GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

# Comparative and joint analyses of gene expression profiles under drought and rewatering in *Arabidopsis*

**Z.H. Xu[1] and W.R. Wu[1,2]**

[1]Department of Agronomy, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China
[2]College of Crop Science, Fujian Agriculture and Forestry University, Fuzhou, China

Corresponding author: W.R. Wu
E-mail: wuwr@zju.edu.cn

Genet. Mol. Res. 12 (3): 3622-3629 (2013)
Received November 14, 2012
Accepted February 1, 2013
Published September 13, 2013
DOI http://dx.doi.org/10.4238/2013.September.13.6

**ABSTRACT.** Drought is a major limiting factor in crop production. Rewatering is a process opposite to drought, allowing plants to recover to their normal physiological state. To understand more thoroughly the set of genes involved in plant response to drought, we comparatively and jointly analyzed the microarray data of drought and rewatering experiments in *Arabidopsis*. A total of 3833 differentially expressed genes (DEGs) were identified. Among them, ~74% were proven to be co-regulated by drought and rewatering. Drought and rewatering showed contrary regulatory effects on almost all of these co-regulated genes. Only ~6% of the DEGs were significantly regulated by drought alone, and the remaining ~20% were significantly regulated by rewatering alone. However, gene ontology analysis suggested that those "rewatering-only" genes also appeared to be related, either directly or indirectly, to drought response.

**Key words:** Drought; Rewatering; Microarray; *Arabidopsis*; Differentially expressed genes

 

## INTRODUCTION

The microarray technique can simultaneously assess the transcription levels of tens of thousands of genes in a single experiment. Therefore, the technique provides a popular and powerful tool for genome-wide analysis of gene expression profiles. One of the most important goals in gene expression experiments is to identify differentially expressed genes (DEGs) under different biological conditions.

Drought is a major limiting factor in crop production. To improve crop tolerance to drought, it is necessary to understand the molecular mechanism of the plant response to drought. For this purpose, several studies were conducted to analyze gene expression profile changes under drought in the model plant *Arabidopsis*, using the microarray technology (Seki et al., 2001, 2002; Kawaguchi et al., 2004; Swindell, 2006). However, the microarrays used in these studies only included part of the genes in *Arabidopsis* and therefore could not reveal the overall expression profile of *Arabidopsis*. More recently, using dual-labeled whole-genome oligonucleotide microarrays, Huang et al. (2008) identified nearly 2000 drought-responsive genes in *Arabidopsis*. Their results suggested that a large number of genes are involved in drought response in plant.

Rewatering can relieve drought stress, allowing plants to recover to their normal physiological state. Hence, rewatering may have contrary regulatory effects to drought on gene expression. Based on the same microarray platform, Huang et al. (2008) also investigated the gene expression profile under rewatering. Although they did not perform statistical analysis of the data, they still noticed by simple examination that most of the genes significantly regulated by drought (identified in the drought experiment) appeared to be inversely regulated by rewatering, suggesting that gene expression under drought and rewatering is negatively correlated. In light of this result, they proposed that rewatering can provide a large-scale validation of the identity of drought-responsive genes. However, to understand the exact relationship between drought and rewatering on gene expression regulation, statistical evidence is required.

Meta-analysis is a classical statistical methodology that has been used in the fields of medicine and sociology for many years (Egger and Smith, 1997; Egger et al., 1997; Smith et al., 2000). It has been used to combine useful information from independent microarray studies aiming at the same or similar scientific questions, so as to improve the statistical power and reliability of DEG detection. Since Rhodes et al. (2002) offered the first case demonstrating the usefulness of meta-analysis in microarray data mining, a number of meta-analysis methods have been proposed and applied to practical studies (Choi et al., 2003; Parmigiani et al., 2004; Stevens and Doerge, 2005; Conlon et al., 2006; Smith et al., 2008). Up to now, however, meta-analysis of microarray data has been limited to the experiments conducted under the same or similar conditions. For the case of two opposite treatments (e.g., drought and rewatering), if the assumption is true that genes responding to one treatment (say, drought) are largely responsive to the other treatment (say, rewatering) in a converse way, the microarray data of the two treatments will have similar patterns as long as the data of one treatment are reversed. Hence, according to its principle, meta-analysis might also be applicable to the microarray data of two opposite treatments.

