GreenPhylDB v2.0: An improved database for plant functional Genomics

Conte M.G[1], Laporte M.A[2], Aluome C[3], Perin C[3], Rouard M[1]

[1] Bioversity International - Commodities for Livelihoods programme Parc Scientifique Agropolis II, 34397Montpellier - Cedex 5, France
[2] CEFE - UMR 5175 - 1919 Route de Mende - F34293 Montpellier cedex 5, France
[3] CIRAD, Department BIOS, UMR DAP - TA40/03, 34398 Montpellier, France

Project–SP4 G4008.21: Large scale phylogenomic analyses to gene function prediction for GCP crops (PI: Mathieu Rouard)

GreenPhylDB is a resource of comparative genomic for plant functional genomics. Here, we present the version 2 which contains a catalogue of gene families for 15 full genomes of plants covering a broad taxonomy of green plants. Results of our automatic clustering is first manually annotated and then analyzed by a phylogeny approach to predict orthologs. This ortholog prediction is particularly useful for functional genomics and candidate gene identification. We describe as well our update platform that allows us to integrate and maintain our database up to date with new genome releases.

From orthologs prediction to function
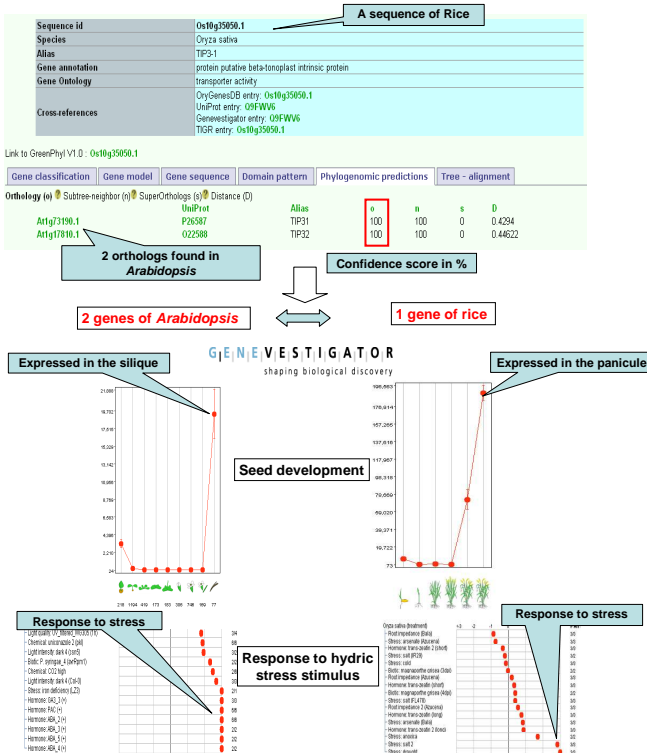
Rationale: The main assumption behind orthologs identification is the conservation of genes functions (at least sharing a similar function) across evolution of closely related species.

Objective: Help biologists with functional validation.

Description: we crossed references our orthologs to *Genevestigator* which is a reference expression database allowing studies of gene-regulation in a wide variety of contexts, such as plants under stress .

Data update platform
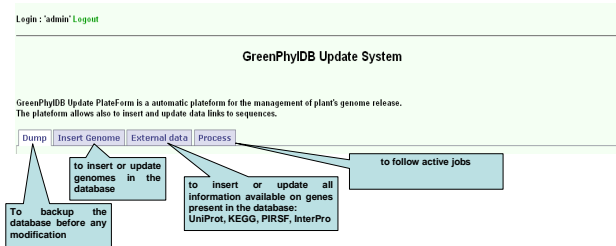
Rationale: Sequencing and annotation projects generate regularly new or updated releases of genome annotation. Moreover, data associated to each gene (domains, classification, cross references…) is then also susceptible to change.

Objective: Provide a regularly updated database totally linked to external data.

Description: This structure grant database administrator to insert or update data in our system. The system allow to track changes in the database.

New genomes
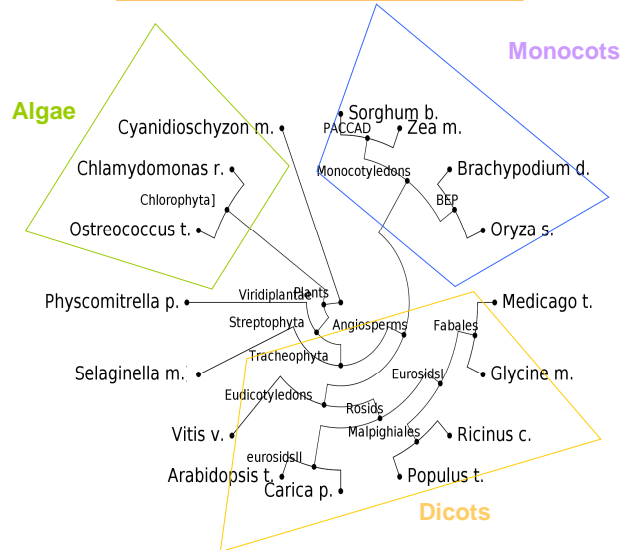
Rationale: Fast progress in sequencing and annotation projects has enabled us to work now on 15 full plant genomes including economically and socially important crops like rice, sorghum and maize

Objectives: Integrate in our database all genomes fully annotated with relatively good annotation to take benefit of all information available on new species

Description: This taxonomy coverage is also useful for accurate orthologs prediction through phylogenetics. These plant genomes should help us to unravel some aspects of green plant evolution.

Family annotation platform
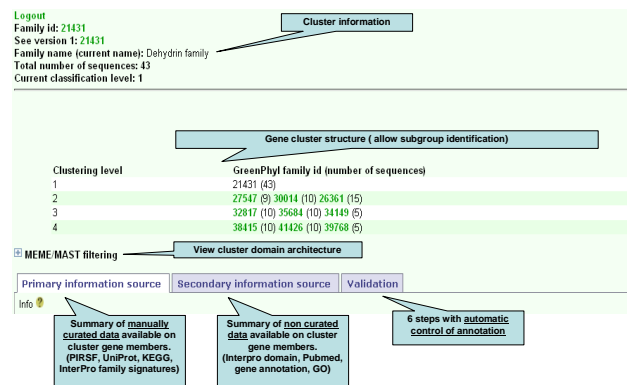
Rationale: Identify proper catalogue of homeomorphic plant gene families - genes that evolved from a common ancestor and sharing full-length sequence similarity and common domain architecture - is a valuable resource for evolution studies, future gene annotation, and orthologs inference.

Objective: Provide a comprehensive list a plant gene families with manually curation

Description: This structure allows granted users to annotate gene clusters that were automatically generated. This platform proposes statistics of the different data available on gene clusters members.

References: (http://greenphyl.cirad.fr)

Conte MG, Gaillard S, Droc G, Périn C (2008). Phylogenomics of plant genomes: a methodology for genome-wide searches for orthologs in plants. BMC Genomics, 2008, 183

Conte MG, Gaillard S, Lanau N, Rouard M, Périn C (2008). GreenPhylDB: a database for plant comparative genomics. Nucleic Acids Research. 2008 January; 36 D991–D998