

## Dataset Paper

# Extending MapMan Ontology to Tobacco for Visualization of Gene Expression

Maurice H. T. Ling,<sup>1,2</sup> Roel C. Rabara,<sup>3</sup> Prateek Tripathi,<sup>3</sup> Paul J. Rushton,<sup>3</sup> and Xijin Ge<sup>1</sup>

<sup>1</sup> Department of Mathematics and Statistics, South Dakota State University, Brookings, SD 57007, USA

<sup>2</sup> Department of Zoology, The University of Melbourne, Parkville, VIC 3010, Australia

<sup>3</sup> Department of Biology and Microbiology, South Dakota State University, Brookings, SD 57007, USA

Correspondence should be addressed to Xijin Ge; [gexijin@gmail.com](mailto:gexijin@gmail.com)

Received 6 July 2012; Accepted 30 July 2012

Academic Editors: B. G. De Los Reyes, T. W. Nattkemper, and G. Ozsoyoglu

Copyright © 2013 Maurice H. T. Ling et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Microarrays are a large-scale expression profiling method which has been used to study the transcriptome of plants under various environmental conditions. However, manual inspection of microarray data is difficult at the genome level because of the large number of genes (normally at least 30 000) and the many different processes that occur within any given plant. MapMan software, which was initially developed to visualize microarray data for *Arabidopsis*, has been adapted to other plant species by mapping other species onto MapMan ontology. This paper provides a detailed procedure and the relevant computing codes to generate a MapMan ontology mapping file for tobacco (*Nicotiana tabacum* L.) using potato and *Arabidopsis* as intermediates. The mapping file can be used directly with our custom-made NimbleGen oligoarray, which contains gene sequences from both the tobacco gene space sequence and the tobacco gene index 4 (NTGI4) collection of ESTs. The generated dataset will be informative for scientists working on tobacco as their model plant by providing a MapMan ontology mapping file to tobacco, homology between tobacco coding sequences and that of potato and *Arabidopsis*, as well as adapting our procedure and codes for other plant species where the complete genome is not yet available.

## 1. Introduction

Plants, being sessile organisms, must react and acclimatize to abiotic stresses to survive in various environmental conditions. Plants have developed various stress tolerance mechanisms, such as physiological and biochemical alterations, that result in adaptive or morphological changes. In crop production, understanding how cultivated crops respond to abiotic stress is crucial in developing new varieties that could tolerate stress without affecting potential yield. With the rapid development of technologies for functional genomics research, comprehensive analyses at the mRNA, protein, and metabolites level have become possible. This is leading to increased understanding of the complex regulatory networks associated with stress adaptation and tolerance [1].

Currently, microarrays are one of the most popular technologies for large-scale expression profiling because they allow the simultaneous detection of tens of thousands of

transcripts at a reasonable cost [2]. The development of gene chips for model plants like *Arabidopsis* and rice and other species that have a sequenced genome has led to genome-wide transcriptional profiling from diverse tissues. This is a key tool for the identification of novel target genes for functional genomics [3]. Studies using microarrays to characterize abiotic stress responses have been reported for model species such as the moss *Physcomitrella patens* [4], *Arabidopsis thaliana* [5, 6], *Medicago truncatula* [7, 8], and rice [9], as well as nonmodel species such as soybean [10] and *Musa* [11]. However, microarrays generate huge amounts of data which are often in the form of lists of differentially expressed genes. Manual inspection of these data is time consuming, and the volume and variety of information creates a problem in interpretation. This is compounded when transcriptomic analysis is being combined with other OMICS data. Development of new, more reliable methods of data analysis and visualization will enable easier interpretation

of results and thus a greater contribution to explaining the biological problem [12]. Prior to MapMan release in 2004 [13], several bioinformatics tools have been developed to visualize datasets in the context of biological pathways. These include GenMAPP (<http://www.genmapp.org/>) and BioMiner [14] among others. However, their application to plant datasets is limited due to the following reasons. Firstly, these tools were developed for microbial and animal systems and, secondly, flexibility is limited in terms of the display of family members (e.g., class of enzymes) [13]. This limitation had been addressed by MapMan software [13], which relied on its own ontology to classify genes and metabolites and visualize the pathways and processes in pictorial diagrams in a modular system [15]. Currently, there are several stand-alone or web-based visualization tools for biological pathways such as Pathway Tools Omics Viewer (<http://biocyc.org/>) [16], KaPPa-View (<http://kpv.kazusa.or.jp/en/>) [17] Blast2Go (<http://www.blast2go.com/b2gome>) [18], and KEGG Atlas (<http://www.kegg.jp/kegg/atlas/>) [19].