In this study, we comparatively and jointly analyzed the microarray data of the drought and rewatering experiments in *Arabidopsis* published by Huang et al. (2008), aiming to 1) identify rewatering-regulated genes, 2) identify additional drought-regulated genes, and 3) clarify the relationship between drought and rewatering on gene expression regulation.

 

## MATERIAL AND METHODS

### Data sources

The microarray data used in this study were from two published experiments on *Arabidopsis* (Huang et al., 2008). The dual-dye hybridization system was adopted for the experiments using a microarray containing >26,000 spotted 70-mer oligo-DNA probes (GEO Serial No. GPL1911) fabricated by the University of Arizona (http://ag.arizona.edu/microarray/) in USA or the University of Alberta (http://www.biology.ualberta.ca/facilities/microarray/) in Canada. The first experiment was the drought treatment (*vs* control), which had four microarrays (dataset 1, denoted as D1); the second experiment was the rewatering treatment (*vs* drought), which had two microarrays (dataset 2, D2). These data had been normalized using RobustSplines in Bioconductor (http://www.bioconductor.org/) and were available in the format of log ratios. After preprocessing (e.g., removing spike probes and merging duplicated probes or spots), 24,132 genes were retained for the subsequent analyses. A mixture of D1 and -D2 (denoted as Dm) was used for the meta-analysis. The negative sign before D2 means that all the data in D2 were multiplied by -1 so as to make the gene expression ratio reversed in D2.

### Detection of DEGs

D1 was analyzed with the popular statistical software Significance Analysis of Microarray (SAM). D2 was analyzed with the software Linear Models for Microarray Data (LIMMA) because its sample size was not large enough for SAM. Dm was analyzed with two methods (softwares), SAM and RankProd. SAM is a popular statistical software for analysis of microarrays (Tusher et al., 2001) and has been demonstrated to be applicable to microarray data meta-analysis (Sims et al., 2008), although it was originally developed for independent analysis. LIMMA is a tool for the analysis of gene expression data arising from microarray or RNASeq technologies (Smyth, 2004), implemented in R language (http://www.r-project.org) as a package of the open resource Bioconductor project (Gentleman et al., 2004). It has features that make the analyses stable, even for experiments with a small number of arrays; this is achieved by borrowing information across genes. RankProd is a microarray data meta-analysis tool based on a non-parametric statistical "rank product" (Hong et al., 2006), also provided as a package in Bioconductor. A significance level of false-discovery rate (FDR) = 0.05 was set in all the analyses. In the analysis with SAM, the "one class" model was employed, and missing data were treated with the method of *k*-nearest neighbor imputation algorithm normalization (using the default $k = 10$) implemented in the SAM software. The default setting of the other parameters was adopted. Similarly, in the analyses with LIMMA and RankProd, the "one class" model was also used, and the default setting of the other parameters was adopted.

### Gene Ontology (GO) analysis

GO analysis was performed using the web server agriGO (Du et al., 2010). Significantly enriched GO terms were identified by the chi-square test using a significance level of FDR = 0.01. The default values for the other parameters were used.

## RESULTS

The analyses on D1 and D2 detected 1860 DEGs (named as set G1) and 2459 DEGs (G2), with 933 and 1186 upregulated and 927 and 1273 downregulated, respectively (Figure 1). In the analysis on Dm (meta-analysis), 2306 and 1978 DEGs were detected by SAM and RankProd, respectively. These two methods together identified 2823 DEGs (Gm), with 1462 upregulated and 1361 downregulated, respectively (Figure 1). Taken together, a total of 3833 DEGs were identified in this study (Table S1).

