

Coexpression characteristics of trehalose-6-phosphate phosphatase subfamily genes reveal different functions in a network context

Pinghua Li[a,†], Shisong Ma[a,‡] and Hans J. Bohnert[a,b,c]*

[a]Department of Plant Biology, University of Illinois at Urbana-Champaign, 1201 West Gregory Drive, Urbana, IL 61801, USA
[b]Department of Crop Sciences, University of Illinois at Urbana-Champaign, 1201 West Gregory Drive, Urbana, IL 61801, USA
[c]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1201 West Gregory Drive, Urbana, IL 61801, USA

Correspondence
*Corresponding author,
e-mail: bohnerth@life.uiuc.edu

Received 17 January 2008; revised 25 February 2008.

doi: 10.1111/j.1399-3054.2008.01101.x

Arabidopsis thaliana databases are available that highlight the behavior of the transcriptome under literally hundreds of experimental manipulations, making attempts possible that integrate this information into gene networks. We present and discuss the functioning of a gene network model generated using deposited microarray experiments. Based on a graphical Gaussian model, the network describes conditional coregulation of genes under a variety of external factors and abiotic, biotic and chemical treatments. In this study, we show an aspect of this network that pertains to functions of genes in families where all members appear to carry out the same biochemical reaction. Chosen in this study were 10 genes in the Arabidopsis genome encoding trehalose-6-phosphate phosphatases (TPPs). Nine of these genes were highlighted by the network. Generally, each TPP formed a network associated with genes that identify different functional categories. Thus, network structures were obtained that identified connections to carbon distribution, drought, cold, pathogen responses, calcium and reactive oxygen species/redox signatures, including transcriptional control genes that separated network graphs seeded with different TPP genes. The structure of the transcript coexpression networks, by associating diverse members of gene families into separate clusters, facilitates hypothesis building and in-depth studies of functions of individual genes in families.

Introduction

Genomics approaches have generated a wealth of data in terms of genome sequences, transcript profiles, proteins and their interactions and metabolite composition and dynamics. On its own, each approach has opened new ways to understanding, but we may now also be on the verge of losing deeper insights because the data flow is in the process of becoming overwhelming and unwieldy. If one were able to sort, integrate and analyze the data intelligently, we might facilitate the gain of knowledge that extends comprehension and, as well, provides helpful information for biological disciplines that up to now have had only a marginal association with genome- and gene-oriented concepts (Barabasi and Oltvai 2004).

Abbreviations – GGM, graphical Gaussian model; ROS, reactive oxygen species; T6P, trehalose-6-phosphate; TPP, trehalose-6-phosphate phosphatase; TPS, trehalose-6-phosphate synthase.

[†]Present address: Department of Molecular Biology and Genetics, 317 Biotechnology Building, Cornell University, Ithaca, NY 14853, USA
[‡]Present address: Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520, USA

Genomics can advance research into, e.g. understanding speciation, ecological adaptation and taxonomical questions. Plant breeding is enhanced by the ways we can now reconcile phenotype to genotype and provide precise functions for specific genes (Zeng et al. 2007). Also, engineering established or new crop species could be attempted with a higher probability of success by converting available data into conceptual frameworks that revealed how groups of genes, and the functions they encode, determine and shape developmental and physiological functions of the entire organism. By integrating genomics data, we then also begin to better understand truly intricate processes, such as trophic interactions by which organisms in different kingdoms evolved allelopathic, commensalic or symbiotic interactions, or descriptions of the competition between plant species in natural ecosystems. Presently missing are stringent analyses of already available data that would allow us to define with high confidence the genetic and gene interactions in metabolic and developmental pathways.

Among plant species, the most complete description of development and biochemical functions exists for the crucifer *Arabidopsis thaliana* ('Arabidopsis' hereafter) and its many ecotypes, including a nearly complete set of lines in which single genes have been either eliminated by insertional mutagenesis or their expression enhanced by activation tagging (Arabidopsis Genome Initiative 2000, Østergaard and Yanofsky 2004, Schmid et al. 2005). In addition to a complete and increasingly better annotated genome sequence, thousands of transcript profiling experiments provide information about the expression characteristics of those genes that can be monitored by microarray hybridizations (http://www.arabidopsis.org, http://affymetrix.arabidopsis.info; Craigon et al. 2004). Changes in gene expression in comparison with a control state have been monitored as the plants adjust to or are challenged by abiotic, biotic and chemical treatment conditions. To foster understanding, the data ultimately require integration with phenotypic observations and the effects of mutational disturbance of development and metabolism.

We have used publicly available transcript profiles for Arabidopsis to build a gene network using more than 2000 microarray experiments, with plants subjected to a large variety of treatment conditions, e.g. under different light, exposure to chemicals and hormones, nutrient deficits and abiotic and biotic stresses (Ma et al. 2007). We used clustering methods originally developed for yeast gene networks (Gasch and Eisen 2002, Gong et al. 2005, Ma and Bohnert 2007, Ma et al. 2006) and the statistical methods developed for the graphical Gaussian model (GGM; Ma et al. 2007, Schäfer and Strimmer 2005b). The aspects of the study were to find genes that specifically reacted to a given treatment and distinguish these genes from those whose expression was less specific to a particular treatment or stress but instead resulted from a more universal stress response. By establishing correlations in expression between groups of genes, the underlying assumption is that coregulated genes contribute to functions that require the presence of multiple coordinately regulated gene products. When querying this gene network for known pathways, a large number of connections emerged that have been previously characterized (see, Ma and Bohnert 2008, with a discussion of three subnetworks related to drought stress), while a number of unanticipated new connections appeared. In this study, we exemplify the potential of GGM by analyzing a family of 10 genes encoding trehalose-6-phosphate phosphatases (TPPs; Leyman et al. 2001, Paul 2007). The encoded proteins generate trehalose from trehalose-6-phosphate (T6P), a metabolite that evidently serves as a signal in carbohydrate metabolism and redox-based communication between organelles (Kolbe et al. 2005, Schluepmann et al. 2003, Smeekens 2000). Assigning genes in this family to different functions or pathways by GGM-based networks could help clarifying the role of individual genes.

Methods

Using publicly available microarray data that are based on the Affymetrix ATH1 platform, we have constructed an Arabidopsis gene expression network, based on the GGM (Ma et al. 2007). For a more detailed description, we refer to Ma et al. (2007). Briefly, data from >2400 ATH1 microarray chips that covered a variety of biotic or abiotic stress conditions and other treatments were extracted from NASCArrays (http://affymetrix.arabidopsis.info/narrays/help/usefulfiles.html). The GGM model was based on calculating partial correlation coefficients between gene pairs, estimated via a shrinkage approach as originally proposed by Schäfer and Strimmer (2005a). Calculations, conducted via the software package 'GeneNet', version 1.0.1, implemented in Bioconductor (Gentleman et al. 2004, Opgen-Rhein et al. 2006), were limited to approximately 2000 genes at a time. To include the entire data set for >22 000 genes, we employed an iterative strategy coupled with random sampling. This then expanded the network to cover the transcriptome included on ATH1, in excess of 22 000 genes included on this platform (Ma et al. 2007). Graphs were extracted from the resulting network by specifying seednodes (genes) and the number of connections (edges) by which a network extended from the seednode. Visualization used the neato program, embedded in the software package Rgraphviz, a component of Bioconductor (Carey and Long

2006, Carey et al. 2005, Gansner and North 2000). For detailed descriptions of the network features, see Ma et al. (2007) and Ma and Bohnert (2008).

