Transcriptome analysis of the aquaporin AtPIP1;2 deficient line in Arabidopsis thaliana

Anastassia Boudichevskaia a,⁎, Marlies Heckwolf a,b,c, Lea Althaus a, Ralf Kaldenhoff a

a Darmstadt University of Technology, Applied Plant Science, Schnittspahnstr. 10, D-64287 Darmstadt, Germany
b Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI 53703, USA
c Department of Agronomy, University of Wisconsin, Madison, WI 53703, USA

⁎ Corresponding author.
E-mail address: boudichevskaia@bio.tu-darmstadt.de (A. Boudichevskaia).

Abstract
Atmospheric CO₂ impacts all aspects of plant development. It has changed in the past and is predicted to change further on. Studies on the response of crop plants to low and elevated CO₂ concerning growth, productivity and physiological processes are intense. In contrast, the molecular mechanisms of cellular CO₂ exchange are still under discussion. At the same time it becomes more and more accepted that carbon dioxide is transported across cellular biomembranes by CO₂ conducting aquaporins. Our recent study (Boudichevskaia et al., 2015) demonstrates that the lack of a single gene product – aquaporin AtPIP1;2 – resulted in massive transcriptional reprogramming in Arabidopsis as a consequence of reduced tissue CO₂ diffusion rates. Therefore, the transcriptome data of the aquaporin AtPIP1;2 deficient line can be used in the comparative expression analyses for better understanding the role of aquaporins with regard to CO₂ and water transport in plants. Here we describe a gene expression dataset generated for three biological replicates per genotype on Affymetrix platform. We provide detailed methods and analysis on microarray data which has been deposited in Gene Expression Omnibus (GEO): GSE62167. Additionally, we provide the R code for data preprocessing and quality control.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Data analysis

After scanning the microarray slides were pre-analyzed using Partek’s Genomic Suite’s workflow designed for Exon-Arrays. All samples were normalized through the Robust Multi-array Average (RMA) algorithm described in [4]. The CEL files together with the corresponding CHP files were deposited on GEO under the accession number GSE62167. The exon signals of these files were summarized to the gene-level. The reliability of data was ensured by the use of three biological replicates per genotype in the experiment. The quality of biological replicates was determined by pairwise correlations [5]. High, positive Pearson’s correlation coefficients were obtained with scores ranging from 0.98 to 0.99 [3]. The correlations were visualized using scatter plots comparisons of the normalized expression values (Fig. 1). The presence of outliers on the gene-wise level was in addition proved before statistical analyses were performed. The data points that differ by more than 50% from the mean expression value of two more uniform values of the same sample were removed from further analysis. Altogether 1190 loci (4.2% from the list of 28,387 genes) were omitted from the analysis. This procedure has been performed in R. The workflow is available in Supplementary material.

In addition to the parametric t-test, the SAM (Significant Analysis of Microarrays) approach embedded in R package was applied to the Affymetrix dataset to confirm gene expression results obtained with the t-test for highly expressed genes (fold-change of two or higher). The SAM was performed using the delta parameter of 0.6 and FDR of 9.3%. The workflow is available in Supplementary material. According to SAM, 152 genes were differentially expressed (wild type vs. atpip1:2−1 line). All genes claimed as significant based on the SAM approach were also significant according to the t-test (Supporting file 2).

Discussion

In this Data in Brief article, we describe a recently obtained dataset of a knockout-line expression profile (AtPIP1:2 deficient). In the related publication [3], these data are compared to profiles that are publicly available and obtained under low CO2 conditions [7,8] or drought stress [9,10]. The analyses demonstrate the relevance of the microarray approach aimed to uncover the physiological function of AtPIP1:2.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2015.04.018.

Conflict of interest

The authors have no conflicts of interest.

Acknowledgments

We are grateful to Dmitri Pescianschi, Progress Inc. Michigan, US, for help in VBA macros development.

Table 1

|                   | Before filtering | After filtering |
|-------------------|------------------|-----------------|
| Total amount of genes | 28,387 | 27,197 |
| Amount of differentially expressed genes |              |                |
| Up-regulated       | 1550       | 1464 |
| Down-regulated     | 83         | 91       |

Fig. 1. XY scatter plots of transcriptome data among three biological replicates. (a) XY scatter plots of the atpip1:2−1 replicates with each other. (b) XY scatter plots of the wild type (N60000) replicates. Identical values are plotted on the blue line. The red lines indicate a log2-fold difference of 1.
References

[1] J.M. Alonso, A.N. Stepanova, T.J. Leisse, C.J. Kim, H. Chen, P. Shinn, et al., Genome wide insertional mutagenesis of Arabidopsis thaliana. Science 301 (2003) 653–657.

[2] M. Heckwolf, D. Pater, D.T. Hanson, R. Kaldenhoff, The Arabidopsis thaliana aquaporin AtPIP1;2 is a physiologically relevant CO2 transport facilitator. Plant J. 67 (2011) 795–804.

[3] A. Boudichevskaia, M. Heckwolf, R. Kaldenhoff, T-DNA insertion in aquaporin gene AtPIP1;2 generates transcription profiles reminiscent of a low CO2 response. Plant Cell Environ. (2015) http://dx.doi.org/10.1111/pce.12547.

[4] R.A. Irizarry, B.M. Bolstad, F. Collin, L.M. Cope, B. Hobbs, T.P. Speed, Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 31 (2003) e15.

[5] M. Lohse, A. Nunes-Nesi, P. Krüger, A. Nagel, J. Hannemann, F.M. Giorgi, et al., Robin: an intuitive wizard application for R-based expression microarray quality assessment and analysis. Plant Physiol. 153 (2010) 642–651.

[6] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57 (1995) 289–300.

[7] O.E. Bläsing, Y. Gibon, M. Günther, M. Höhne, R. Morcuende, D. Osuna, et al., Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis. Plant Cell Online 17 (2005) 3257–3281.

[8] Y. Li, J. Xu, N.U. Haq, H. Zhang, X.G. Zhu, Was low CO2 a driving force of C4 evolution? Arabidopsis responses to long-term low CO2 stress. J. Exp. Bot. 65 (2014) 3657–3667.

[9] D. Huang, W. Wu, S.R. Abrams, A.J. Cutler, The relationship of drought-related gene expression in Arabidopsis thaliana to hormonal and environmental factors. J. Exp. Bot. 59 (2008) 2991–3007.

[10] A. Harb, A. Krishnan, M.M.R. Ambavaram, A. Pereira, Molecular and physiological analysis of drought stress in Arabidopsis reveals early responses leading to acclimation in plant growth. Plant Physiol. 154 (2010) 1254–1271.