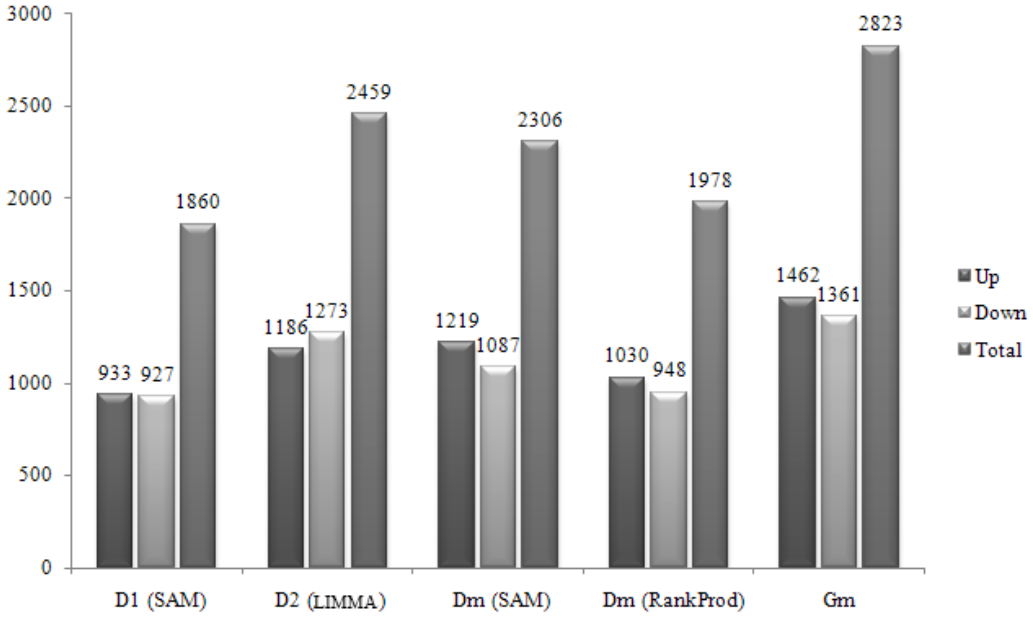


**Figure 1.** Numbers of DEGs detected by individual analysis and meta-analysis.

There were 1109 DEGs in common between G1 and G2, accounting for 59.62% in G1 and 45.10% in G2 (Figure 2). Almost all (1101 DEGs, or 99.28%) of these common DEGs were inversely regulated by drought and rewatering, with 614 up/downregulated and 487 down/upregulated by drought/rewatering, respectively; only 8 (0.72%) of them were regulated in the same directions by drought and rewatering. This result indicated that the regulatory effects of drought and rewatering are generally contrary to each other on the genes that respond to both of them.

As expected, all of the genes in Gm showed clear inverse responses to drought and rewatering (Figure 3). In addition, Gm covered almost all (1098 DEGs, or 99.73%) of the common DEGs between G1 and G2 that were inversely regulated by drought and rewatering, accounting for 38.89% of the genes in Gm. This provided a large-scale validation for the result of the meta-analysis. Based on these results, we can conclude that all of the DEGs detected by the meta-analysis are inversely regulated by drought and rewatering.

It has been mentioned above that only ~60% of the drought-responsive genes (in G1) were regulated by rewatering (also contained in G2). However, comparison indicated that

1632 (87.74%) of the genes in G1 were included in Gm (Figure 2). This suggested that there was a much higher proportion (nearly 90%) of the genes that were regulated by rewatering among the drought-responsive genes. Therefore, meta-analysis can reveal more genes that are responsive to both drought and rewatering.

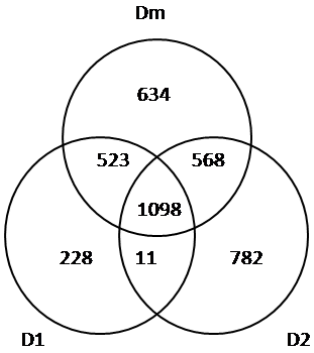


**Figure 2.** Venn diagram analysis between the sets of DEGs detected by the separate analysis (from D1 and D2) and the meta-analysis (from Dm).

