

Determining the transcriptional regulation pattern of *PgTIP1* in transgenic *Arabidopsis thaliana* by constructing gene coexpression networks*

Haiying Chen¹, Lu Ying¹, Jing Jin¹, Qi Li², Weiming Cai^{1*}

¹Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Graduate School of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China;

²Genminix Informatics Co., Ltd.

Email: wmc@sis.ac.cn

Received 29 October 2010; revised 1 November 2010; accepted 1 November 2010.

ABSTRACT

The seed size, seed mass, and growth rate of transgenic *Arabidopsis* plants containing *PgTIP1*, a ginseng tonoplast aquaporin gene, are significantly higher than those of wild-type *Arabidopsis* plants. Whole genome expression and bioinformatics analysis, including analysis of co-expression networks and transcription factors (Tfscan), were used to determine the key genes that are activated after the expression of *PgTIP1* and the transcription factors that play important roles in the regulation of the genes controlling growth of *Arabidopsis thaliana* seeds by using transgenic *Arabidopsis* plants containing *PgTIP1*. Differential gene analysis showed that transformation of exogenous *PgTIP1* to *Arabidopsis* induced endogenous gene expression changes. Analysis of gene co-expression networks revealed 2 genes, *PIP1* (plasma membrane aquaporin 1 gene) and *RD26* (responsive to desiccation 26 gene; a NAC transcription factor), that were localized in the core of the networks. Analysis of the transcriptional regulation network of transgenic *Arabidopsis* plants containing *PgTIP1* showed that *PIP1* and *RD26* were regulated via DNA binding with a finger domain on transcription factor 2 (*Dof2*). In this study, we demonstrated that *Dof2* induces up-regulation of *PIP1* and *RD26* after transformation with *PgTIP1*. The results of this study provide a new means for conducting research into and controlling growth of *Arabidopsis thaliana* seeds.

*Sponsors: the National High Technology Research and Development Program of China(2006AA10Z112); the National Transgenic Program (2009ZX08004-008B); the Knowledge Innovation Program of the Chinese Academy of Sciences (KJ951-A1-001); the China Manned Space Flight Technology Project and the National Natural Science Foundation of China (30770197, 90917009, 31070237).

Keywords: *Arabidopsis thaliana*; Microarray; Coexpression Network; Transcription Factor Analysis; *PgTIP1*

1. INTRODUCTION

We screened differentially expressed genes using suppression subtractive hybridization (SSH) between hormone-autotrophic and hormone-dependent ginseng callus lines and isolated and characterized an aquaporin gene *PgTIP1* (GenBank accession number DQ237285) that was specifically and highly expressed in hormone-autotrophic ginseng cells [1]. We also demonstrated that, when expressed in *Arabidopsis thaliana*, *PgTIP1* substantially altered vegetative and reproductive growth and development. Compared to wild-type (WT) *Arabidopsis* plants, transgenic (Tg) *Arabidopsis* plants containing *PgTIP1* showed significantly increased seed size, seed mass, and growth rates. Moreover, the fatty acid content of seeds from the Tg *Arabidopsis* plants was 1.85-fold higher than that of seeds from the wild-type control. These results demonstrated that *PgTIP1* is important in the growth and development of plant cells.

In this study, we determined the key genes that were activated after the expression of *PgTIP1*, promoted seed growth, and activated transcription factors, which play important roles in the regulation of the genes related to growth control of *Arabidopsis thaliana* seeds. Whole genome expression and bioinformatics analyses, including coexpression network and transcription factor analyses, were conducted using *Arabidopsis* plants expressing *PgTIP1*.

2. METHOD AND MATERIALS

2.1. Plant Materials and Growth Conditions

Wild type and Tg *PgTIP1 Arabidopsis* seeds (ecotype:

Columbia) were germinated and grown in soil at a photon flux density of $150 \mu\text{mol m}^{-2} \text{s}^{-1}$, 60-80% relative humidity (RH), and a 16/8 h day/night (D/N) cycle at 20-22°C in a phytotron. All experiments were performed with the seedlings in **Figure 1**. Whole plant samples were quickly removed from soil, washed with distilled water, frozen in liquid nitrogen, and stored at -70°C until RNA extraction.