In this study, we extracted subnetworks for the 10 genes in the Arabidopsis genome that specify the TPP gene family (Schluepmann et al. 2003, 2004). Data concerning TPP genes, transcripts and microarray results were taken from the TAIR, ATH1 and Genevestigator databases (http://www.arabidopsis.org/, https://www.genevestigator.ethz.ch; Zimmermann et al. 2004).

Results

Phylogeny of plant TPP sequences

The Arabidopsis genome includes, according to TAIR, 11 genes encoding bifunctional trehalose-6-phosphate synthases (TPS), with 7 of these including both a synthase and a, possibly inoperative, phosphatase domain (TPS). In addition, 10 genes for TPP and a single trehalase gene constitute a surprisingly complex complement of genes for this metabolism in Arabidopsis (Eastmond et al. 2003, Leyman et al. 2001). The 10 TPPs in Arabidopsis (Table 1) can be compared with 10 genes in the rice (*Oryza sativa* L. ssp. *japonica*) genome, and six genes appear to be present in poplar (*Populus trichocarpa*). They have been included in a phylogenetic analysis (Fig. 1). The sequences are grouped into two major branches, with each branch including three subgroups. Within each major branch, TPPs from Arabidopsis and poplar, e.g. subgroups III and VI, were clearly separated from rice sequences, indicating dicot/monocot divergence. Several pairs of poplar TPPs were closely grouped together with pairs of Arabidopsis TPPs, e.g. At*TPPI*/At*TPPJ* and PtTPP5/PtTPP6, suggesting the existence of orthologs in these species. Closely paired Arabidopsis TPPs, e.g. AtTPPE/AtTPPH, appear to indicate more recent duplications and the appearance of paralogous sequences separating poplar and Arabidopsis.

Developmental and tissue-specific expression of Arabidopsis TPP genes

Two genes, *TPPA* and *TPPB*, in Arabidopsis have been functionally characterized as TPPs (Vogel et al. 1998). Table 2 lists data on functional and transgenic studies with TPPs from Arabidopsis, yeast, rice, maize and *Escherichia coli*. Several publications indicate involvement of the trehalose biosynthesis and degradation pathway in sugar/carbohydrate sensing, and to some degree, they identify T6P as a signaling molecule that may transmit redox-generated signals (Garg et al. 2002, Kolbe et al. 2005, Miranda et al. 2007, Schluepmann et al. 2003, 2004). The Genevestigator Meta-Analyzer program (https://www.genevestigator.ethz.ch/at/; Zimmermann et al. 2004) has been designed to represent gene expression profiles of genes in the context of environmental stresses, organs and growth stages. Signal intensities for each gene, normalized to the highest intensity of this gene, were used to visualize Arabidopsis TPP expression during different developmental stages (Fig. 2A) and in various tissues (Fig. 2B). As indicated in Fig. 2A, AT1G35910 (*TPPD*), AT1G78090 (*TPPB*), AT4G39770 (*TPPH*), AT5G10100 (*TPPI*) and AT5G51460 (*TPPA*) showed highest expression during the development of rosette leaves, whereas AT4G12430 (*TPPF*), AT4G22590 (*TPPG*) and AT5G65140 (*TPPJ*) were preferentially expressed during senescence and in siliques and developing seeds.

Fig. 2B shows the tissue-specific expression of the Arabidopsis TPP family. AT1g35910 (*TPPD*) alone is specifically expressed in suspension cells, while AT1g78090 (*TPPB*), AT5g51460 (*TPPA*) and AT5g65140 (*TPPJ*) are highly expressed in different region of roots, and AT4g12430

Table 1. TPP genes in *Arabidopsis thaliana*. With 10 TPP genes in the Arabidopsis genome, 9 are represented in the GGM network, only AtTPPA and AtTPPB have been studied. AGI, Arabidopsis Genome Initiative.

AGI	Annotation	Short name
AT5G51460	Trehalose-6-phosphate phosphatase	AtTPPA
AT1G78090	Trehalose-6-phosphate phosphatase	AtTPPB
AT1G22210	Trehalose-6-phosphate phosphatase, putative	AtTPPC (not in network)
AT1G35910	Trehalose-6-phosphate phosphatase, putative	AtTPPD
AT2G22190	Similar to trehalose-6-phosphate phosphatase	AtTPPE
AT4G12430	Trehalose-6-phosphate phosphatase, putative	AtTPPF
AT4G22590	Trehalose-6-phosphate phosphatase, putative	AtTPPG
AT4G39770	Trehalose-6-phosphate phosphatase, putative	AtTPPH
AT5G10100	Trehalose-6-phosphate phosphatase, putative	AtTPPI
AT5G65140	Trehalose-6-phosphate phosphatase, putative	AtTPPJ

