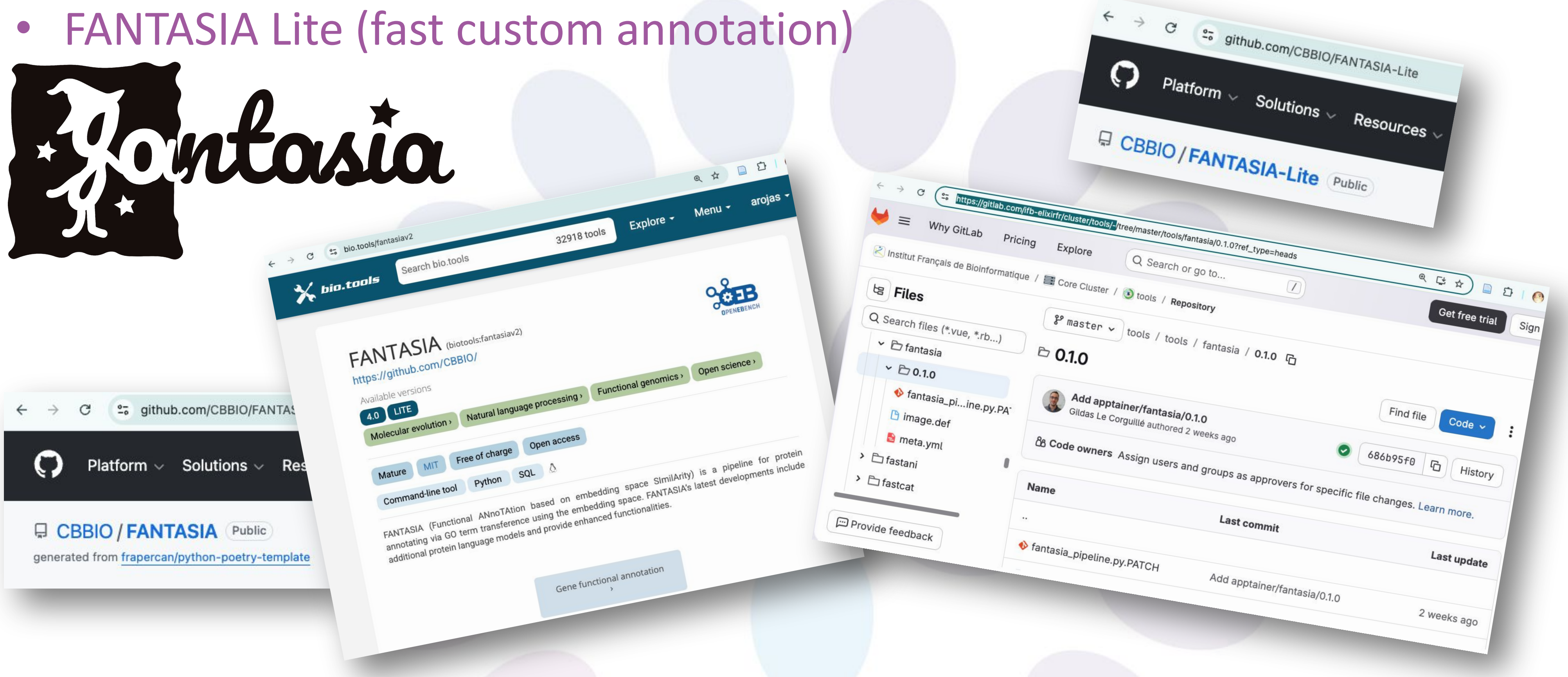
OSCARS

- **FANTASIA development**

- To accommodate different user needs (i.e. fast and quick or fully exploitable)
- To make it FAIR and fully accessible to the community

RESULTS

- **Development** and implementation of a FAIR workflow to perform
- FANTASIA full (reference data control and fully parametrizable)
- FANTASIA Lite (fast custom annotation)



github.com/CBBIO/FANTASIA-Lite

CBBIO / FANTASIA-Lite Public

github.com/CBBIO/FANTAS

CBBIO / FANTASIA Public

generated from frapercan/python-poetry-template

bio.tools/fantasiav2

FANTASIA (biotools:fantasiav2)

https://github.com/CBBIO/

Available versions

4.0 LITE

Molecular evolution › Natural language processing › Functional genomics › Open science ›

Mature MIT Free of charge Open access

Command-line tool Python SQL

FANTASIA (Functional ANnoTAtion based on embedding space SImilArity) is a pipeline for protein annotating via GO term transference using the embedding space. FANTASIA's latest developments include additional protein language models and provide enhanced functionalities.