2.2. Generation of *PgTIP1*-Overexpressing *Arabidopsis* Plants

The Tg plants containing *PgTIP1* were generated as described in Lin *et al.* [1]. Briefly, the ORF of *PgTIP1* was cloned into the pHB vector [2] using the HindIII and an XbaI restriction sites to generate a double 35S:*PgTIP1* transgene. Six-week-old *Arabidopsis* (ecotype Columbia) plants were transformed with *Agrobacterium* using the Xoral dip method [3,4]. Seeds were screened in 0.8% selection medium containing 50 $\mu\text{g/ml}$ hygromycin for 7 days and were then transferred to a 1.0% selection medium for an additional 7 days. Resistant plants were transferred to soil and analyzed further. Positive T_0

plants were self-pollinated and T_1 seeds were collected. Individual T_1 plants were tested for expression of *PgTIP1* using RT-PCR. T_1 plants expressing the *PgTIP1* gene were self-pollinated to produce a homozygous generation which was subsequently used for chip analysis.

2.3. RNA Isolation and Microarray Hybridization

Total RNA was extracted using a QIAGEN RNeasy mini kit (Qiagen, CA) according to the manufacturer's instructions and incubated with oligo dT/T7 primers and reverse-transcribed into double-stranded cDNA. *In vitro* transcription of the purified cDNA was performed with T7 RNA polymerase at 42°C for 6 h. The amplified RNA was purified and subjected to a second round of amplification and biotin labeling with Affymetrix's IVT labeling kit. Biotin-labeled RNA was fragmented and hybridized to whole-genome *Arabidopsis* GeneChips (Affymetrix) for 16 h, washed, stained, and scanned.

2.4. Differential Gene Expression Analysis

Two-Factor Analysis of Variance is used to filter differentially expressed genes according to two factors (here, the factors are time and transgene). A random variant model corrected (RVM) t-test [5] and f-test was used to filter significant differentially expressed genes using the time factor, the transgene factor, and the union of these 2 factors. The RVM can raise the degree of freedom to effectively decrease deviation due to small sample size [6,7].

2.5. Construction and Topological Attributes of Coexpression Networks

We built gene coexpression networks to identify gene interactions [8]. Gene coexpression networks were built according to the normalized signal intensity of differentially expressed genes. For each pair of genes, we calculated the Pearson correlation and chose significant correlation pairs with which to construct the network [9]. The purpose of network structure analysis is to locate core regulatory factors (genes). In one network, these factors connect most adjacent genes and have the highest degrees, or connectivity values. For different networks, core regulatory factors were determined by degree differences between 2 class samples [10].

In network analysis, degree centrality is the simplest and most important measure of gene centrality within a network built for determining relative importance. Degree centrality is defined as the link numbers one node has to another. Moreover, to study various properties of networks, k-cores were introduced in the graph theory as a method of simplifying graph topology analysis. A k-core of a network is a sub-network in which all nodes