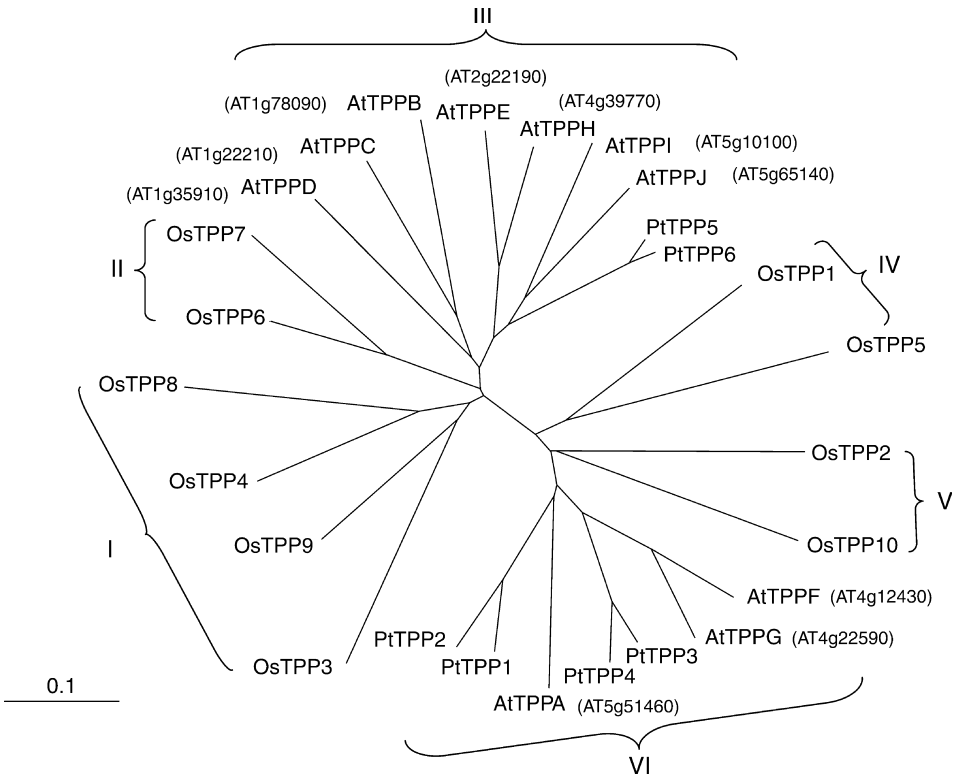

Fig. 1. Phylogeny of functionally identified and putative TPPs from Arabidopsis, rice and poplar. The tree was generated with PHYML maximum likelihood software. The accession numbers of TPPs are listed. Arabidopsis: AtTPPA (AT5g51460), AtTPPB (AT1g78090), AtTPPC (AT1g22210), AtTPPD (AT1g35910), AtTPPE (AT2g22190), AtTPPF (AT4g12430), AtTPPG (AT4g22590), AtTPPH (AT4g39770), AtTPPI (AT5g10100) and AtTPPJ (AT5g65140). Rice: OsTPP1 (Os02g0661100), OsTPP2 (Os10g0553300), OsTPP3 (Os07g0624600), OsTPP4 (Os02g0753000), OsTPP5 (OSJNBa0010H02.18), OsTPP6 (Os08g0409100), OsTPP7 (Os09g0369400), OsTPP8 (Os06g0222100), OsTPP9 (Os03g0386500) and OsTPP10 (Os07g0485000). Poplar (DOE-JGI database http://genome.jgi-psf.org/pub/JGI_data): PtTPP1 (estExt_Genewise1_v1.C_LG_XV2838), PtTPP2 (eugene3.00121254), PtTPP3 (estExt_fgenesh4_pm.C_1270030), PtTPP4 (gw1.III.870.1), PtTPP5 (grail3.0011016001) and PtTPP6 (grail3.0026014501).

(*TPPF*), AT4g22590 (*TPPG*) and AT5g51460 (*TPPA*) are highly expressed particularly in pollen. The figures also indicate that AT1g22210 (*TPPC*) is hardly expressed in any tissue and under all developmental conditions studied and does not perceptibly respond to any experimental manipulation of the plants. As well, *TPPC* is not represented in the GGM network.

Expression pattern of Arabidopsis TPPs under different stresses and hormone treatments

Expression of Arabidopsis TPPs in shoots and roots under different stresses and hormone treatments, calculated from AtGenExpress data, is displayed in the form of heatmaps (Figs 3A–C). Omitted was the extremely lowly expressed *TPPC* gene that showed no discernable variation in expression under any condition. Differential expression changes of the other TPPs could be roughly categorized into four groups. Genes in group 1, including AT1g35910 (*TPPD*), AT5g10100 (*TPPI*) and AT4g12430 (*TPPF*), were significantly induced by osmotic/salt stresses, with two of them (AT1g35910 and AT5g10100) only induced in the roots. AT4g12430 was also highly induced by cold stress in roots. The second group included genes At1g78090 (*TPPB*) and At4g39770 (*TPPH*) with significant downregulation under osmotic/salt stress. However, different from AT4g39770, which responded to the stresses in shoots and roots, AT1g78090 showed a root-specific response. Group 3 genes, AT2g22190 (*TPPE*), AT4g22590 (*TPPG*) and AT5g65140 (*TPPJ*), were induced by cold and salt in shoots and roots but induced by osmotic stress in shoots alone. The sole gene in group 4, AT5g51460 (*TPPA*), was repressed by osmotic stress in shoots and roots but induced by salt stress in the roots. Only AT4g22590 showed a response to wounding.

Several Arabidopsis TPPs also responded to hormone treatment (Fig. 3C). In seedlings, AT2G22190, AT4G12430 and AT5G10100 were induced by ABA, and AT4G22590 was induced by ABA and jasmonate.

Table 2. Functional identification of TPPs in plants. AGPase, ADP-glucose pyrophosphorylase.

Sources of TPPs	Genes	Function in plants	Reference
Arabidopsis	*AtTPPA*, *AtTPPB*	Functional complement the yeast tps2 mutant, expressed in flowers and young developing tissue of Arabidopsis	Vogel et al. (1998)
	AtTPPA, AtTPPB and *AtTPPJ*	Induced by hypoxia	Liu et al. (2005)
	AtTPPA, AtTPPB, AtTPPI, *AtTPPJ*	Induced by nitrate treatment in roots	Wang et al. (2003)
	AtTPPB and *AtTPPF*	Expression induced by 100 m*M* trehalose treatment	Schluepmann et al. (2004)
	AtTPPJ	Induced by Cadmium stress in roots	Roth et al. (2006)
Rice	*OsTPP1 and OsTPP1 2*	Induced by severe chilling stress as well as ABA, salt and drought stresses, involved in transient induction of trehalose biosynthesis during stress	Pramanik and Imai (2005), Shima et al. (2007)
Maize	*RAMOSA3*	A TPP that controls inflorescence architecture by modification of a sugar signal	Satoh-Nagasawa et al. (2006)
E. coli	*OtsB*	Attenuate sucrose- and trehalose-dependent redox activation of AGPase when overexpressed in Arabidopsis	Kolbe et al. (2005)
	OtsA–OtsB	Rice that overexpressed TPSP fusion gene exhibited sustained plant growth, less photooxidative damage and more favorable mineral balance under salt, drought and low-temperature stress conditions	Garg et al. (2002)
Yeast	*ScTPS1–TPS2*	Overexpression of the yeast TPS–TPP chimeric translational fusion protein increased drought, freezing, salt and heat tolerance in Arabidopsis	Miranda et al. (2007)
Tobacco	*ScTPS1–TPS2*	Improved drought tolerance/trehalose accumulation by overexpression	Karim et al. (2007)
Arabidopsis	*AtTPS6*	Mutation in *TPS* gene affecting cell shape and plant architecture	Chary et al. (2008)

GGM gene network

The network, with >2000 microarray experiments incorporated, provides options about the statistical confidence that can be applied to the data (Ma and Bohnert 2008, Ma et al. 2007). In the version used in this study, we represent data with a value ($|\text{pcor}| > 0.05$) that includes the largest number of genes among the three models that have been generated (Ma and Bohnert 2008). We accepted this model as biologically significant because it includes for many biochemical pathways the largest number of pathway genes.

TPPs in GGM

Trehalose metabolism (Ramon and Rolland 2007) provides an example for the potential value of the gene network model. The majority of genes, 18 of 22 genes, in the synthesis and metabolism of trehalose are included in GGM network graphs (not shown). Without a network representation among the TPPs is only *TPPC*, which is expressed at very low levels. The nine genes in the network (Figs 4A–I) identify separate networks that do not, or do only very distantly, overlap, in those cases separated by an extended number of edges. This separation of TPPs into distinct subnetworks, as recorded by GGM, appears to highlight how the functionally presumably identical enzymes react to different stimuli that affect metabolism and, in particular, stress responses.

The *TPPA*-centered network is associated with several genes whose functions are ill defined or unknown, including a RNA-binding protein that has not been characterized functionally. The *TET10* gene included has been associated with pattering in Arabidopsis early leaf development, while the serine carboxypeptidase 2 that is included has been associated with brassinosteroid signaling. In contrast, the network modeled by the *TPPB* seed gene identifies a larger number of genes in carbon and hormone (cytokinin) metabolism, redox-based processes and in signaling, such as two clavata-like genes and *CIPK13* and *SnRK2.9* genes. The *CLE2/6* genes are involved in apoplastic receptor-binding functions. The subnetwork surrounding *TPPD* shows as an interesting feature coexpression with a number of transcription factors in the *WRKY, ERF, MYB*, heat shock and homeobox leucine zipper categories. Further included are signaling functions. The *TPPE* network was enriched in transcription factors involved in embryo development, e.g. MEE3 and BEL1, indicating a possible function for *TPPE* in seed maturation and/or embryogenesis. *TPPE* is most highly expressed in late silique and seed maturation stages 8–10 (see Genevestigator). *TPPE* also included genes encoding glycosyl hydrolases, an invertase and

