

# INSTITUT PASTEUR | PARIS

06>09 JUIL

· IIIIar

202

# Proceedings > Posters



## AgroLD: A Knowledge Graph Database for plant functional genomics

Pierre Larmande,<sup>1,2</sup> Ndomassi Tando,<sup>1,2</sup> Bertrand Pitollat,<sup>2,3</sup> Valentin Guignon,<sup>2,4</sup> Mathieu Rouard,<sup>2,4</sup> Gaetan Droc,<sup>2,3</sup> Manuel Ruiz<sup>2,3</sup>

1 - DIADE, IRD, Univ. Montpellier, 911 av Agropolis, 34398 Montpellier, France 2 - French Institute of Bioinformatics (IFB)-South Green Bioinformatics Platform, Bioversity, CIRAD, INRAE, IRD, F-34398 Montpellier, France 3 - AGAP, CIRAD, INRAE, Univ. Montpellier, av Agropolis, 34398 Montpellier, France 4 - Bioversity International, Parc Scientifique Agropolis II, 34397 Montpellier, France

Corresponding Author: pierre.larmande@ird.fr

Exploring the links between genetic and phenotypic traits is an important area of research in agronomy. One of the main objectives of this is to accelerate the development of important traits that can positively impact the agricultural economy. However, due to the existence of complex molecular interactions, to gain complete understanding will warrant data analyses performed at different molecular and environmental levels for a given (plant) subject. For instance, to understand how rice genes involved in metabolism or signaling of growth regulators control the rice panicle architecture. While high-throughput technologies have played a key role in accelerating and generating the much-needed data, these can only partially capture the dynamics in genotypephenotype relations. Consequently, our knowledge of the complex relationships between the different molecular actors responsible for the expression of the phenome in various plant systems remains fragmented. Hence, there is an urgent need to effectively integrate and assimilate complementary information to understand the biological system in its entirety.

We have developed AgroLD [1] (www.agrold.org), a knowledge graph system that exploits the Semantic Web technology and FAIR principles [2], to integrate information to integrate data about plant species of high interest for the plant science community e.g., rice, wheat, Arabidopsis and in this way facilitating the formulation of new scientific hypotheses. We present some integration results of the project, which currently focused on genomics, proteomics and phenomics. AgroLD is now an RDF knowledge base of 900M triples created by annotating and integrating more than 100 datasets coming from 15 data sources –such as Ensembl plants [3], Gramene.org [4] and TropGeneDB [5]- with 15 ontologies -such as the Gene Ontology [6] and Plant Ontology [7]. Our objective is to offer a domain specific knowledge platform to solve complex biological and agronomical questions related to the implication of genes in, for instances, plant disease resistance or high yield traits. We expect the resolution of these questions to facilitate the formulation of new scientific hypotheses to be validated with a knowledge-oriented approach.

### Acknowledgements

Authors thank the technical staff of the South Green Bioinformatics platform for their support.

### References

1. Venkatesan A, Tagny Ngompe G, Hassouni NE, Chentli I, Guignon V, Jonquet C, et al. Agronomic Linked Data (AgroLD): a Knowledge-based System to Enable Integrative Biology in Agronomy. PLoS ONE. 2018;:13:17. 2. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles

for scientific data management and stewardship. Sci Data. 2016;3.

3. Bolser D, Staines DM, Pritchard E, Kersey P. Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. Methods Mol Biol Clifton NJ. 2016;1374:115-40.

4. Tello-Ruiz MK, Naithani S, Stein JC, Gupta P, Campbell M, Olson A, et al. Gramene 2018: Unifying comparative genomics and pathway resources for plant research. Nucleic Acids Res. 2018.

5. Hamelin C, Sempere G, Jouffe V, Ruiz M. TropGeneDB, the multi-tropical crop information system updated and extended. Nucleic Acids Res. 2013;41.

6. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015;43:D1049-56. 7. Cooper L, Walls RL, Elser J, Gandolfo MA, Stevenson DW, Smith B, et al. The plant ontology as a tool for comparative plant anatomy and genomic analyses. Plant Cell Physiol. 2013;54:e1.