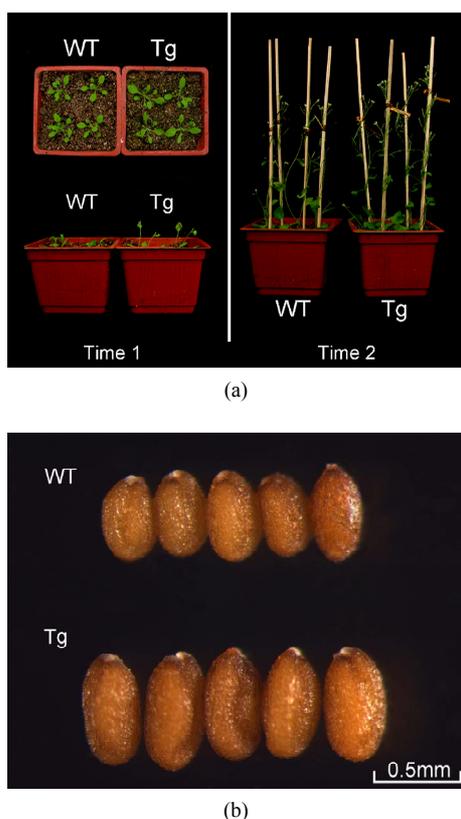


Figure 1. (a) Wild type and transgenic *Arabidopsis thaliana* at time 1 and time 2. Time 1 is the bolting time and time 2 is 10 days after time 1; (b) Mature dried seeds from wild type and transgenic *Arabidopsis* plants. (Bar = 0.5 mm).

are connected to at least k other genes in the sub-network. A k -core of a protein-protein interaction network usually contains cohesive groups of proteins [11,12].

2.6. Transcription Factor Analysis (Tfscan)

Transcription factor analysis (Tfscan) reveals how transcription factors regulate genes. First, the sequences of differentially expressed genes are searched, and then, using the Jemboss software the relationship between genes and transcription factors is determined by counting the correlation between the gene sequence and transcription factor sequence. Next, we built a transcription factor regulation network (TF-Gene-Network) with the interactions between genes and transcription factors. The network's core transcription factor is the most important center and has the largest degree [9,13]. Pearson correlation analysis [9] is used to measure the regulatory ability of transcription factors by calculating the correlation between transcription factors and the genes they regulate and the correlations between the genes regulated by the same factors.

3. RESULTS

3.1. Important Roles of *PIP1* and *RD26* in the Coexpression Network

Results indicated that Tg *Arabidopsis thaliana* grow faster and stronger than the WT and have more plump seeds “**Figure 1**”. The most intriguing phenotype of Tg *Arabidopsis* is the size of mature seeds, which is significantly larger than that of WT seeds “**Figure 1(b)**”.

After two-factor RVM analysis with a threshold of $p < 0.05$, 5796 genes were identified according to time factor (group A), 391 differentially expressed genes were selected with the Tg factor (group B), and 388 differentially expressed genes were chosen taking both Tg and time factors into account (group C). The union of group B and group C was used to build coexpression networks.

One particular coexpression network was constructed using the signal intensity of wild type of *Arabidopsis thaliana* “**Figure 2(a)**” and the other was constructed using the signal intensity of Tg *Arabidopsis thaliana* “**Figure 2(b)**”. The correlation significance level, or interaction between genes in coexpression networks, was calculated by Pearson correlation analysis and found to be greater than 0.99 with a correlation significance of less than 0.0001.

To further understand the effect of the water channel gene *PgTIP1* transfer into *Arabidopsis thaliana*, the Tg coexpression network was simplified to a sub-network containing only core regulatory factors and their interactions “**Figure 3**”. Core regulatory factors have large degrees and k -core in the Tg coexpression network and have the biggest degree distance in the wild type and Tg