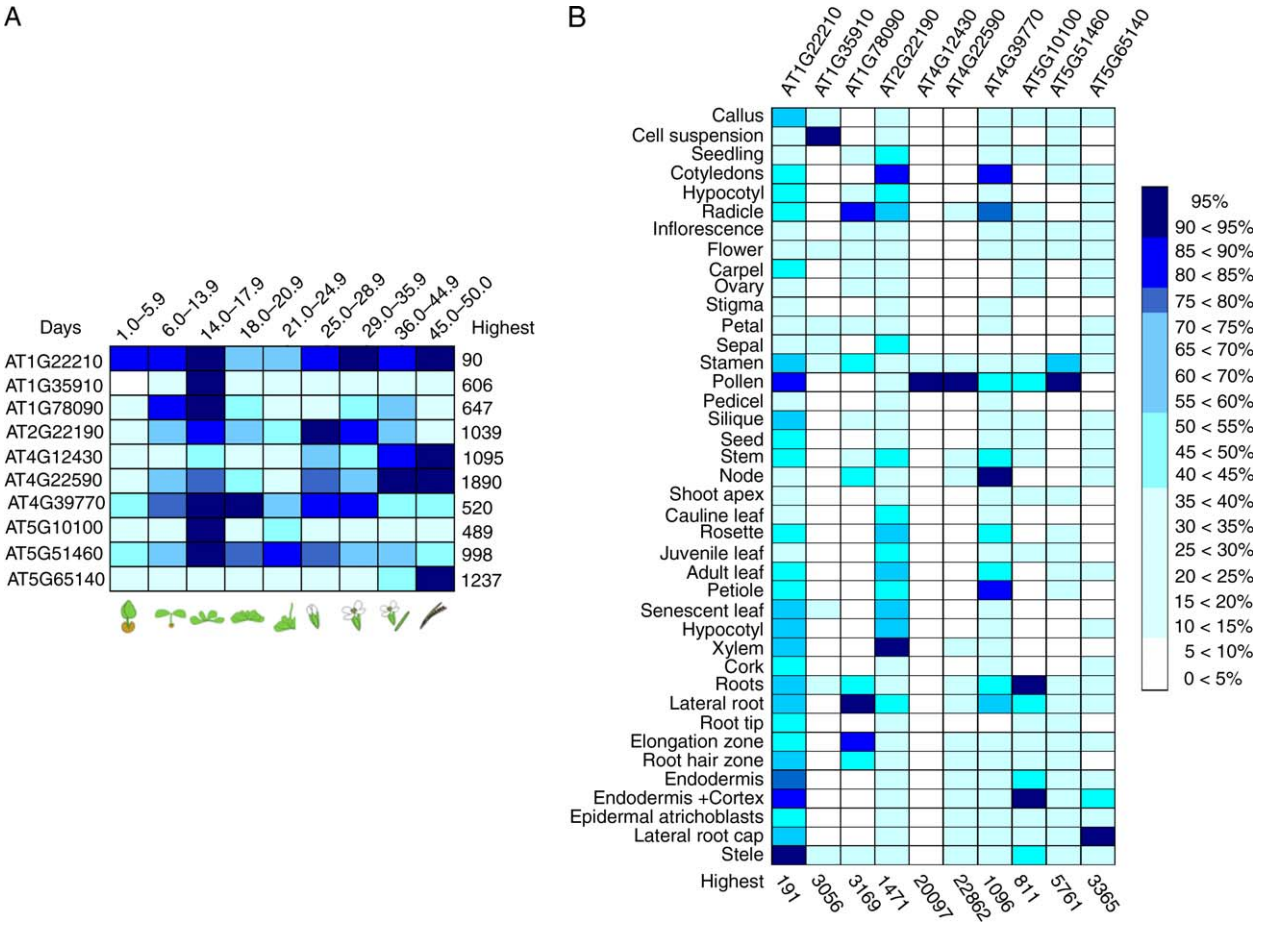

Fig. 2. Expression levels of TPPs from Arabidopsis. (A) Expression levels during the life cycle. (B) Organ-specific expression. The figures show transcript expression deduced from Arabidopsis gene expression microarray experiments. All gene-level profiles were normalized such that for each gene the highest signal intensity reported was given the value 100% (dark blue). Absence of signals received the value 0% (white) as displayed by Genevestigator (https://www.genevestigator.ethz.ch). The intensity listed in the figure represents the highest intensity listed in Genevestigator.

a His-phosphotransfer protein. The *TPPF* is only remotely connected to a network, through a sucrose phosphate synthase gene. Extending the network further, several transcription factors appear, as well as signaling components and metabolic functions. The *TPPG* network includes four genes for transcription factors/activators, among which *DREB2A* and *DREB2B* have clearly been associated with drought responses (Agarwal et al. 2006). Further included are a protein phosphatase 2C gene and a lipase-3 gene, a multi-copper oxidase and an RNA helicase, all of which may be involved in signaling functions. The remaining genes, an ATPase, *PARG*, a poly(ADP-ribose) glycohydrolase and a functionally unknown gene with a WD-40 domain (indicative of protein–protein and/or protein–DNA interactions), complete this subnetwork dominated by genes with a documented or strongly suggested signaling and control function. *TPPI* was also associated with several transcription factor genes, signaling functions and many functionally unknown genes. *TPPJ* expression has been shown to increase in stages 7–10 of silique and seed development, with a peak in stage 9 (see Genevestigator). It is associated with DOG1, a quantitative trait locus involved in the control of seed dormancy, dehydrin and transcription factor expression related to ABA and signaling.

As an example of a larger graph, the subnetwork, extending two edges from *TPPH* is included. Next to functionally unknown genes and genes encoding transcription factors, the largest group of genes identifies genes involved in reactive oxygen species (ROS) generation and detoxification.

T6P is generated by TPS, most of which include both a synthase and a phosphatase domain. The Arabidopsis genome includes 11 of these TPS genes, of which 8 can be modeled by GGM (not included). Interestingly, TPS are not closely associated with TPPs, the only exception