MapMan was initially developed to analyze two sets of 22K Affymetrix arrays that investigated the response of *Arabidopsis* rosettes to low sugar [13]. Rapid advances in sequencing have resulted in full-genome sequences for an increasing number of important crop species (e.g., soybean, rice, maize, papaya, sorghum, and corn). These genome sequences have facilitated the development of large-scale whole-genome arrays. MapMan software can be applied in new species by transferring the MapMan ontology to the transcripts and proteins of the studied species [15]. Several studies have been reported in extending MapMan ontology to other species such as soybean [20], cotton [21, 22], grapevine [23], maize [15], *Musa* [11], potato [24], and tomato [25].

MapMan as a complement to existing visualization tools offers several advantages. Aside from the ease of use, MapMan can display time-based experiments in pictorial format. It can superimpose different datasets as overlay plots which simplifies the identification of shared features on a global and gene-to-gene basis [26].

Tobacco is a popular model plant in recombinant technology because of its well-established gene transfer and regeneration methodologies as well as the availability of many robust expression cassettes for the control of transgene expression [27]. However, the disadvantage of tobacco is that it is an allotetraploid and its genome is not yet fully sequenced. A large gene space sequence project was performed for tobacco (<http://www.tobaccogenome.org/>), and the gene space reads have been deposited as individual unassembled reads at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This is an excellent resource but does not cover the entire genome. The tobacco cultivar Bright Yellow 2 (BY-2) cell line is an important model system to study cell physiology, hormone signaling, cell cycle, cell growth, and stress situations [28]. However, tobacco is still a mostly unsequenced and relatively unannotated plant system in which identification of the proteins and their interactions relies on cross-species identification based on homology and orthology [29].

In this paper, we extend MapMan ontology of sequenced dicot plants to generate a mapping dataset for tobacco. The

dataset we have generated from this study will be a tool for scientists working on tobacco as their model plant by providing a MapMan ontology mapping file for tobacco and homology comparisons between tobacco coding sequences and those of potato and *Arabidopsis*. In addition, we provide a method and the required computer codes to generate MapMan mapping files which may be adapted for other plant species where the complete genome is not yet available. The code and data package can be downloaded from <http://maurice.vodien.com/datasets/MapMan-Tobacco.rar>.

## 2. Methodology

Data files used are the following. Tobacco transcript sequences are from two sources: tobacco gene index 4 (NTGI4) ([ftp://occams.dfci.harvard.edu/pub/bio/tgi/data/Nicotiana\\_tabacum/](ftp://occams.dfci.harvard.edu/pub/bio/tgi/data/Nicotiana_tabacum/)) and tobacco genomic survey sequences (<http://www.pnng.org/tgi/>). (A dataset of 1 159 022 genomic survey sequences was downloaded from the TGI in 2008. These genome survey sequences have subsequently been deposited at the National Center for Biotechnology Information <http://www.ncbi.nlm.nih.gov/>.) Two mapping files from the MapMan Store were used: *Arabidopsis* Information Resource version 9 (TAIR9) to MapMan ontology and genome release version 3.2 from Potato Genome Sequence Consortium to MapMan ontology (downloaded from MapMan Store <http://mapman.gabipd.org/web/guest/mapmanstore>). In addition, potato DNA coding sequence (downloaded from [http://solanaceae.plantbiology.msu.edu/pgsc\\_download.shtml](http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml)) and TAIR9 coding sequence ([ftp://ftp.arabidopsis.org/home/tair/Sequences/blast\\_datasets/TAIR9\\_blastsets/TAIR9\\_cds\\_20090619](ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/TAIR9_blastsets/TAIR9_cds_20090619)) were used to generate nucleotide BLAST database using a local installation of NCBI BLAST version 2.2.25+.