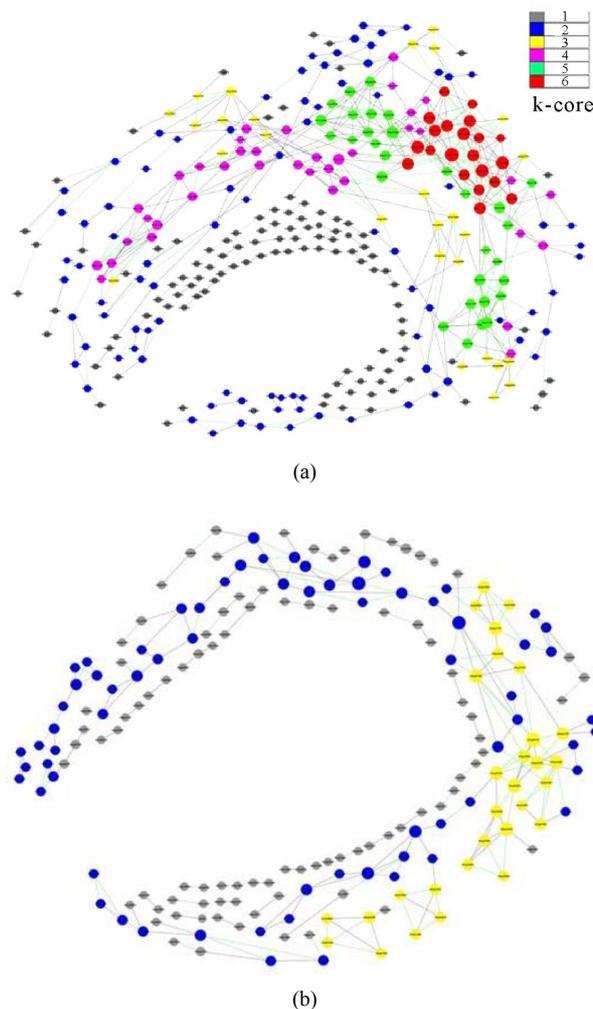


Figure 2. Coexpression networks of wild type (a) and transgenic (b) *Arabidopsis thaliana*. Colors represent the same sub-network with similar k -core values. Comparing network complexity, transgenic species are clearly more complex than wild type.

coexpression networks “**Figure 4**”.

In the concentration area of genes related to the water channel and growth “**Figure 3**”, there are two important genes, *PIP1* and *RD26*. These genes are up-regulated core regulatory factors in Tg and both related with water channel. Additionally, these genes regulate several genes related to growth according to the Tg sub-network “**Figure 3**”.

3.2. Transcription Factor Dof2 and Dof3

Transcription factor analysis (Tfscan) was enlisted for the genes in WT and Tg coexpression networks. In the Tfscan of genes in the Tg sub-network, two important transcription factors, Dof2 and Dof3, were found in the transcription regulation network (TF-Gene-Network) “**Figure 5**”. Both Dof2 and Dof3 were co-expressed