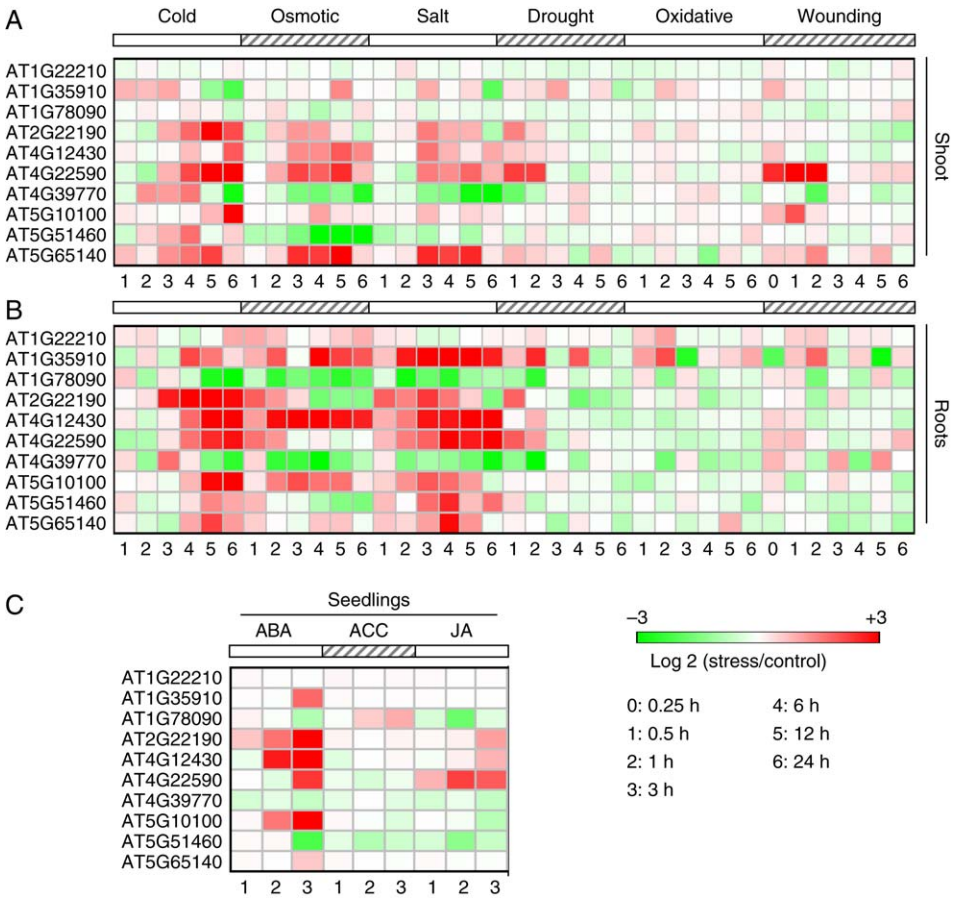

Fig. 3. Expression levels of Arabidopsis TPP genes under different stresses and hormone treatments. (A) Shoots, (B) roots and (C) hormones. Log2 expression ratio (stress/control) of Arabidopsis TPPs under different stresses and hormone treatments was calculated from AtGenExpress data.

being *TPS8*, which is separated by two edges from *TPPH* (AT4G39770) (Fig. 4I). However remotely (two edges), *TPS8* is also connected to three other TPS (*TPS9, TPS10* and *TPS11*). The network in which these four TPS transcripts are located is further characterized by transcripts for a number of calcium-binding proteins, several transcription factors, sugar and ion transporters and (uncharacterized putative) signaling intermediates. Three of the remaining TPS genes are difficult to place: *TPS6* is associated with a network that includes genes for proteins involved in chromatin structure, cyclins and RNA-binding proteins. The identity of these genes seems to correspond well with a function for *TPS6* in the determination of cell shape (Chary et al. 2008). *TPS1* may be associated with ovule development based on the genes that populate its network. Indeed, a number of *TPS1* alleles show an embryo-lethal phenotype, disturbances in flowering transition and less meristem activity (van Dijken et al. 2004). Last, *TPS7* is based on a network that includes relationships to mitochondrial functions. *TPS5* is strongly associated with glutredoxins, a pyrophosphatase, tyrosine phosphatase-interacting protein, isoamylase, glucose-1-P adenylyltransferase and *ATMPK19* (in addition to several functionally unknown genes).

Discussion

Trehalose and T6P

Genes for trehalose synthesis and metabolism are ubiquitous, while trehalose accumulates only in a few organisms, typically associated with osmotic stress (Avonce et al. 2006, Paul 2007, Smeekens 2000). In plants, while only trace amounts accumulate, the pathway genes are induced by several conditions that also constitute stresses: cold, drought, salinity, among other treatments. The activity of TPS has been associated with enhanced rates of photosynthesis and altered carbohydrate metabolism in high light, while TPP overexpression abolishes

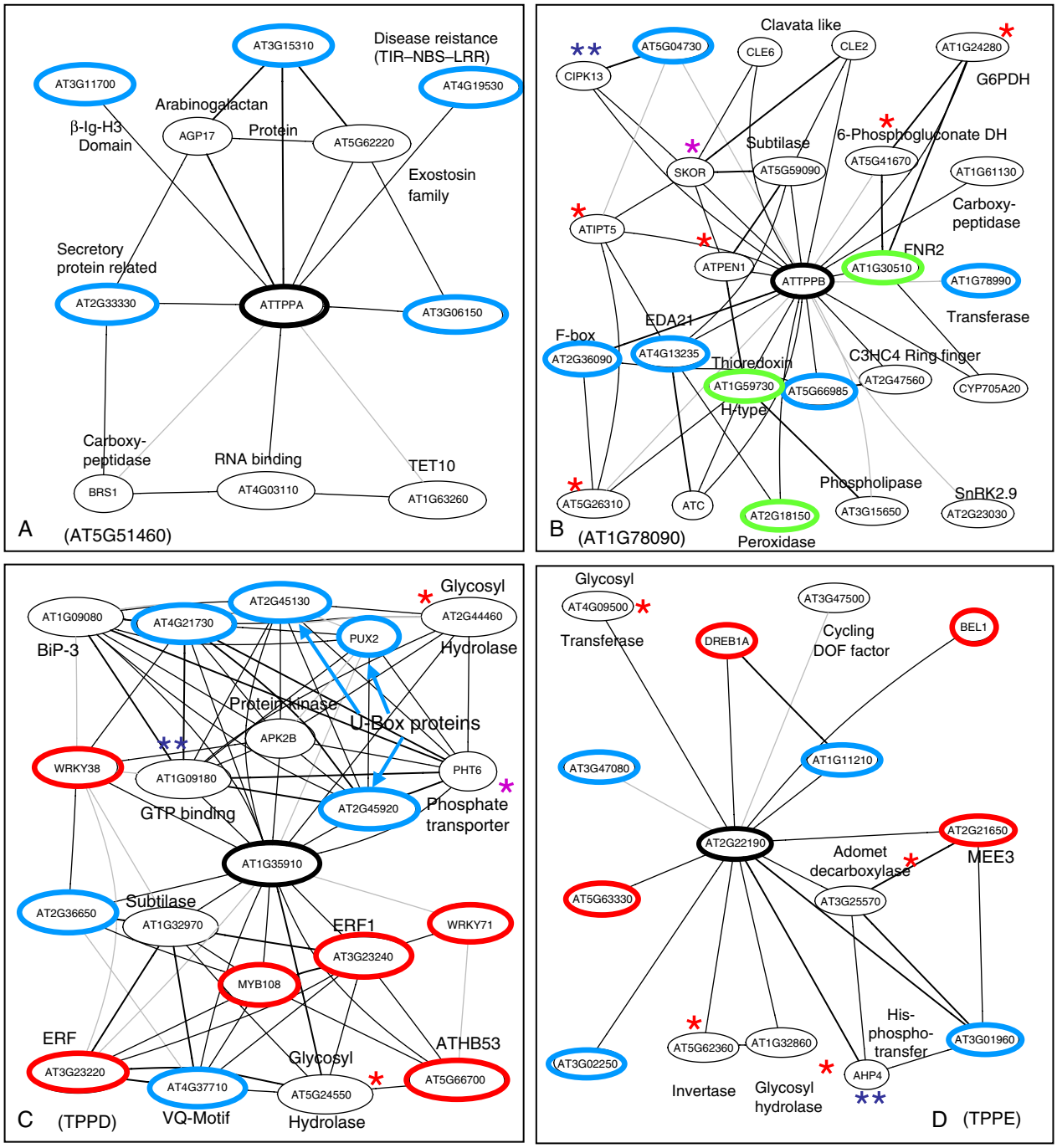

Fig. 4. GGM network centered on Arabidopsis TPP genes. (A) TPPA, (B) TPPB, (C) TPPD, (D) TPPE, (E) TPPF, (F) TPPG, (G) TPPI, (H) TPPJ and (I) TPPH. The network model used is of relaxed stringency and includes >14 000 genes, i.e. approximately 50% of the transcribed Arabidopsis genome (Ma and Bohnert 2008). Most genes are modeled to extend one edge from the TPP genes; TPPH and TPPF are modeled two edges from the seed gene. Blue circles, functionally unknown genes; red circles, transcription factor genes; green circles, ROS generation or scavenging; *(red), metabolism; *(purple), transporters; **(blue), signaling. Additional abbreviations and gene names are included in the Supplementary material Table S1.

