

Poster B-51
Arabidopsis Functional Genomics
Data Integration



Authors:

Nicolas Tsesmetzis (*Computational Biology Department, John Innes Centre*)

Matthew Couchman (*Computational Biology Department, John Innes Centre*)

Sean Walsh (*Computational Biology Department, John Innes Centre*)

Short Abstract: We are currently working on two systems: i) The Arabidopsis thaliana Integration Database (<http://atidb.org>) integrates the latest gene models and annotation with the location of insertional mutagens, the Gene Ontology Database, InterProScan results, primer locations etc. ii) The Arabidopsis Reactome is a knowledge base that models biological processes as a series of events and it is based on the human Reactome data-structure (<http://www.reactome.org>).

Long Abstract:

The integration of molecular biology data on a DNA backbone by reference to biological entities encoded therein is a common theme in bioinformatics. These database systems generally provide genome browsing capabilities, integrated tools and programming interfaces that are utilised by bench biologists and computational biologists alike. Continued development of these systems to include emerging datasets is essential and here we describe a genome database for Arabidopsis. To complement the genome database we have embarked upon the creation of a knowledge base which present the inner workings of the cell from the perspective of the biological phenomena.

Genome View

The ATIDB system (<http://atidb.org>) for Arabidopsis is built with Genome Browser from the Generic Model Organisms Database Project (GMOD, <http://www.gmod.org>). The system integrates the Arabidopsis Genome gene models and annotation with the location of insertional mutagens (transposons and T-DNAs), the Gene Ontology Database, small RNA locations, InterProScan results, primer locations etc. Apart from providing a browsing and search interface, the system is distributed with a powerful scripting interface which allows sophisticated queries and reporting. We have used this facility to answer numerous questions from JIC scientists and international collaborators. We aim to maintain this tool and extend it in the near future to include whole genome tiling array results, promoter location and protein-protein interactions.

Biological Phenomena View

Reactome (<http://www.reactome.org>) is a knowledge-base that models biological processes as a series of events. Events include catalysis, complex formation, inhibition etc. are organised into pathways which allow cell signaling, genetic regulation and metabolic processes to be captured. Our Reactome instances for Arabidopsis have been seeded with orthologous events from human using pair-wise best hit computational approaches (OrthoMCL, Inparanoid). These electronic inferences are then being confirmed or modified based on experimental evidence from our target species (or from other species within the

same kingdom). The newly created pathways are then passed onto expert biologists for inspection.

Our Reactome instance will have many scientific applications once fully established:

- As a reference on the state of knowledge
- As a reference to working hypothesis (with the appropriate implementation of evidence ontologies)
- For overlaying “-omics” data in order to visualise sets of entities (genes, proteins, metabolites) involved in certain processes
- Establish a high quality reference genome for the rapid annotation of function in newly sequenced plant genomes
- As a basis for systems biology applications (parameter estimation in metabolic flux analysis, search for common network motifs etc).

Reactome and the genome database are available under open source terms.

To fully develop Reactome’s potential, the engagement and contribution of the scientific community is essential.