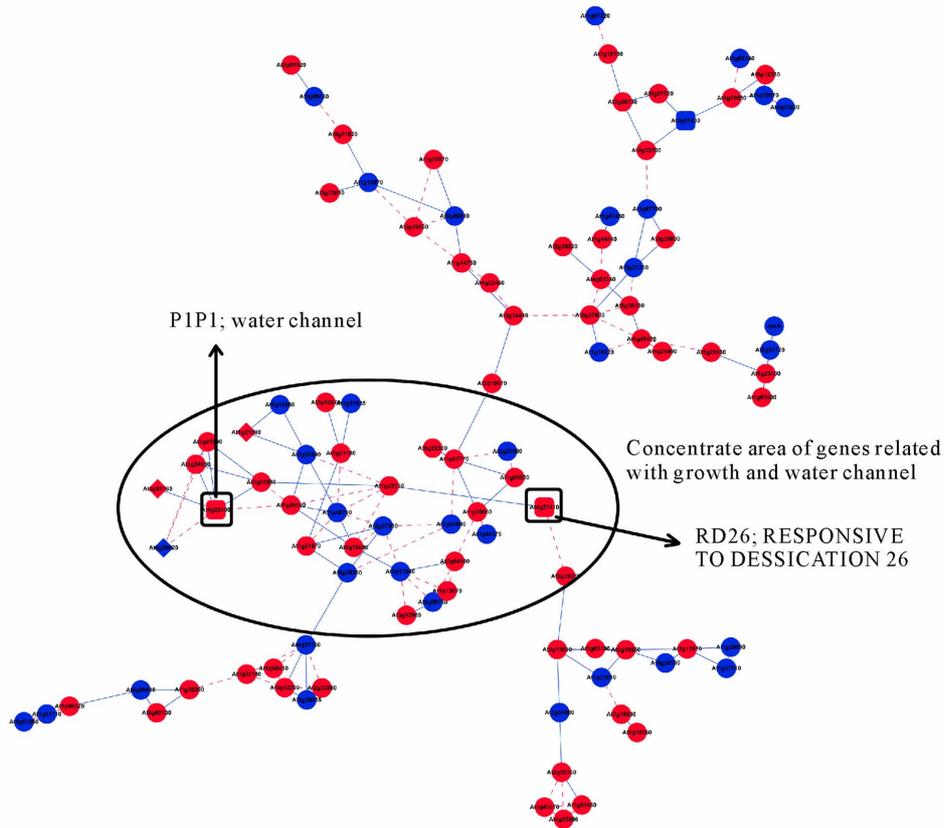


Figure 3. Transgenic (Tg) sub-network. Red node represents an up-regulated gene, blue node represents a down-regulated gene; regular node represents a gene related to water channel, and diamond-shaped node represents a gene related to growth; solid line represents positive correlation between 2 genes, and dashed line represents negative correlation between 2 genes.

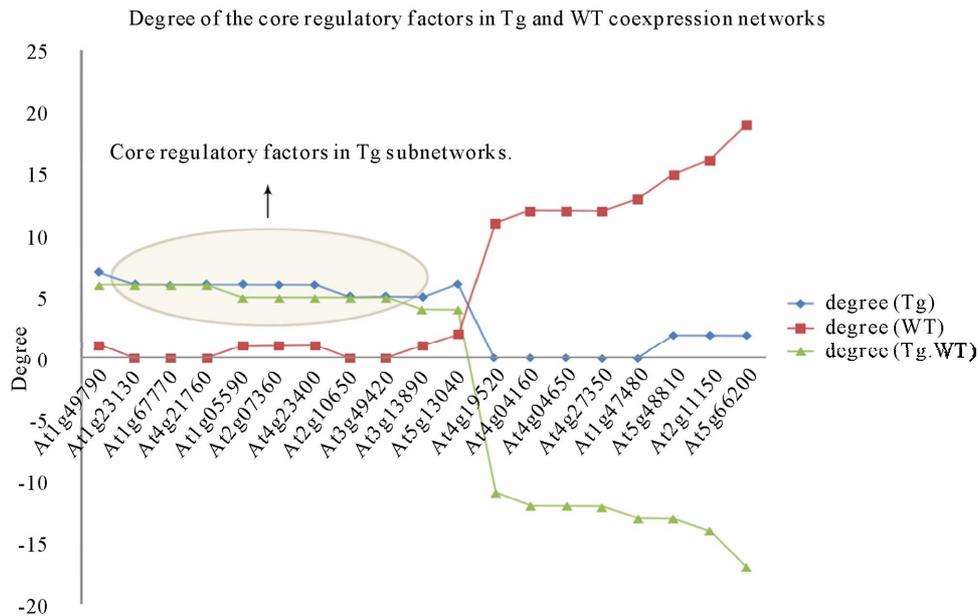


Figure 4. Degree polygon of core regulatory factors in wild type (WT) and transgenic (Tg) coexpression networks. Horizontal axis represents gene name. Vertical axis represents degree value of genes in WT and Tg types. Degree (Tg-WT) = degree (Tg) – degree (WT).

with other genes in the network. Furthermore, Dof2 displayed positive correlation with *PIP1* and *RD26*, the core regulatory factors in the coexpression network. More importantly, expression of Dof2 increased in group A. From correlation analysis of the TF-Gene-Network, the correlation coefficient of the Tg TF-Gene-Network is larger than that of the WT TF-Gene-Network. These results show the regulatory ability of Dof2 and Dof3 to greatly decrease and the genes regulated by these transcription factors are not closely related the Tg in the transcription regulation network of WT “**Figure 6**”.