this effect, indicating T6P as the important metabolite (Paul et al. 2001). T6P is now recognized as a signaling molecule that responds to changes in cellular redox state and is involved in the regulation of sugar and starch metabolism (Karim et al. 2007, Kolbe et al. 2005, Sokolov et al. 2006) and coordinating carbon supply with

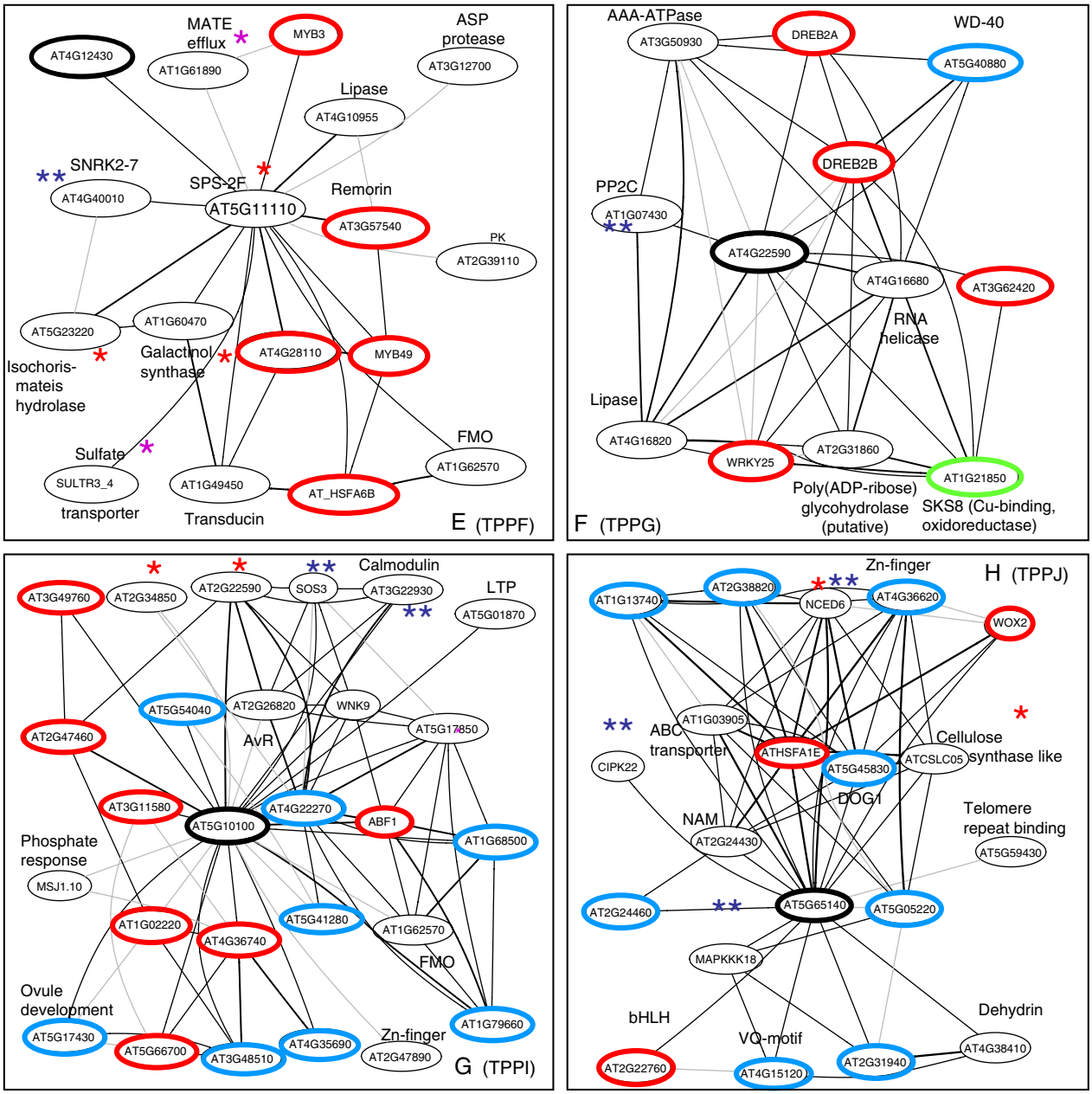

Fig. 4. Continued

development and growth (Chary et al. 2008, Gómez et al. 2006, Satoh-Nagasawa et al. 2006). Among a number of questions about trehalose metabolism, one topic that remains to be understood is the rationale for the complexity of the genes in this pathway. Especially, one would like to understand the pathways that connect the biosynthesis and degradation of T6P to specific events, presumably the perception of some stressful condition, with the transcriptional (and posttranslational) induction or repression of T6P and trehalose biosynthesis or turnover. Such information could be useful for the design of crop protection schemes as they have, e.g. been contemplated through the transgenic modification of trehalose metabolism (Karim et al. 2007, Miranda et al. 2007, Kolbe et al. 2005, Romero et al. 1997). We see the possibility of the coregulation identified by the GGM network to reveal connections between genes that could be exploited in plant engineering.