BLAST processing and ontology file assembly were as follows. Both tobacco transcript sequence files were used as input files for nucleotide BLAST (blastn) and translated nucleotide BLAST (tblastx) against the potato DNA coding sequence and TAIR9 coding sequence. The XML files from blastn and tblastx were read using BioPython, and hits with an expectation threshold of less than 1e-9 were saved as comma-delimited files for mapping onto MapMan ontology (Python codes in Listing 1). Duplicate hits in the generated comma-delimited files were removed and all the BLAST results were combined in Dataset Item 5 (Table) for mapping to MapMan ontology using the potato and *Arabidopsis* mapping file from MapMan store. In addition to an expectation threshold of less than 1e-9, the threshold for global sequence identity is set at 30% (Python codes in Dataset Item 8 (Source Code)). The resulting MapMan ontology map file is presented in Dataset Item 9 (Table). Duplicate mapping with the same MapMan bincode (ontology) and identifier (NTGI4 IDs or tobacco genomic survey sequence IDs) was removed. A flowchart of the steps is shown in Figure 1.

For completeness, we had included the necessary source files, Python scripts, and potato and *Arabidopsis* mapping file from MapMan store in our dataset.

Running the tobacco mapping file in MapMan was as follows. To test how well the generated tobacco mapping file

```

from BioBlast import NCBIXML
expect_threshold 1e9
outfile openblast.outputtxt w
for record in NCBIXMLparseopenblast.outputxml
    query_title record.querysplit[0]
    query_length float(record.query_length)
    for alignment in record.alignments
        for hsp in alignment.hsps
            if float(hsp.expect) < expect_threshold
                data query.title
                alignment.title = split[1]
                str(float(hsp.identities)/query_length)
                str(query_length)
                str(hsp.expect)
                outfile.writelines(join(data, '\n'))
outfile.close

```

LISTING 1: Processing BLAST XML to comma-delimited file.

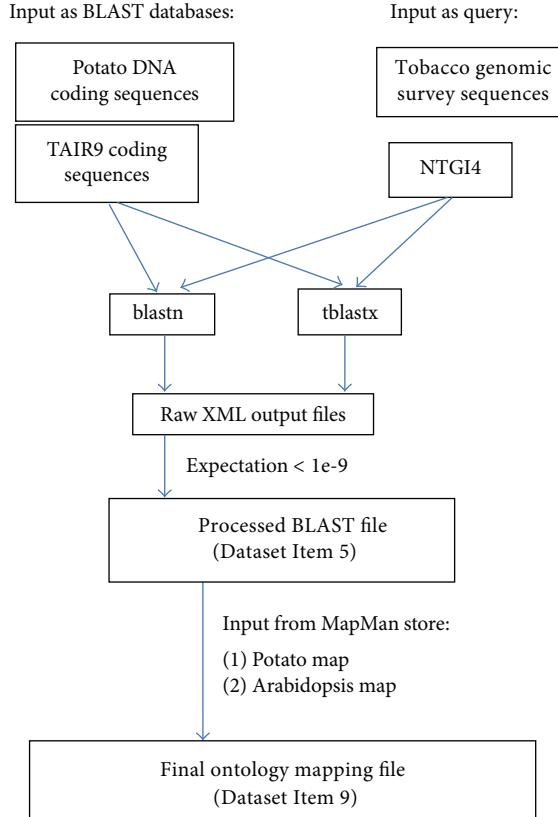


FIGURE 1: Schematic flowchart of analysis.

will visualize microarray data, the mapping file was added into MapMan. The microarray data that was used to test the mapping file is the transcriptional profile of tobacco plants subjected to time-based dehydration. Figure 2 shows the MapMan overview of genes involved in regulation and water stress signaling. The changes in transcript levels in different tissues and at varying times of dehydration stress can easily

be compared in MapMan since the data is presented in a pictorial format. MapMan visualized the changes in gene expression in the families of transcription factors and interestingly revealed the involvement of calcium signaling, receptor kinases, and the plant hormones abscisic acid, jasmonates, and ethylene in tobacco responses to water stress. The visualization provided by MapMan concurred with the results