4. DISCUSSION

The Tg *Arabidopsis* plants containing *PgTIP1* demonstrated promising traits for agricultural application. In this study, we aimed to understand the gene expression changes due to transformation with *PgTIP1*. Whole genome expression analysis using gene chips was employed to identify differences in transcription profiles between Tg and WT plants. A random variance model corrected *t*-test was used to assess the detection value of samples and to effectively reduce the residual caused by small sample sizes by sufficiently increasing the degrees of freedom [5]. We identified 5796 genes categorized

according to the time factor, 391 differentially expressed genes categorized according to the Tg factor, and 388 differentially expressed genes characterized according to both Tg and time factors. Thus, transformation of exogenous *PgTIP1* into *Arabidopsis* induces expression changes of endogenous *Arabidopsis* genes.

Data from gene coexpression network analysis revealed 2 genes, *PIP1* and *RD26*, localized to the network cores. *PIP1* is a plasma membrane aquaporin. *PIP1* members increase the water permeability of cells expressing these aquaporins [14,15]. *RD26* encodes a NAC transcription factor. Seedlings of *RD26*-overexpressed plants have large leaf blades and short petioles, while *RD26*-overexpressed plants have small leaf blades and long petioles [16]. *PIP1* and *RD26* are up-regulated to coordinate *PgTIP1* expression.

Tfscan illuminates 2 important transcription factors, Dof2 and Dof3, regulating gene networks. Dof proteins are DNA-binding proteins with one finger domain transcription factor. Dof-domain proteins play critical roles as transcriptional regulators in plant growth and development [17]. The TF-Gene-Network of Tg reveals that *PIP1* and *RD26* are regulated by Dof2. This suggests that synergism of *PgTIP1* expression and Dof2 enhance

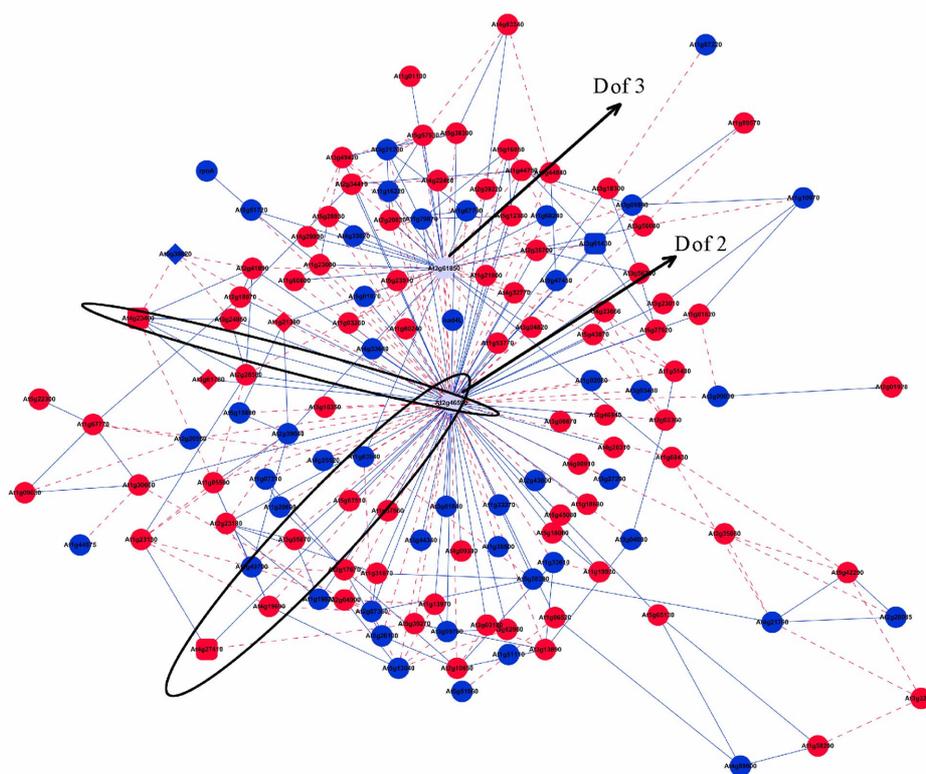


Figure 5. TF-Gene-Network of transgenic (Tg); red nodes represents up-regulated genes and blue nodes represent down-regulated genes in Tg type. The blue boundary circle shows that *PIP1* and *RD26* are regulated by Dof2.