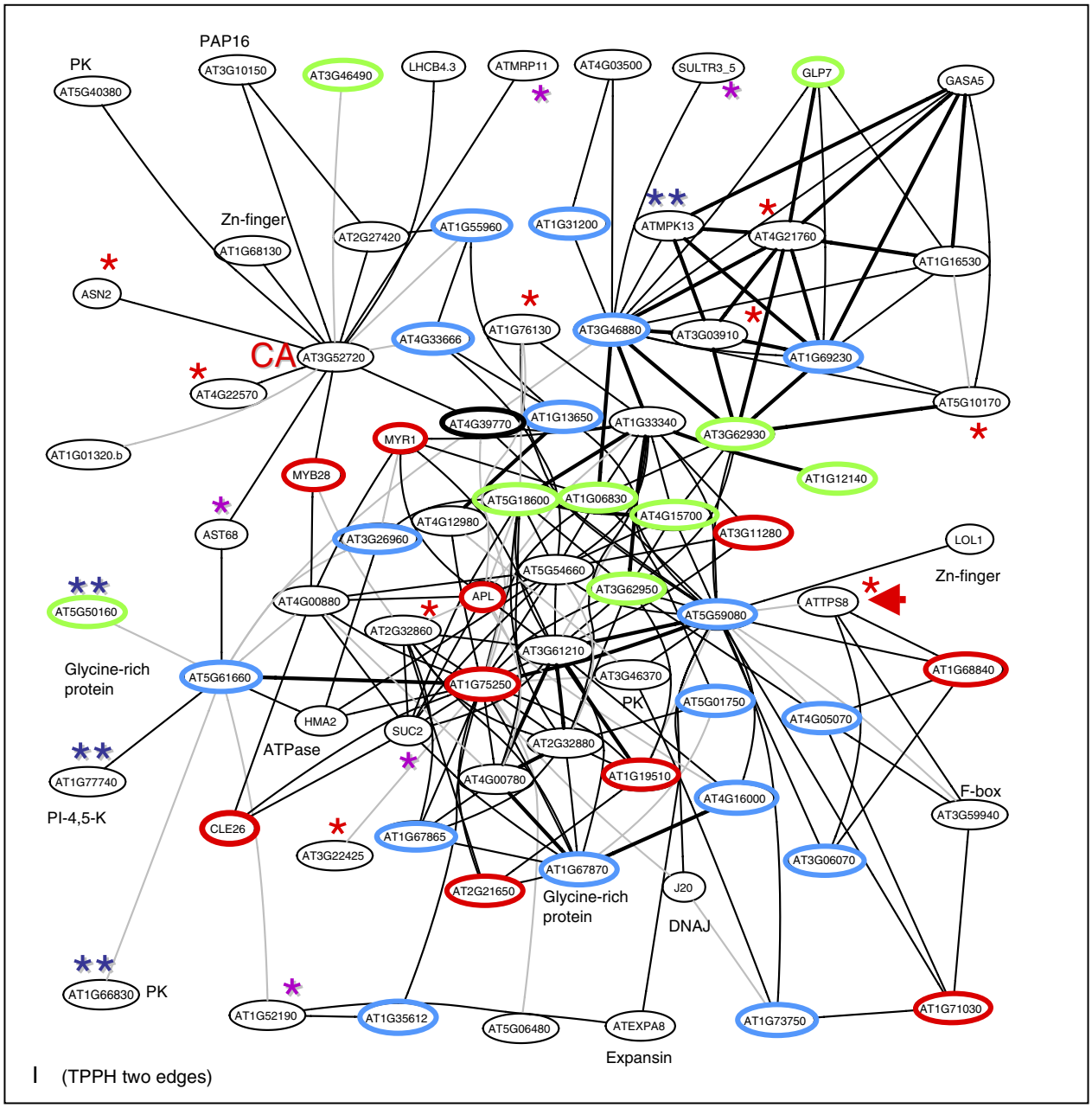

Fig. 4. Continued.

Gene networks

Transcript profiling has generated many complex data and requires extensive analysis that is often incomplete. However, with advanced presentation tools that, at their best, visualize results and facilitate developmental, cellular and conditional integration of the profiles into a whole organism view, transcriptomics allows for insights that inspire new hypotheses and continued studies. The challenge then must organize the clustered genes into networks, preferably with a predictive assessment of coregulation characteristics. Several such tools exist which in their majority depend on Pearson's correlation coefficients (e.g. Obayashi et al. 2007). The GGM network solutions that we present exceed the models that are generated by the correlation coefficient tool because the expression of each gene pair is made contingent on the expression of the genes in a pair in

comparison with all genes in the genome. Irrespectively, such a gene-based network will generate models that are by necessity incomplete because the expression of genes is not always controlled at the level of transcription. Networks that sample protein–protein interaction, protein modifications and metabolite dynamics must eventually be reconciled with the transcription-based network structures. At present, only the information on transcript profiles of Arabidopsis is advanced enough so that one might consider the generation of networks, and tools for an advanced statistical analysis of the transcript profiling data exist (e.g. Schäfer and Strimmer 2005a, 2005b). Further augmentation can now fill in biological information, through expression analysis and biochemical, treatment-specific and developmental pathways (TAIR: http://www.arabidopsis.org; Genevestigator: Zimmermann et al. 2004; http://gabi.rzpd.de/projects/MapMan: Usadel et al. 2005). Once network solutions have been generated for the Arabidopsis genome, the results can be compared with the, less complete, transcript profiles in other species. For example, the information on maize cell wall, 'maizewall' (Guillaumie et al. 2007), biosynthesis genes and their expression characteristics could be modeled into the Arabidopsis GGM network to a large extent because of extensive correlation (data not shown).

TPPs and the gene network

Many network graphs associate individual isoforms of genes in pathway into models that separate orthologs for the same function. Thus, the GGM appears to distinguish pathways in which isoforms of gene families exert their action. The graphs outlining the network for nine TPP genes in Arabidopsis (Fig. 4) are no exception – each isoform is identified by a distinct set of genes with which it is coregulated and associated. This predictive aspect and quality of the gene network will assist in the formation of testable hypotheses about isoform functions. Another feature of the model is that it supports with high significance pathways in developmental, biochemical and stress-regulated functions that have over the past decade been established by analyses centered on individual genes/mutants (Ma et al. 2007). In addition, the GGM invariably includes additional genes that have not previously been associated with a specific pathway or process. In most cases, the additionally identified genes have not been studied. Finally, in a number of biochemical pathway models, GGM includes most of the pathway genes, while several genes in the pathway that are constitutively expressed are missing. Instead, these pathways typically include genes – most of unknown functions – that by their domain structure appear to encode functions that lead to posttranslational modifications in other proteins. It seems possible that these modifiers target the pathway proteins that are not included in the coexpression network.

In applying these apparent rules to the TPPs, we cannot assign a function to all genes. Several inferences appear possible that may be pursued by additional studies. *ATTPPB* and *ATTPPH* appear as genes situated in networks that show a strong relationship to redox, ROS and kinase-related signaling pathways. *ATTPPG* (and possibly also *ATTPPF*) shows a close relationship to drought-related transcription factors, and *ATTPPE* may play a role in seed maturation and/or embryogenesis. It also seems possible that *ATTPPI* could be placed into a pathway, or pathways, that shows less of a relationship to stresses and sugar sensing, carbohydrate biochemistry and photosynthesis.

Acknowledgements – Support at some time by the US National Science Foundation, plant genome program, DBI-9813360 and DBI-0223905, and by University of Illinois at Urbana-Champaign institutional funding is gratefully acknowledged.

References

Agarwal PK, Agarwal P, Jain P, Jha B, Reddy MK, Sopory SK (2006) Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. Plant Cell Rep 25: 1263–1274

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408: 796–815.

Avonce N, Mendoza-Vargas A, Morett E, Iturriaga G (2006) Insights on the evolution of trehalose biosynthesis. BMC Evol Biol 6: 109

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5: 101–113

Carey VJ, Gentry J, Whalen E, Gentleman R (2005) Network structures and algorithms in Bioconductor. Bioinformatics 21: 135–136

Carey VJ, Long L (2006) RBGL: interface to boost C++ graph lib R package version 1100. Available at http://www.bioconductor.org

Chary SN, Hicks GR, Choi YG, Carter D, Raikhel NV (2008) Trehalose-6-phosphate synthase/phosphatase regulates cell shape and plant architecture in Arabidopsis. Plant Physiol 146: 97–107

Craigon DJ, James N, Okyere J, Higgins J, Jotham J, May S (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. Nucleic Acids Res 32: D575–D577

Eastmond PJ, Li Y, Graham IA (2003) Is trehalose-6-phosphate a regulator of sugar metabolism in plants? J Exp Bot 54: 533–537

Gansner ER, North SC (2000) An open graph visualization system and its applications to software engineering. Softw Pract Exper 30: 1203–1233

Garg AK, Kim JK, Owens TG, Ranwala AP, Choi YD, Kochian LV, Wu RJ (2002) Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proc Natl Acad Sci USA 99: 15898–15903