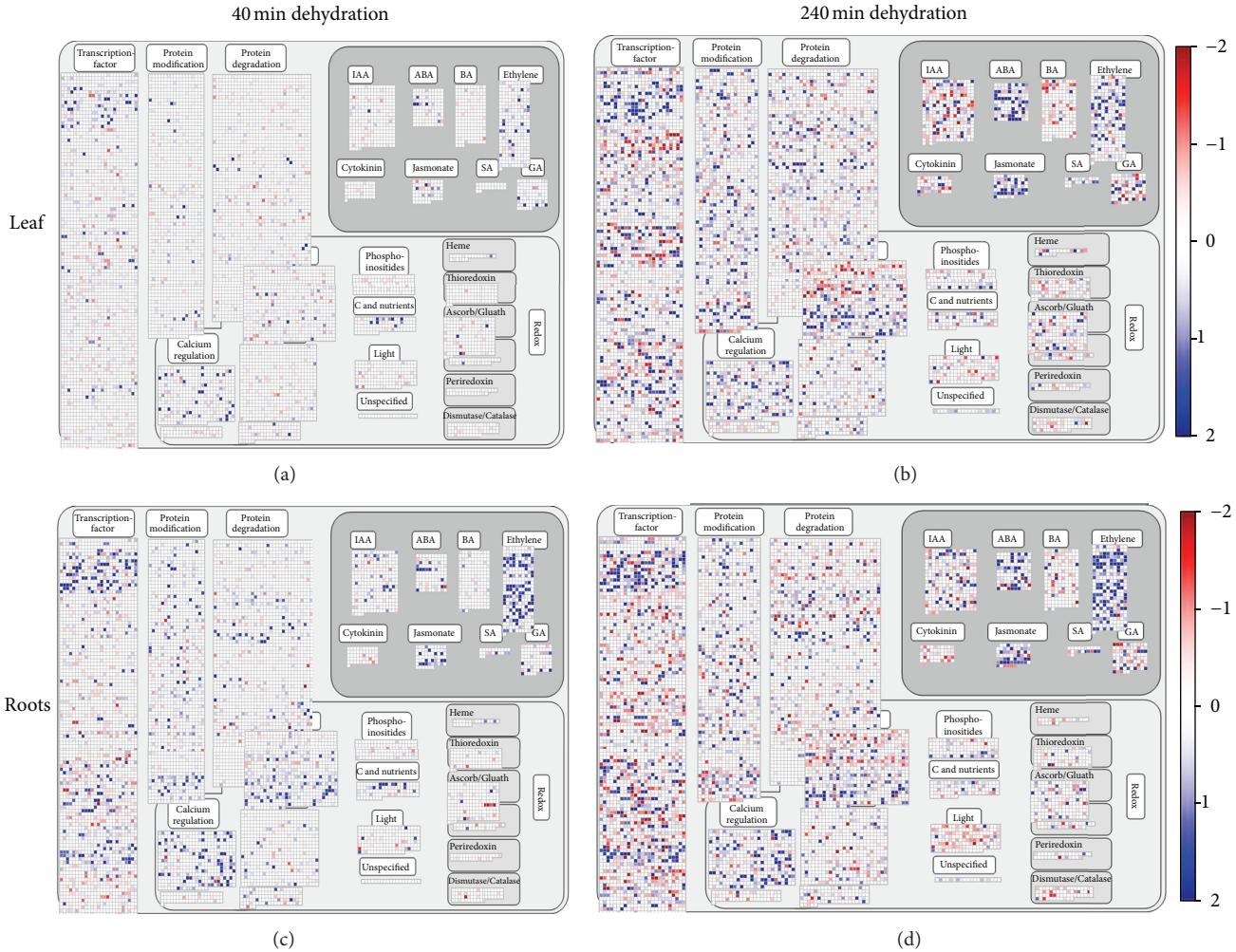


FIGURE 2: MapMan visualization of changes in expression levels of genes associated with regulatory processes using our MapMan ontology mapping file for tobacco (*Nicotiana tabacum* L.). Blue denotes upregulation and red denotes downregulation. Intense blue or red denotes fold changes of 8-fold or more.

from the identification of individual differentially regulated genes (Figure 3). The tobacco MapMan mapping file based on the tobacco gene index 4 (NTGI4) and tobacco genomic survey sequences therefore revealed new insights into water stress responses in tobacco.