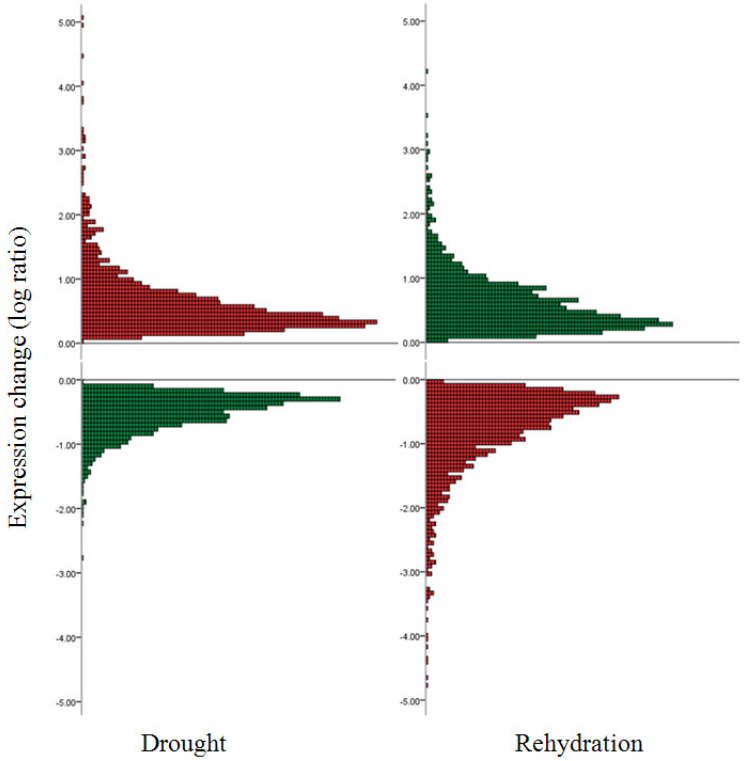


**Figure 3.** Response of DEGs to drought (right panel) and rewatering (left panel) detected by the meta-analysis (from Dm). The color is based on the relative expression in drought versus control (left panel): red = upregulated; green = downregulated.

In summary, according to the comparison among G1, G2, and Gm (Figure 2), the total 3833 DEGs identified in this study can be divided into three subsets; namely, 228 (5.95%) drought-regulated-only genes (subset S1), 782 (20.40%) rewatering-regulated-only genes (S2), and 2823 (73.65%) drought-rewatering co-regulated genes (S3). Obviously, the genes in S1 and S3 are all regulated by drought. A question here is whether the genes in S2 are also related to drought response. To investigate this issue, we conducted GO analysis on S2. A total of 32 GO terms concerning responses to various stimuli or stresses were found to be enriched in S2 (Table 1). These GO terms are all related to drought, either directly or indirectly, as many studies have shown that cross-talk exists among various abiotic and biotic stresses, probably because there are convergence points among different signaling pathways (Fujita et al., 2006). Hence, the GO analysis results suggested that at least most of the genes in S2 are also related to drought response.

