

MapMen

the Future, building tools for integrative crop analysis

Nagel A.¹, Usadel B.¹, Gibon Y.¹, Blaesing O.E.^{1,2}, Thimm O.^{1,2}, Hannemann J.¹, Lodka D.¹, Krüger P.¹, Selbig J.³, Walther D.¹, Kersten B.¹, Stitt M.¹

1 Max-Planck Institute of Molecular Plant Physiology, Am Muehlenberg 1, D-14476 Golm

2 Present Address: Metanomics, Tegeler Weg 33, 10589 Berlin

3 Universität Potsdam

<http://www.gabipd.de/projects/MapMan>

Introduction

The explosion in the use of microarray technology and current advances in genome sequencing in crop plants and model organisms requires new ways to tackle the generated data. We propose to extend MapMan by adding further tools to deal with microarray data, starting on the user side with a tool to facilitate experiment design (MiaMan), and a tool to assess the quality of microarrays and statistically determine differential expression (Robin). Additionally we will build new tools to help quicker classification of new unigene builds in crop plants or to support visualization in hitherto unknown crop plants (Scavenger F). Finally the data that has been analysed with Robin is stored with statistical information in a database for further reference.

Robin

MapMan as a desktop application was developed for the analysis and condensed visualization of large datasets. Whereas typically microarray experiments encompassed very few slides and simple statistics such as t-tests, publication standards for microarrays have risen and novel ways to evaluate data have been developed. Most notably, with the BioConductor framework of R, powerful tools became available for microarray analysis. Unfortunately, installing R and BioConductor is not easy and later assessment requires in most cases learning the R programming language.

To enable the biologist to use R/BioConductor without any further needs, we are currently developing Robin, MapMan's helper, (<http://bioinformatics.mpimp-golm.mpg.de/projects/own/robin/>). Robin aids in microarray data quality control as well as statistical evaluation. An early technology demo release version of Robin already features:

- RNA degradation plots
- Intensity plots
- Boxplots
- Scatterplots of microarray slides
- Weighting image plots showing artefacts
- Moderated t-statistics
- A multiplicity of measures to tackle the multiple testing problem (Benjamini-Hochberg, Benjamini-Yekutieli, Hommel, Bonferroni,...)

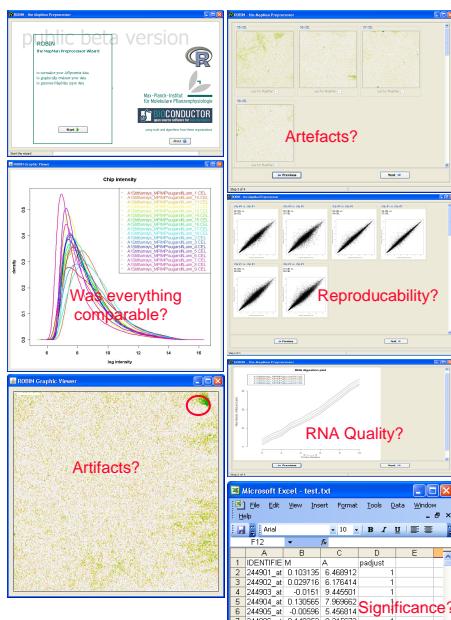


Figure 2: Affymetrix microarray analysis including quality and assessment for differential expression.

XemlLab

Currently, microarray deposition requires the user to supply Minimal Information about Microarray Experiments (MIAME), but often enough this does not even describe plant experiments well. Therefore, a plant extension MIAME-Plant has been proposed which is starting to become accepted. However, currently it is difficult to annotate an experiment with such descriptions and on the other hand it is hard to immediately understand such experiments, given the raw data and/or description. For this reason we are developing a tool to annotate, store and visualize (not necessarily microarray) experiment designs (MiaMan). Here annotation is supported using a graphical interface, so that experiments can be described based on mouse interactions, instead of lengthy manual sessions.



Figure 1: MiaMan showing an experiment design including harvesting time points (green arrow heads)

Mercator

To facilitate automatic generation of MapMan classes for unknown sequences (e.g. from EST collections of new full genome sequences from crops) we are building a tool kit that integrates many different information resources to classify sequences. The search starts with similarity information to reference genomes such as Arabidopsis, adds reference information from Swiss-Prot, then queries known, well-classified protein families and domains, and finally looks for key-words in similarity searches.

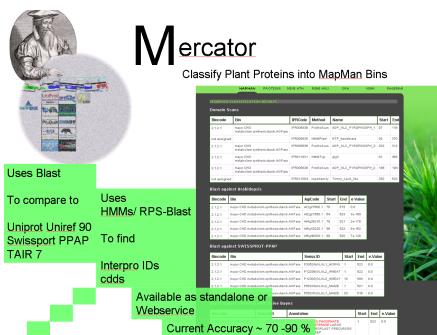


Figure 3: Auto Classification of an "unknown" AGPase from barley



Figure 4: KEGG data harvester

As an extension, we will also develop a pathway importer for pathways from KEGG which will help making MapMan available for KEGG supported organisms, even including non-plant species.

MOBY out MOBY in

The current avalanche of bioinformatics tools leaves the user wondering how to use these. Also it is difficult for a user to navigate through possible tools and format queries according to the database's needs. BioMOBY is a framework to chain different databases and/or webservices together using a machine-readable interface. Thus tools can be linked together, and it is possible for programs to connect to web services or invoke them. Currently MapMan can connect to a Meta Publication Search Engine which uses BioMOBY technology, and more services will be included in the future.

Furthermore, some MapMan classifications will be made available via BioMOBY, so that in return other services can use the MapMan ontology.

Figure 5: LitRep, a BioMOBY Aggregator for scientific publications, currently available using the MapMan link-out feature with AGI mappings. LitRep is the work of Dr. Haase in the group of Heiko Schoof (MPIZ Cologne). It queries TAIR, AtDB, Aramemnon, and the MapMan auto-literature database

Array DB

As quality assessments are often not stored in microarray databases together with the experimental results, several microarray experiments from a multitude of crop plants will be stored within a database together with their Robin based quality assessment, so they can be used as reference.

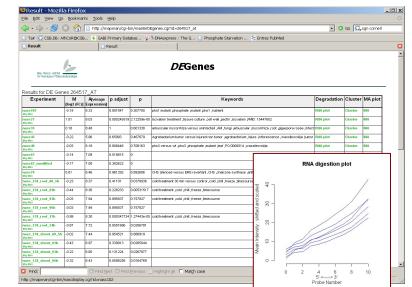


Figure 6: Completed microarray experiments that have been processed with Robin, including Robin's quality assessment

Supporting Database

A large curated data base will be developed that can be used on line in MapMan in combination with filtering and other user interfaces to allow the user to interpret data against a background of a large amount of functional data. This may be especially helpful to crop scientists as it will provide a easy-to-use portal to a large amount of functional information from reference species like Arabidopsis.

■ References

- Thimm et al. 2004 Plant J. v37, 914
- Urbanczyk-Wochniak et al. 2006 Plant Mol Biol 60, 773
- Usadel et al. 2005 Plant Physiol 138, 1195
- Usadel et al. 2006 BMC Bioinformatics 7, 535