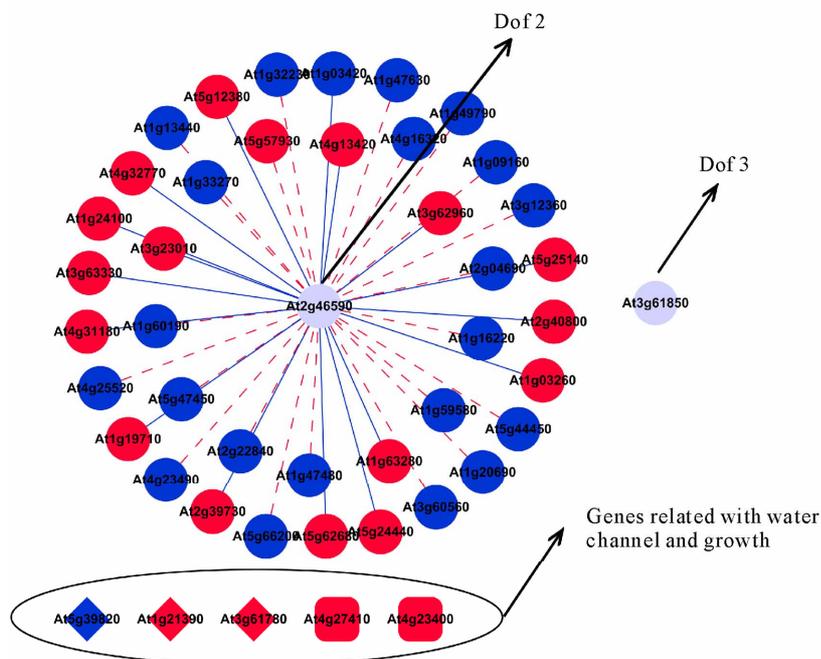


Figure 6. TF-Gene-Network of wild type (WT); red nodes represent up-regulated genes and blue nodes represent down-regulated genes in the WT. The blue boundary circle shows that genes related to water channel and growth are independent from Dof2.

Tg seed growth in *Arabidopsis*.

In conclusion, from coexpression networks, we deduced that the genes *PIP1* and *RD26* are key expression genes activated following transformation of *PgTIP1*. Genes *PIP1* and *RD26* are activated to influence growth of seeds concurrently with *PgTIP1* expression. Transcription factor Dof2 plays an important role in regulation of *PIP1* and *RD26* and their relative genes, which revealed a way to study and control *Arabidopsis thaliana* seed growth. Studying Tg plants with altered expression of key genes using the proposed model is necessary to better understand functions of key genes in Tg *Arabidopsis* plants containing *PgTIP1*.

5. ACKNOWLEDGEMENTS

We thank Dr. Aiping Zang for maintaining the *PgTIP1* transgenic *Arabidopsis* plants. This work was supported by grants from the National High Technology Research and Development Program of China (863 Program) (Grant No. 2006AA10Z112), the National Transgenic Program (Grant No. 2009ZX08004-008B), the Knowledge Innovation Program of the Chinese Academy of Sciences (Grant No. KJCX2-YW-L08), the China Manned Space Flight Technology Project, and the National Natural Science Foundation of China (Grant Nos. 30770197, 90917009, 31070237).

REFERENCES

[1] Lin, W.L., Peng, Y.H., Li, G.W., Arora, R., Tang, Z.C., Su,

W.A. and Cai, W.M. (2007) Isolation and functional characterization of *PgTIP1*, a hormone-autotrophic cell-specific tonoplast aquaporin in ginseng. *Journal of Experimental Botany*, **58**(5), 947-956.