### 3. Dataset Description

The dataset associated with this Dataset Paper consists of 9 items which are described as follows.

**Dataset Item 1 (Nucleotide Sequences).** Tobacco gene index 4 (NTGI4) used as an input file for nucleotide BLAST (blastn) and translated nucleotide BLAST (tblastx).

**Dataset Item 2 (Nucleotide Sequences).** Tobacco genomic survey sequences used as an input file for nucleotide BLAST (blastn) and translated nucleotide BLAST (tblastx).

**Dataset Item 3 (Nucleotide Sequences).** Potato DNA coding sequence used to generate nucleotide BLAST database.

**Dataset Item 4 (Nucleotide Sequences).** TAIR9 coding sequence used to generate nucleotide BLAST database.

**Dataset Item 5 (Table).** The processed BLAST result file for regenerating the required MapMan ontology map file. The Query ID column is the identifier for the query sequence from either NTGI transcript or Tobacco GSS, which is also the source of the Identifier column in the ontology map file, while the Mapped ID column is the identifier of the hit sequence from either TAIR9 CDS or Potato CDS. Query Length, Global Identity, E-Value, Query Source, BLAST Database, and BLAST Method are attributes containing the source data for concatenation into the Description column in the ontology map file.

*Column 1:* Query ID

*Column 2:* Mapped ID

*Column 3:* Global Identity

*Column 4:* Query Length

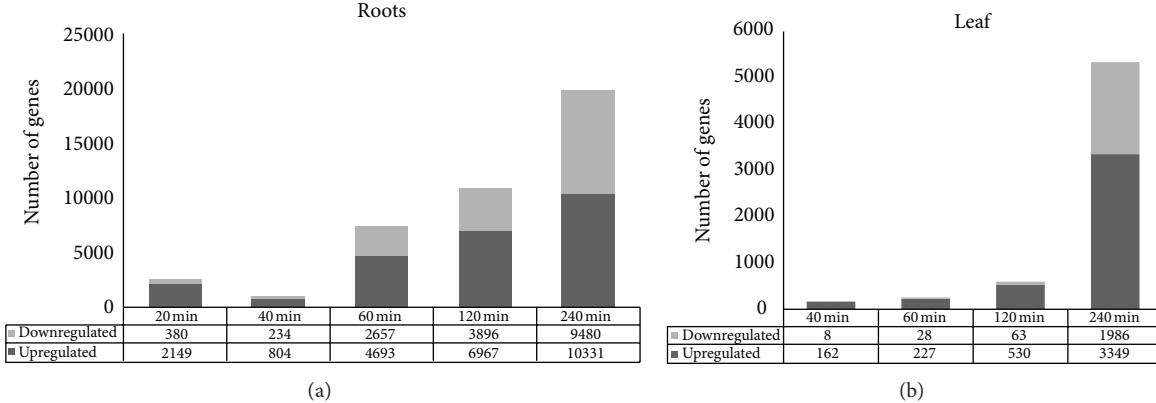


FIGURE 3: Differentially regulated genes in roots and leaves under different dehydration timepoints. Student's *t*-test (with false discovery rate correction using Benjamini-Hochberg method) was conducted and genes with *P* value of  $\leq 0.05$  were selected.

*Column 5:* E-Value

*Column 6:* Query Source

*Column 7:* BLAST Database

*Column 8:* BLAST Method

*Column 9:* Description

*Dataset Item 6 (Table).* Genome release version 3.2 from Potato Genome Sequence Consortium used for mapping to MapMan ontology.