Gene functional annotation ›

https://gitlab.com/ifb-elixirfr/cluster/tools/-/tree/master/tools/fantasia/0.1.0?ref_type=heads

Institut Français de Bioinformatique / Core Cluster / tools / Repository

Files

- fantasia
  - 0.1.0
    - fantasia_pi...ine.py.PA
    - image.def
    - meta.yml
- fastani
- fastcat

master tools / tools / fantasia / 0.1.0

0.1.0

Add apptainer/fantasia/0.1.0

Gildas Le Corguillé authored 2 weeks ago

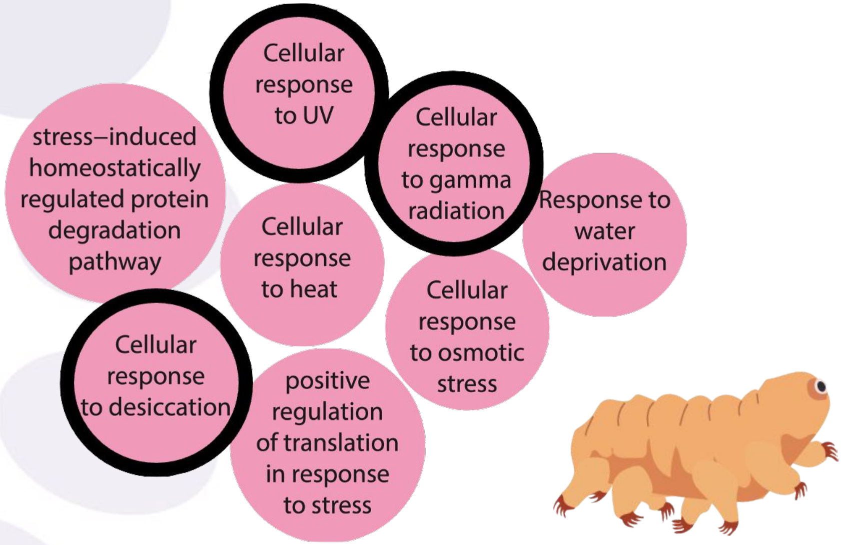
Code owners Assign users and groups as approvers for specific file changes. Learn more.

Name	Last commit	Last update
..		
fantasia_pipeline.py.PATCH	Add apptainer/fantasia/0.1.0	2 weeks ago

RESULTS

- **Scientific dissemination**:

- About the method implementation (publication: 10.1007/978-1-0716-4623-6_8)
- **Biodiversity** settings: ~24M genes across ~1,000 species (10.1038/s42003-025-08651-2)
- Applicable to noise detection in model organisms (plant benchmark), 200 plant genomes are annotated (to be released soon)



Cellular response to UV

stress–induced homeostatically regulated protein degradation pathway

Cellular response to gamma radiation

Response to water deprivation

Cellular response to heat

Cellular response to osmotic stress

Cellular response to desiccation

positive regulation of translation in response to stress

RESULTS

- **Exploitation and usage by the community**
  - EarthBiogenome https://www.earthbiogenome.org/report-on-annotation-recommended-tools
  - Repository: https://gitlab.com/ifb-elixirfr/cluster/
  - Bio.tools: https://bio.tools/fantasiav2

- **ONGOING…**
  - Enhanced applicability, expansion modules
  - Fully and massive exploration of latent spaces
  - Increasing the network reach, >5 active collaborations around the world

- **Outreach** actions
  - General media (National press, National radio with 2 events)
  - Scientific outreach

TEAM

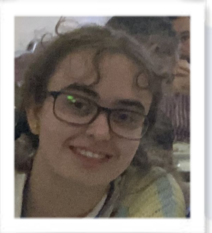
Fran Pérez
Comp. Sci, AI

Àlex Domínguez
Comp. Sci, AI

Ildefonso Cases
Comp. Bio
0000-0003-0750-9099

Ana M. Rojas
Comp. Bio.
0000-0003-0750-9099

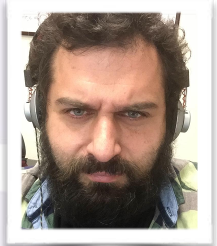
Belen Carbonetto
Genomics, Bioinformatics
0000-0002-2784-5324



Gemma Martínez
Genomics, Bioinformatics
0000-0001-5633-8690

Rosa Fernández*
Genomics, Biodiversity
0000-0002-4719-6640

Iñigo de Martin-Agirre
Genomics, Bioinformatics
0009-0006-0374-6504



Aureliano Bombarely
Genomics,
Bioinformatics
0000-0001-6257-8914