

P0932

Function Prediction in Plant Genomes from Large Scale Phylogenomics Analyses

Date: Monday, January 14, 2013

Room: Grand Exhibit Hall

Valentin Guignon , Bioversity International, Montpellier, France

Jean-François Dufayard , CIRAD - UMR AGAP, Montpellier, France

Sebastien Briois , Syngenta Seeds, Toulouse, France

Matthieu G. Conte , Syngenta Seeds, Toulouse, France

Mathieu Rouard , Bioversity International, Montpellier, France

With the increasing number of plant genomes being sequenced, a major challenge is to accurately transfer annotations from well characterized genomes to newly obtained sequences. GreenPhylDB is a database designed for comparative and functional genomics based on complete genome-derived gene sequences. The database currently includes gene families of protein sequences from 22 plant species, including socio-economically important crops like rice, sorghum, maize, cassava and banana. Genes from all these species are organized in clusters based on sequence similarity. The clusters are manually annotated (i.e. properly named and classified) and sequences included in each cluster are characterized by phylogenetic analysis in order to elucidate evolutionary relationships (e.g. orthologs, super-orthologs, in/out-paralogs) among genes. GreenPhyl provides a reliable and stable catalog of gene families useful for annotation on new genome sequences in plants. With its improved user interface, the new release of GreenPhyl keeps the previous gene clustering quality and introduces additional features such as specific search engines (quick search, deep search, InterPro domain combination and GO family browser). The GreenPhyl's pipeline relies on RapGreen, a new version of the RAP reconciliation tool (Dufayard et al, 2005) that allows us to root gene trees and infer orthology relationships between sequences of a family. GreenPhyl version 3 is available at <http://www.greenphyl.org> and is a collaborative resource of SouthGreen (southgreen.cirad.fr), a bioinformatics platform applied to the genetic and genomic resources analyses of the South and Mediterranean plants.