

Poster L-11

Development of a Rice Genome Database for the Interpretation of Cereal Microarray Data



Authors:

Philip Law (*Bioinformatics and Computational Biology Unit, University of Pretoria*)

Clotilde Claudel-Renard (*Bioinformatics and Computational Biology Unit, University of Pretoria*)

Fourie Joubert (*Bioinformatics and Computational Biology Unit, University of Pretoria*)

Dave Berger (*Department of Botany, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria*)

Short Abstract: The MADIBA (MicroArray Data Interface for Biological Annotation) toolkit identifies GO terms, metabolic pathways, transcriptional co-regulation and genomic localisations, in clusters of Plasmodium or rice genes. These are identified using microarray experiments showing significant expression variation. Planned improvements include a pre-processing option and plant defence maps.

Long Abstract:

Rice (*Oryza sativa*) is an important crop, not only as food source for approximately half the world's population, but also as a model organism for monocotyledons, in particular cereals. With the sequence of the rice genome completed, this not only provides a base for crop improvements, but also a means to study the synteny between the various cereal genomes (including rice, millet, wheat and maize) and possibly lead to the elucidation of the functions of orthologous genes and proteins.

Microarrays have made it possible to investigate the transcriptional changes in thousands of genes, all at once. However, such an approach generates large gene lists, and after standard post-processing steps, such as normalisation and clustering, several sets of genes can be found. Still, an additional step is required to determine the biological meaning of these gene sets. Many tools have been developed in an attempt to solve this problem. One such tool is MADIBA (MicroArray Data Interface for Biological Annotation), developed at the University of Pretoria, South Africa. MADIBA consists of a relational database and a web interface with tools to: 1) identify Gene Ontology (GO) terms; 2) visualise the metabolic pathways where the genes are implicated; 3) identify sets of genes that are possibly co-ordinated by a common set of regulators; and 4) visualise the genes' genomic localisations. Although MADIBA was initially built for Plasmodium microarray data, it was the aim of this project to extend the program to include rice data, and potentially other plant species, such as Arabidopsis.

Rice genomic data and annotations were obtained from The Institute for Genome Research (TIGR), and the data incorporated into MADIBA's database. A BLASTX interface was added to allow the use of other cereals' genes.

The functionality of MADIBA was tested on published data by Fujiwara et al, 2004, and found to give similar results with regard to the clusters' annotations. The initial results indicated that the rice data were successfully incorporated into MADIBA and that functionality was maintained.

Further improvements that are in development include a pre-processing option, using the R

program SSHscreen (<http://www.stats.ox.ac.uk/~vos/SSHscreen>) to identify genes that are significantly differentially expressed and to determine if they were rare or abundant in the original sample. Another feature, specifically for plant genomes, is the development of a plant defence map, where the genes in the cluster are highlighted on a map to ascertain whether the genes are involved in a particular known defence pathway.