Gasch AP, Eisen MB (2002) Exploring the conditional co-regulation of yeast gene expression through fuzzy k-means clustering. Genome Biol 3: R0059

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80

Gómez LD, Baud S, Gilday A, Li Y, Graham IA (2006) Delayed embryo development in the Arabidopsis trehalose-6-phosphate synthase 1 mutant is associated with altered cell wall structure, decreased cell division and starch accumulation. Plant J 46: 69–84

Gong Q, Li P, Ma S, Indu Rupassara S, Bohnert HJ (2005) Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. Plant J 44: 826–839

Guillaumie S, San-Clemente H, Deswarte C, Martinez Y, Lapierre C, Murigneux A, Barrière Y, Pichon M, Goffner D (2007) MAIZEWALL Database and developmental gene expression profiling of cell wall biosynthesis and assembly in maize. Plant Physiol 143: 339–363

Karim S, Aronsson H, Ericson H, Pirhonen M, Leyman B, Welin B, Mäntylä E, Palva ET, Van Dijck P, Holmström KO (2007) Improved drought tolerance without undesired side effects in transgenic plants producing trehalose. Plant Mol Biol 64: 371–386

Kolbe A, Tiessen A, Schluepmann H, Paul M, Ulrich S, Geigenberger P (2005) Trehalose 6-phosphate regulates starch synthesis via posttranslational redox activation of ADP-glucose pyrophosphorylase. Proc Natl Acad Sci USA 102: 11118–11123

Leyman B, Van Dijck P, Thevelein JM (2001) An unexpected plethora of trehalose biosynthesis genes in Arabidopsis thaliana. Trends Plant Sci 6: 510–513

Liu F, Vantoai T, Moy LP, Bock G, Linford LD, Quackenbush J (2005) Global transcription profiling reveals comprehensive insights into hypoxic response in Arabidopsis. Plant Physiol 137:1115–1129

Ma S, Bohnert HJ (2007) Integration of *Arabidopsis thaliana* stress-related transcript profiles, promoter structures, and cell-specific expression. Genome Biol 8: R49

Ma S, Bohnert HJ (2008) Gene networks in *Arabidopsis thaliana* for metabolic and environmental functions. Mol BioSyst, in press. doi: 10.1039/b715811b

Ma S, Gong Q, Bohnert HJ (2006) Dissecting salt stress pathways. J Exp Bot 57: 1097–1107

Ma S, Gong Q, Bohnert HJ (2007) An Arabidopsis gene network based on the graphical Gaussian model. Genome Res 17: 1614–1625

Miranda JA, Avonce N, Suárez R, Thevelein JM, Van Dijck P, Iturriaga G (2007) A bifunctional TPS-TPP enzyme from yeast confers tolerance to multiple and extreme abiotic-stress conditions in transgenic Arabidopsis. Planta 226: 1411–1421

Obayashi T, Kinoshita K, Nakai K, Shibaoka M, Hayashi S, Saeki M, Shibata D, Saito K, Ohta H (2007) ATTED-II: a database of co-expressed genes and *cis*-elements for identifying co-regulated gene groups in Arabidopsis. Nucleic Acids Res 35: D863–D869

Opgen-Rhein R, Schafer J, Strimmer K (2006) GeneNet: modeling and inferring gene networks R package version 101. Available at http://www.strimmerlab.org/software/genenet/

Østergaard L, Yanofsky MF (2004) Establishing gene function by mutagenesis in *Arabidopsis thaliana*. Plant J 39: 682–696

Paul M (2007) Trehalose 6-phosphate. Curr Opin Plant Biol 10: 303–309

Paul M, Pellny T, Goddijn O (2001) Enhancing photosynthesis with sugar signals. Trends Plant Science 6: 197–200

Pramanik MH, Imai R (2005) Functional identification of a trehalose 6-phosphate phosphatase gene that is involved in transient induction of trehalose biosynthesis during chilling stress in rice. Plant Mol Biol 58:751–762

Ramon M, Rolland F (2007) Plant development: introducing trehalose metabolism. Trends Plant Sci 12: 185–188

Romero C, Belles JM, Vaya JL, Serrano R, Culianez-Macia FA (1997) Expression of the yeast trehalose-6-phosphate synthase gene in transgenic tobacco plants: pleiotropic phenotypes include drought tolerance. Planta 201: 293–297

Roth U, von Roepenack-Lahaye E, Clemens S (2006) Proteome changes in *Arabidopsis thaliana* roots upon exposure to Cd^{2+}. J Exp Bot 57: 4003–4013

Satoh-Nagasawa N, Nagasawa N, Malcomber S, Sakai H, Jackson D (2006) A trehalose metabolic enzyme controls inflorescence architecture in maize. Nature 441: 227–230

Schäfer J, Strimmer K (2005a) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat Appl Genet Mol Biol 4: Article32

Schäfer J, Strimmer K (2005b) An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics 21: 754–764

Schluepmann H, Pellny T, van Dijken A, Smeekens S, Paul M (2003) Trehalose 6-phosphate is indispensable for carbohydrate utilization and growth in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 100: 6849–6854

Schluepmann H, van Dijken A, Aghdasi M, Wobbes B, Paul M, Smeekens S (2004) Trehalose mediated growth inhibition of Arabidopsis seedlings is due to trehalose-6-phosphate accumulation. Plant Physiol 135: 879–890

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. Nat Genet 37: 501–506

Shima S, Matsui H, Tahara S, Imai R (2007) Biochemical characterization of rice trehalose-6-phosphate phosphatases supports distinctive functions of these plant enzymes. FEBS J 274: 1192–1201

Smeekens S (2000) Sugar-induced signal transduction in plants. Annu Rev Plant Physiol Plant Mol Biol 51: 49–81

Sokolov LN, Dominguez-Solis JR, Allary AL, Buchanan BB, Luan S (2006) A redox-regulated chloroplast protein phosphatase binds to starch diurnally and functions in its accumulation. Proc Natl Acad Sci USA 103: 9732–9737

Usadel B, Nagel A, Thimm O, Redestig H, Blaesing OE, Palacios-Rojas N, Selbig J, Hannemann J, Piques MC, Steinhauser D, Scheible WR, Gibon Y, Morcuende R, Weicht D, Meyer S, Stitt M (2005) Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. Plant Physiol 138: 1195–1204

van Dijken AJ, Schluepmann H, Smeekens SC (2004) Arabidopsis trehalose-6-phosphate synthase 1 is essential for normal vegetative growth and transition to flowering. Plant Physiol 135: 969–977

Vogel G, Aeschbacher RA, Müller J, Boller T, Wiemken A (1998) Trehalose-6-phosphate phosphatases from *Arabidopsis thaliana*: identification by functional complementation of the yeast tps2 mutant. Plant J 13: 673–683

Wang R, Okamoto M, Xing X, Crawford NM (2003) Microarray analysis of the nitrate response in Arabidopsis roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. Plant Physiol 132: 556–567

Zeng H, Luo L, Zhang W, Zhou J, Li Z, Liu H, Zhu T, Feng X, Zhong Y (2007) PlantQTL-GE: a database system for identifying candidate genes in rice and Arabidopsis by gene expression and QTL information. Nucleic Acids Res 35: 879–882

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR Arabidopsis microarray database and analysis toolbox. Plant Physiol 136: 2621–2632

Supplementary material

The following supplementary material is available for this article:

Table S1. Networks seeded by TPPs genes.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1399-3054.2008.01101.x

(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Edited by K.-J. Dietz