*Column 1:* Bincode

*Column 2:* Name

*Column 3:* Identifier

*Column 4:* Description

*Column 5:* Type

*Dataset Item 7 (Table).* Arabidopsis Information Resource version 9 (TAIR9) used for mapping to MapMan ontology.

*Column 1:* Bincode

*Column 2:* Name

*Column 3:* Identifier

*Column 4:* Description

*Column 5:* Type

*Dataset Item 8 (Source Code).* Python script used for processing and generating MapMan ontology file.

*Dataset Item 9 (Table).* The resulting MapMan ontology map file, containing 5 columns as per MapMan ontology map file format. They are Bincode, Name, Identifier, Description, and Type. Through our own map file generation and use, we found two limitations in the format of the map file. Firstly, only Bincode, Identifier, and Type columns are mandatory and used by MapMan [13]. Bincode and Identifier columns are the

MapMan ontology bin identifier and microarray probe identifier, respectively. In our case, the Identifier column refers to NTGI4 or tobacco genomic survey sequence identifier, which is also used as a probe identifier in our custom microarray. The Type column is default to "T". Secondly, the Name column is the name descriptor used by MapMan [13] for displaying the ontology in a tree format, together with the Bincode, even though it is not mandatory. However, if the Name column is used, each Bincode can only be mapped to one Name. As we had combined two primary map files (*potato* and *Arabidopsis*) to generate a tobacco map file, we found that the Name column may not be consistent with the Bincode and resulted in error. Thus, the Name column is not used and left blank. We used the Description column as a composite of 6 attributes to describe the BLAST process. The six attributes are as follows: Query Length to denote the length of query sequence in a number of bases; Global Identity to denote the global sequence identity between the query sequence and the matched sequence in the BLAST database; E-value to denote the expectation value from the BLAST hit; Query Source to denote the source of the query sequence; hence, the source of the Identifier; which is either "NTGI transcript" or "Tobacco GSS"; BLAST Database to denote the source of sequences to generate the custom BLAST database, which is either "TAIR9 CDS" or "Potato CDS"; and BLAST Method to denote the BLAST program used, which is either "blastn" or "tblastx".

*Column 1:* Bincode

*Column 2:* Name

*Column 3:* Identifier

*Column 4:* Description

*Column 5:* Type

## 4. Concluding Remarks

Figure 2 shows that the MapMan ontology mapping file for tobacco that we have generated does indeed work. It shows the changes in expression level of genes associated with regulatory processes after water stress in both roots and leaves.

The blue color shows genes that are upregulated at the mRNA level and the red color shows genes that are downregulated. The darkest color represents at least 8-fold change in mRNA level. Based on the MapMan results, we can clearly identify areas of primary and secondary metabolisms that are subject to regulation during water stress. These genes are therefore identified as potential targets for improving drought responses.

The successful use of the MapMan ontology mapping file for tobacco (Figure 2) illustrates that our strategy of going via potato has been a good one. This is because each unassembled gene space read from tobacco that is present on the oligo array may only contain a short part of an exon and this may not correspond to any protein sequence in the more distantly related *Arabidopsis* proteome. Potato is much more closely related to tobacco as it is also a member of the Solanaceae. This means that most fragmentary tobacco sequences can be assigned to a corresponding full-length potato sequence that will contain conserved domains that allow identification of the protein. This full-length potato protein sequence will in the majority of cases have a similar type of protein in *Arabidopsis* for mapping purposes. We propose that adapting our procedure and codes for other plant species where the complete genome is not yet available will facilitate MapMan ontology mapping for those plant species.

## Dataset Availability

The dataset associated with this Dataset Paper is dedicated to the public domain using the CC0 waiver and is available at <http://dx.doi.org/10.7167/2013/706465>. In addition, the code and data package can be downloaded from <http://maurice.vodien.com/datasets/MapMan-Tobacco.rar>.

## Conflict of Interests

The authors declare that they have no conflict of interests.