**Table 1.** Enriched Gene Ontology (GO) terms in the sets of differentially expressed genes detected only from D2 (S2).

| GO ID | GO term | FDR |
|---|---|---|
| GO:0050896 | Response to stimulus | $8.6 \times 10^{-105}$ |
| GO:0009717 | Response to endogenous stimulus | $1.8 \times 10^{-6}$ |
| GO:0009753 | Response to jasmonic acid stimulus | $4.9 \times 10^{-5}$ |
| GO:0009725 | Response to hormone stimulus | 0.00018 |
| GO:0009733 | Response to auxin stimulus | $4.9 \times 10^{-70}$ |
| GO:0009737 | Response to abscisic acid stimulus | 0.0028 |
| GO:0009739 | Response to gibberellin stimulus | $1.1 \times 10^{-32}$ |
| GO:0009723 | Response to ethylene stimulus | $1.1 \times 10^{-11}$ |
| GO:0009755 | Hormone-mediated signaling pathway | $4.7 \times 10^{-12}$ |
| GO:0009628 | Response to abiotic stimulus | $6.8 \times 10^{-77}$ |
| GO:0009314 | Response to radiation | 0.0044 |
| GO:0009416 | Response to light stimulus | 0.0028 |
| GO:0009639 | Response to red or far red light | $5.1 \times 10^{-6}$ |
| GO:0009266 | Response to temperature stimulus | $2.8 \times 10^{-113}$ |
| GO:0009409 | Response to cold | $2.1 \times 10^{-115}$ |
| GO:0006970 | Response to osmotic stress | $5.6 \times 10^{-44}$ |
| GO:0009651 | Response to salt stress | $2.0 \times 10^{-42}$ |
| GO:0009415 | Response to water | $8.5 \times 10^{-16}$ |
| GO:0009414 | Response to water deprivation | $1.4 \times 10^{-18}$ |
| GO:0009605 | Response to external stimulus | $3.0 \times 10^{-18}$ |
| GO:0009611 | Response to wounding | $5.5 \times 10^{-25}$ |
| GO:0048583 | Regulation of response to stimulus | $1.9 \times 10^{-8}$ |
| GO:0033554 | Cellular response to stress | 0.0042 |
| GO:0070887 | Cellular response to chemical stimulus | $1.2 \times 10^{-22}$ |
| GO:0006950 | Response to stress | $8.7 \times 10^{-81}$ |
| GO:0042221 | Response to chemical stimulus | $8.8 \times 10^{-78}$ |
| GO:0006979 | Response to oxidative stress | $1.3 \times 10^{-11}$ |
| GO:0010035 | Response to inorganic substance | $1.3 \times 10^{-85}$ |
| GO:0010038 | Response to metal ion | $6.6 \times 10^{-64}$ |
| GO:0010033 | Response to organic substance | $1.0 \times 10^{-79}$ |
| GO:0009743 | Response to carbohydrate stimulus | $2.7 \times 10^{-19}$ |
| GO:0010200 | Response to chitin | $6.4 \times 10^{-20}$ |

FDR = false-discovery rate.

## DISCUSSION

In this study, although there were two microarrays in the rewatering experiment, we still detected as many as 2459 DEGs using an appropriate statistical approach, even more than those detected in the drought experiment, and 1350 (~55%) of them were not detected in the drought experiment (Figure 2). This suggests that more DEGs can be detected by setting

the rewatering experiment when studying drought. It can be inferred from the relationship between drought and rewatering that a large proportion of DEGs detected in the rewatering experiment are related to drought response. Hence, rewatering experiments are very useful for the study of molecular mechanism of drought response.

Although drought and rewatering are different processes, it is still possible to perform joint analysis (or meta-analysis) on the gene expression profiles induced by them, owing to their reverse physiological effects. In this study, apart from detecting almost all the genes that were significant in both drought and rewatering, the meta-analysis also detected 523 and 568 genes that were significant only in drought and rewatering, respectively, revealing that these genes were actually regulated by both drought and rewatering; meanwhile, meta-analysis also detected an extra of 634 genes that were not significant in the separate analyses of drought and rewatering (Figure 2). Therefore, the meta-analysis not only could reveal the relationship between drought and rewatering, but it could also detect more genes co-regulated by drought and rewatering. This indicates the merit of meta-analysis.

In this study, we have found from the genes simultaneously significant in the separate analyses of drought and rewatering that drought and rewatering act inversely on almost all the genes co-regulated by them (Figure 3). This is consistent with the fact that drought and rewatering have opposite physiological effects, and is also the basis for the successful meta-analysis of drought and rewatering microarray data. We also tried to perform meta-analysis using the direct mixture of drought and rewatering microarray data, and only seven genes were detected. This also indicates that the regulatory effects of drought and rewatering on gene expression are basically contrary to each other.