- [2] Mao, J., Zhang, Y.C., Sang, Y., Li, Q.H. and Yang, H.Q. (2005) A role for *Arabidopsis* cryptochromes and COP1 in the regulation of stomatal opening. *Proceedings of the National Academy of Sciences, USA*, **102**(34), 12270-12275.
- [3] Clough, S.J. and Bent, A.F. (1998) Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant Journal*, **16**(6), 725-742.
- [4] Weigel, D. and Glazebrook, J. (2002) *Arabidopsis: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.
- [5] Wright, G.W. and Simon, R.M. (2003) A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics*, **19**(15), 2448-2455.
- [6] Yang, H., Crawford, N., Lukes, L., Finney, R., Lancaster, M. and Hunter, K.W. (2005) Metastasis predictive signature profiles pre-exist in normal tissues. *Clinical and Experimental Metastasis*, **22**(7), 593-603.
- [7] Clarke, R., Ransom, H.W., Wang, A., Xuan, J., Liu, M.C., Gehan, E.A. and Wang, Y. (2008) The properties of high-dimensional data spaces: Implications for exploring gene and protein expression data. *Nature Reviews Cancer*, **8**(1), 37-49.
- [8] Pujana, M.A., Han, J.D., Starita, L.M., Stevens, K.N., Tewari, M., Ahn, J.S., Rennert, G., Moreno, V., Kirchhoff, T., Gold, B., Assmann, V., Elshamy, W.M., Rual, J.F., Levine, D., Rozek, L.S., Gelman, R.S., Gunsalus, K.C., Greenberg, R.A., Sobhian, B., Bertin, N., Venkatesan, K.,

- Ayivi-Guedehoussou, N., Sole, X., Hernandez, P., Lazaro, C., Nathanson, K.L., Weber, B.L., Cusick, M.E., Hill, D.E., Offit, K., Livingston, D.M., Gruber, S.B., Parvin, J.D. and Vidal, M. (2007) Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nature Genetics*, **39(11)**, 1338-1349.
- [9] Prieto, C., Risueno, A., Fontanillo, C. and De las Rivas, J. (2008) Human gene coexpression landscape: Confident network derived from tissue transcriptomic profiles. *PLoS One*, **3(12)**, e3911.
- [10] Carlson, M.R., Zhang, B., Fang, Z., Mischel, P.S., Horvath, S. and Nelson, S.F. (2006) Gene connectivity, function, and sequence conservation: Predictions from modular yeast coexpression networks. *BMC Genomics*, **7**, 40.
- [11] Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, **5(2)**, 101-113.
- [12] Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. and Barabási, A.L. (2002) Hierarchical organization of modularity in metabolic networks. *Science*, **297(5586)**, 1551-1555.
- [13] Vermeirssen, V., Barrasa, M.I., Hidalgo, C.A., Babon, J.A., Sequerra, R., Doucette-Stamm, L., Barabási, A.L. and Walhout, A.J. (2007) Transcription factor modularity in a gene-centered *C. elegans* core neuronal protein-DNA interaction network. *Genome Research*, **17(7)**, 1061-1071.
- [14] Franka, S., Melvin, T. T., Claudio, L., Andrea, S. and Ralf, K. (2002) PIP1 plasma membrane aquaporins in tobacco: From cellular effects to function in plants. *Plant Cell*, **14(4)**, 869-876.
- [15] Olivier, P., Colette, T., Alexandre, G., Yann, B., Raphaël, M., Anton, R.S. and Christophe, M. (2010) A PIP1 Aquaporin contributes to hydrostatic pressure-induced water transport in both the root and rosette of arabidopsis. *Plant Physiology*, **152(3)**, 1418-1430.
- [16] Fujita, M., Fujita, Y., Maruyama, K., Seki, M., Hiratsu, K., Ohme-Takagi, M., Tran, L.S., Yamaguchi-Shinozaki, K. and Shinozaki, K. (2004) A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *Plant Journal*, **39(6)**, 863-876.
- [17] Shuichi, Y. (2004) Dof domain proteins: Plant-specific transcription factors associated with diverse phenomena unique to plants. *Plant and Cell Physiology*, **45(4)**, 386-391.