## Acknowledgments

The authors would like to thank the Administrative and Research Computing at the South Dakota State University for providing computational resources. This project was supported by the National Research Initiative Grants 2008-35100-04519 and 2008-35100-05969 from the USDA National Institute of Food and Agriculture. The research in the Rushton laboratory is also supported by the United Soybean Board, The Consortium for Plant Biotechnology Research, the South Dakota Soybean Research and Promotion Council, and the North Central Soybean Research Program. M. H. T. Ling and X. Ge were supported in part by the National Institutes of Health (GM083226 to X. Ge).

## References

- [1] K. Urano, Y. Kurihara, M. Seki, and K. Shinozaki, "Omics' analyses of regulatory networks in plant abiotic stress responses," *Current Opinion in Plant Biology*, vol. 13, no. 2, pp. 132–138, 2010.
- [2] W. Busch and J. U. Lohmann, "Profiling a plant: expression analysis in *Arabidopsis*," *Current Opinion in Plant Biology*, vol. 10, no. 2, pp. 136–141, 2007.
- [3] S. Hafidh, K. Breznenová, P. Ružička, J. Feciková, V. Čapková, and D. Honys, "Comprehensive analysis of tobacco pollen transcriptome unveils common pathways in polar cell expansion and underlying heterochronic shift during spermatogenesis," *BMC Plant Biology*, vol. 12, supplement 24, Article ID 24, 2012.
- [4] A. C. Cuming, S. H. Cho, Y. Kamisugi, H. Graham, and R. S. Quatrano, "Microarray analysis of transcriptional responses to abscisic acid and osmotic, salt, and drought stress in the moss, *Physcomitrella patens*," *New Phytologist*, vol. 176, no. 2, pp. 275–287, 2007.
- [5] M. Seki, M. Narusaka, J. Ishida et al., "Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray," *Plant Journal*, vol. 31, no. 3, pp. 279–292, 2002.
- [6] Y. Oono, M. Seki, M. Satou et al., "Monitoring expression profiles of *Arabidopsis* genes during cold acclimation and deacclimation using DNA microarrays," *Functional and Integrative Genomics*, vol. 6, no. 3, pp. 212–234, 2006.
- [7] D. Li, Z. Su, J. Dong, and T. Wang, "An expression database for roots of the model legume *Medicago truncatula* under salt stress," *BMC Genomics*, vol. 10, article no. 517, 2009.
- [8] D. Li, Y. Zhang, X. Hu et al., "Transcriptional profiling of *Medicago truncatula* under salt stress identified a novel CBF transcription factor MtCBF4 that plays an important role in abiotic stress responses," *BMC Plant Biology*, vol. 11, article no. 109, 2011.
- [9] S. Li, Q. Fu, W. Huang, and D. Yu, "Functional analysis of an *Arabidopsis* transcription factor WRKY<sub>25</sub> in heat stress," *Plant Cell Reports*, vol. 28, no. 4, pp. 683–693, 2009.
- [10] R. Stolf-Moreira, E. G. M. Lemos, L. Carareto-Alves et al., "Transcriptional profiles of roots of different soybean genotypes subjected to drought stress," *Plant Molecular Biology Reporter*, vol. 29, no. 1, pp. 19–34, 2011.
- [11] M. W. Davey, N. S. Graham, B. Vanholme, R. Swennen, S. T. May, and J. Keulemans, "Heterologous oligonucleotide microarrays for transcriptomics in a non-model species: a proof-of-concept study of drought stress in *Musa*," *BMC Genomics*, vol. 10, article no. 1471, p. 436, 2009.
- [12] A. Rotter, B. Usadel, S. Baebler, M. Stitt, and K. Gruden, "Adaptation of the MapMan ontology to biotic stress responses: application in solanaceous species," *Plant Methods*, vol. 3, no. 1, article no. 10, 2007.
- [13] O. Thimm, O. Bläsing, Y. Gibon et al., "MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes," *Plant Journal*, vol. 37, no. 6, pp. 914–939, 2004.
- [14] M. Šírava, T. Schäfer, M. Eiglsperger et al., "BioMiner—modeling, analyzing, and visualizing biochemical pathways and networks," *Bioinformatics*, vol. 18, no. 2, pp. S219–S230, 2002.
- [15] B. Usadel, F. Poree, A. Nagel, M. Lohse, A. Czedik-Eysenberg, and M. Stitt, "A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize," *Plant, Cell and Environment*, vol. 32, no. 9, pp. 1211–1229, 2009.
- [16] S. M. Paley and P. D. Karp, "The pathway tools cellular overview diagram and Omics viewer," *Nucleic Acids Research*, vol. 34, no. 13, pp. 3771–3778, 2006.
- [17] T. Tokimatsu, N. Sakurai, H. Suzuki et al., "KaPPA-view. A web-based analysis tool for integration of transcript and metabolite