Although most of the DEGs detected in this study were co-regulated by drought and rewatering, there were still some genes that were significant only in drought or in rewatering (Figure 2). The results of GO analysis suggest that most of the genes that were significant in only the rewatering treatment are also related either directly or indirectly to drought response. Thus, we can believe that the overwhelming majority of DEGs detected from the rewatering experiment are regulated by drought or related to drought response. Whether and how many of these genes are specifically regulated by rewatering remains to be investigated further.

## ACKNOWLEDGMENTS

Research supported by the National Basic Research Program ("973" Program) of China (Grant #2004CB117306).

## Supplementary material

## REFERENCES

Choi JK, Yu U, Kim S and Yoo OJ (2003). Combining multiple microarray studies and modeling interstudy variation. *Bioinformatics* (19 Suppl) 1: i84-i90.

Conlon EM, Song JJ and Liu JS (2006). Bayesian models for pooling microarray studies with multiple sources of replications. *BMC Bioinformatics* 7: 247.

Du Z, Zhou X, Ling Y, Zhang Z, et al. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38: W64-W70.

Egger M and Smith GD (1997). Meta-analysis: potentials and promise. *BMJ* 315: 1371-1374.

Egger M, Smith GD and Phillips AN (1997). Meta-analysis: principles and procedures. *BMJ* 315: 1533-1537.
Fujita M, Fujita Y, Noutoshi Y, Takahashi F, et al. (2006). Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Curr. Opin. Plant Biol.* 9: 436-442.
Gentleman RC, Carey VJ, Bates DM, Bolstad B, et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5: R80.
Hong F, Breitling R, McEntee CW, Wittner BS, et al. (2006). RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 22: 2825-2827.
Huang D, Wu W, Abrams SR and Cutler AJ (2008). The relationship of drought-related gene expression in *Arabidopsis thaliana* to hormonal and environmental factors. *J. Exp. Bot.* 59: 2991-3007.
Kawaguchi R, Girke T, Bray EA and Bailey-Serres J (2004). Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in *Arabidopsis thaliana*. *Plant J.* 38: 823-839.
Parmigiani G, Garrett-Mayer ES, Anbazhagan R and Gabrielson E (2004). A cross-study comparison of gene expression studies for the molecular classification of lung cancer. *Clin. Cancer Res.* 10: 2922-2927.
Rhodes DR, Barrette TR, Rubin MA, Ghosh D, et al. (2002). Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res.* 62: 4427-4433.
Seki M, Narusaka M, Abe H, Kasuga M, et al. (2001). Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell.* 13: 61-72.
Seki M, Narusaka M, Ishida J, Nanjo T, et al. (2002). Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.* 31: 279-292.
Sims AH, Smethurst GJ, Hey Y, Okoniewski MJ, et al. (2008). The removal of multiplicative, systematic bias allows integration of breast cancer gene expression datasets - improving meta-analysis and prediction of prognosis. *BMC Med. Genomics* 1: 42.
Smith DD, Givens GH and Tweedie RL (2000). Meta-analysis in Medicine and Health Policy. In: Adjusting for Publication and Quality Bias in Bayesian Meta-Analysis. (Stangl DK and Berry DA, eds.). Marcel Dekker, New York, 277-304.
Smith DD, Saetrom P, Snove O Jr, Lundberg C, et al. (2008). Meta-analysis of breast cancer microarray studies in conjunction with conserved cis-elements suggest patterns for coordinate regulation. *BMC Bioinformatics* 9: 63.
Smyth GK (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3: Article3.
Stevens JR and Doerge RW (2005). Combining Affymetrix microarray results. *BMC Bioinformatics* 6: 57.
Swindell WR (2006). The association among gene expression responses to nine abiotic stress treatments in *Arabidopsis thaliana*. *Genetics* 174: 1811-1824.
Tusher VG, Tibshirani R and Chu G (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U. S. A.* 98: 5116-5121.

 