Genomics Data 4 (2015) 162–164

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

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

ARTICLE INFO

Article history:
Received 15 April 2015
Accepted 21 April 2015
Available online 30 April 2015

Keywords:
Affymetrix
Microarray
Arabidopsis thaliana

ABSTRACT

Atmospheric CO_2 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 CO_2 concerning growth, productivity and physiological processes are intense. In contrast, the molecular mechanisms of cellular CO_2 exchange are still under discussion. At the same time it becomes more and more accepted that carbon dioxide is transported across cellular biomembranes by CO_2 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 CO_2 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 CO_2 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.

Specifications	
Organism/cell line/tissue	*Arabidopsis thaliana*; leaf material
Sex	
Sequencer or array type	GeneChip®AraGene-1.0 ST arrays (Affymetrix, USA)
Data format	CEL and CHP files
Experimental factors	Expression in *AtPIP1;2* T-DNA insertion line compared to the wild type (N60000)
Experimental features	We performed microarray analysis on control and *AtPIP1;2* T-DNA insertion line to determine differentially expressed genes. The aim was to compare the transcriptomic profile of the *atpip1;2-1* plants with data, available in public domains, related to water stress or low CO_2 content. Biological replicates: 3 control replicates and 3 *atpip1;2-1* replicates
Consent	N/A
Sample source location	Darmstadt, Germany

Direct link to deposited data

The deposited data can be found at: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62167.

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

Experimental design, materials and methods

Arabidopsis thaliana wild type (N60000) and *atpip1;2-1* (N519794) plant lines were obtained from the Nottingham Stock Centre [1] and backcrossed. The *atpip1;2–1* plants were previously characterized by [2]. Plants were grown under conditions described in [3]. The leaf material of three week-old plants was used for microarray analysis. Total RNA was isolated from 5 individual plants for each microarray (15 plants from each line in total) using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, USA) according to the manufacturer's protocol. RNA quality and quantity were determined by using a Nanodrop 2000 (Peqlab). RNA was processed for use on the GeneChip® AraGene-1.0 ST arrays (Affymetrix, USA). In total, 100 ng of total RNA was subjected to Ambion WT Expression Kit (Ambion, USA) and Affymetrix GeneChip WT Terminal Labeling Kit (Affymetrix, USA), following the manufacturer's protocols for whole genome gene expression analysis. The hybridization cocktail was prepared according to the Hybridization, Wash and Stain Kit manual (Affymetrix, USA). The GeneChip®AraGene-1.0 ST arrays (3 per genotype) were hybridized in a rotating oven at 45 °C for 17 h. Standard post-hybridization wash and stain were used (Fluidics Station 450), followed by scanning of the array strips using the Affymetrix GeneChip® Scanner 3000 7G.

http://dx.doi.org/10.1016/j.gdata.2015.04.018

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. The data about remaining genes can be found in Supporting file 1. Statistical analyses including *t*-test and false discovery rate (FDR) correction [6] are described in [3]. All genes with FDR corrected *p*-value < 0.05 were identified as significant (Supporting file 2). Totally, of the 27,197 genes quantified by microarray 3014 genes were scored as significantly regulated. Among these, 91 genes displayed twofold or more decrease in the expression rate and 83 genes revealed twofold or more increase in level of transcript in the *AtPIP1*;*2* T-DNA insertion line (Table 1).

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).

Table 1
Analysis of genes significantly regulated in the mutant line *atpip1*;*2-1* based on the GeneChip® AraGene-1.0 ST arrays (Affymetrix). The information about the whole gene number and amount of genes after removal of outliers is represented. The data are based on three biological replicates. The differentially expressed genes were determined with the parametric *t*-test and further corrected for multiple comparisons (FDR with p-value cutoff of <0.05).

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

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 CO_2 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.

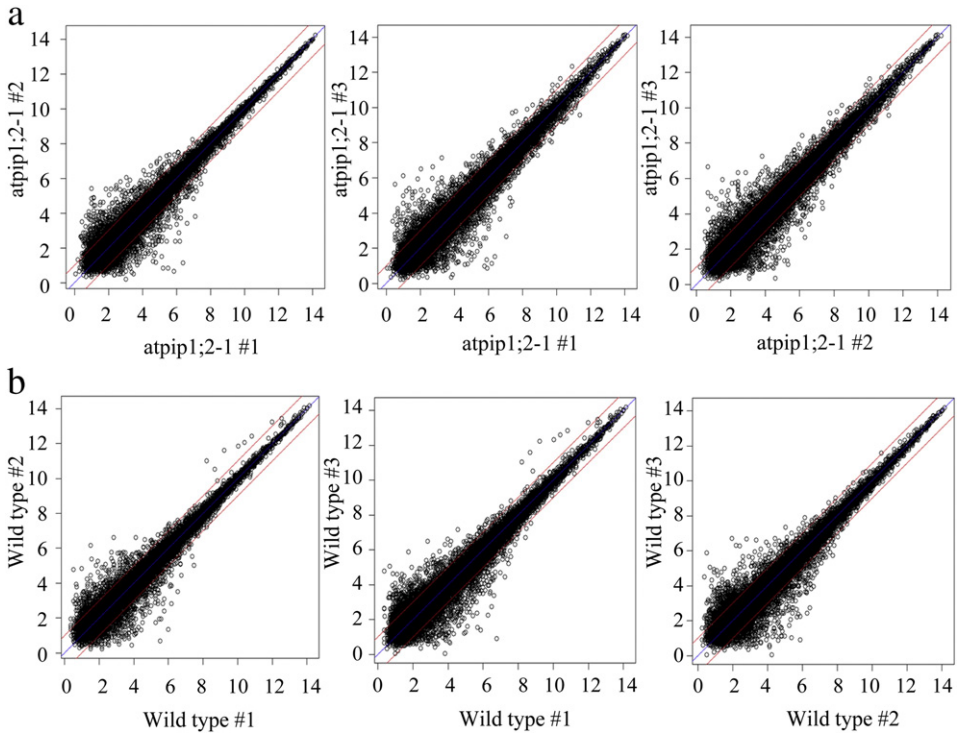

Fig. 1. XY scatter plots of transcriptome data among three biological replicates. (a) XY scatter plots of the *atpip1*;*2-1* replicates with each other. (b) XY scatter plots of the wild type (N60000) replicates. Identical values are plotted on the blue line. The red lines indicate a $\log_2$-fold difference 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 CO_2 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 CO_2 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 CO_2 a driving force of C4 evolution: *Arabidopsis* responses to long-term low CO_2 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.