- data on plant metabolic pathway maps,” *Plant Physiology*, vol. 138, no. 3, pp. 1289–1300, 2005.
- [18] A. Conesa, S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles, “Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research,” *Bioinformatics*, vol. 21, no. 18, pp. 3674–3676, 2005.
- [19] S. Okuda, T. Yamada, M. Hamajima et al., “KEGG Atlas mapping for global analysis of metabolic pathways,” *Nucleic Acids Research*, vol. 36, pp. W423–W426, 2008.
- [20] Y. Nanjo, K. Maruyama, H. Yasue, K. Yamaguchi-Shinozaki, K. Shinozaki, and S. Komatsu, “Transcriptional responses to flooding stress in roots including hypocotyl of soybean seedlings,” *Plant Molecular Biology*, pp. 129–144, 2011.
- [21] Y. Al-Ghazi, S. Bourot, T. Arioli, E. S. Dennis, and D. J. Llewellyn, “Transcript profiling during fiber development identifies pathways in secondary metabolism and cell wall structure that may contribute to cotton fiber quality,” *Plant and Cell Physiology*, vol. 50, no. 7, pp. 1364–1381, 2009.
- [22] J. A. Christianson, D. J. Llewellyn, E. S. Dennis, and I. W. Wilson, “Global gene expression responses to waterlogging in roots and leaves of cotton (*Gossypium hirsutum* L.),” *Plant and Cell Physiology*, vol. 51, no. 1, pp. 21–37, 2010.
- [23] A. Rotter, C. Camps, M. Lohse et al., “Gene expression profiling in susceptible interaction of grapevine with its fungal pathogen *Eutypa lata*: extending MapMan ontology for grapevine,” *BMC Plant Biology*, vol. 9, article no. 104, 2009.
- [24] M. Kondrák, F. Marincs, B. Kalapos, Z. Juhász, and Z. Bánfalvi, “Transcriptome analysis of potato leaves expressing the trehalose-6-phosphate synthase 1 gene of yeast,” *PLoS One*, vol. 6, no. 8, Article ID e23466, 2011.
- [25] C. Barsan, P. Sanchez-Bel, C. Rombaldi et al., “Characteristics of the tomato chromoplast revealed by proteomic analysis,” *Journal of Experimental Botany*, vol. 61, no. 9, pp. 2413–2431, 2010.
- [26] A. Pitzschke and H. Hirt, “Bioinformatic and systems biology tools to generate testable models of signaling pathways and their targets,” *Plant Physiology*, vol. 152, no. 2, pp. 460–469, 2010.
- [27] R. M. Twyman, “Host plants, systems and expression strategies for molecular farming,” in *Molecular Farming: Plant-Made Pharmaceuticals and Technical Proteins*, R. Fischer and S. Schillberg, Eds., p. 338, John Wiley & Sons, New York, NY, USA, 2006.
- [28] T. Nagata, Y. Nemoto, and S. Hasezawa, “Tobacco BY-2 cell line as the “HeLa” cell in the cell biology of higher plants,” in *International Review of Cytology*, W. J. Kwang and F. Martin, Eds., pp. 1–30, Academic Press, 1992.
- [29] N. Remmerie, T. De Vijlder, D. Valkenborg et al., “Unraveling tobacco BY-2 protein complexes with BN PAGE/LC-MS/MS and clustering methods,” *Journal of Proteomics*, vol. 74, no. 8, pp. 1201–1217, 2011.