

Coexpression patterns indicate that GPI-anchored non-specific lipid transfer proteins are involved in accumulation of cuticular wax, suberin and sporopollenin

Monika M. Edstam · Kristina Blomqvist ·
Anna Eklöf · Uno Wennergren · Johan Edqvist

Received: 10 February 2012 / Accepted: 12 July 2013 / Published online: 27 July 2013
© Springer Science+Business Media Dordrecht 2013

Abstract The non-specific lipid transfer proteins (nsLTP) are unique to land plants. The nsLTPs are characterized by a compact structure with a central hydrophobic cavity and can be classified to different types based on sequence similarity, intron position or spacing between the cysteine residues. The type G nsLTPs (LTPGs) have a GPI-anchor in the C-terminal region which attaches the protein to the exterior side of the plasma membrane. The function of these proteins, which are encoded by large gene families, has not been systematically investigated so far. In this study we have explored microarray data to investigate the expression pattern of the LTPGs in Arabidopsis and rice. We identified that the LTPG genes in each plant can be arranged in three expression modules with significant coexpression within the modules. According to expression patterns and module sizes, the Arabidopsis module AtI is functionally equivalent to the rice module OsI, AtII corresponds to OsII and AtIII is functionally comparable to OsIII. Starting from modules AtI, AtII and AtIII we generated extended networks with Arabidopsis genes coexpressed with the modules. Gene ontology analyses of the obtained networks suggest roles for LTPGs in the synthesis or deposition of cuticular waxes, suberin and sporopollenin. The AtI-module is primarily involved with cuticular wax, the AtII-module with suberin and the AtIII-module with sporopollenin. Further transcript analysis revealed that several transcript forms exist for several of the LTPG genes

in both Arabidopsis and rice. The data suggests that the GPI-anchor attachment and localization of LTPGs may be controlled to some extent by alternative splicing.

Keywords LTP · Lipid transfer protein · Wax · Sporopollenin · Suberin · Coexpression · Microarray · Alternative splicing

Introduction

The non-specific lipid transfer proteins (nsLTPs) constitute a large protein family specific for plants. This protein family evolved when plants colonized land, as they are found in all land plants, but not in green alga (Edstam et al. 2011). In the nsLTPs eight conserved cysteines are localized in a motif with the general form C-X_n-C-X_n-CC-X_n-CXC-X_n-C-X_n-C (José-Estanyol et al. 2004). The cysteines form four disulphide bonds that stabilize the tertiary structure of the proteins, making them very resistant to heat denaturation and proteolytic digestion (Lindorff-Larsen and Winther 2001). The compact structure consists of four to five α -helices with a central hydrophobic cavity that is suitable for binding hydrophobic ligands (Shin et al. 1995; Lascombe et al. 2008). The nsLTPs that have been examined for lipid binding are promiscuous and bind many different hydrophobic or amphiphilic molecules, including alkanes, fatty acids, fatty acyl-coenzym A and phospholipids (Sodano et al. 1997; Zachowski et al. 1998; Guerbet et al. 1999).

Initially nsLTPs were thought to be involved in intracellular lipid trafficking, but this has later been excluded due to the fact that nsLTPs possess an N-terminal signal sequence leading them to the extracellular space. Their exact in vivo functions have not yet been clarified, even

Electronic supplementary material The online version of this article (doi:10.1007/s11103-013-0113-5) contains supplementary material, which is available to authorized users.

M. M. Edstam · K. Blomqvist · A. Eklöf · U. Wennergren ·
J. Edqvist (✉)
IFM, Linköping University, 581 83 Linköping, Sweden
e-mail: johan.edqvist@liu.se

though the nsLTPs have been known for almost three decades (Kader et al. 1984). One nsLTP from Arabidopsis has been suggested to be involved in long-distance signaling during pathogen defense and there are also several examples of nsLTPs showing antifungal or antibacterial properties in vitro (Nielsen et al. 1996; Maldonado et al. 2002; Wang et al. 2004; Kirubakaran et al. 2008). Further, there are several papers reporting an involvement in formation of the protective cuticle and also some reports suggesting roles in reproduction, e.g. pollen tube adhesion and pollen wall development (Sterk et al. 1991; Thoma et al. 1994; Park et al. 2000; Cameron et al. 2006; Zhang et al. 2008; DeBono et al. 2009; Lee et al. 2009). Some nsLTPs are also expressed abundantly during seed germination and possibly have a role in lipid recycling (Edqvist and Farbos 2002; Eklund and Edqvist 2003).

We have previously divided the nsLTPs into four major and several minor types according to sequence similarity, intron position and spacing between the cysteine residues (Edstam et al. 2011). In one of the major types, Type G, the transcripts encode a C-terminal signal sequence in addition to the N-terminal one, leading to a posttranslational modification where a glycosylphosphatidylinositol(GPI)-anchor is added to the protein. The GPI-anchor attaches the protein to the extracellular side of the plasma membrane. GPI-anchored proteins are found in all eukaryotic organisms and are involved in different functions including cell-to-cell interactions, immune recognition and polarized cell growth (Wang et al. 2002; Ahmad et al. 2003; Ghiran et al. 2003). In plants, proteins with GPI-anchors are involved in many different processes, like callose deposition and metabolism in bud dormancy release, in cell-to-cell communication and in polarized pollen tube growth (Lalanne et al 2004; Simpson et al. 2009; Rinne et al. 2011).

Systematic functional analysis of the nsLTPs have been hampered due to the fact that they are encoded by large gene families and that the genes are likely to be functionally redundant. This complicates the usage of genetic tools, such as T-DNA insertion mutants. In this work, we decided to build a platform for further research of the biological function of these enigmatic proteins by using microarray data to investigate coexpression patterns. Coexpression of genes may indicate an involvement in the same biological processes. Therefore the identification of coexpression networks may open for discoveries of gene function. Here, we focused our attention to the Type G nsLTPs (LTPGs). This selection was done to limit the number of genes in the investigation, but also due to the fact that there are reports that associate a phenotype, less cuticular wax or less wax components, with lowered expression of LTPGs in Arabidopsis (DeBono et al. 2009; Lee et al. 2009; Kim et al. 2012). We constructed modules of coexpressed LTPGs in rice and Arabidopsis. From the

Arabidopsis modules we built extended networks by searching the whole Arabidopsis transcriptome for genes coexpressed with each LTPG-module. The networks were analyzed for enrichments in Gene Ontology terms in order to obtain clues to biological function of the LTPG modules. Our data suggest that the LTPGs are involved in the accumulation of cuticular waxes as indicated previously. However, we also show that the LTPGs may be involved in the biosynthesis or deposition of suberin and sporopollenin. We also characterized the splicing pattern of the Arabidopsis and rice LTPG transcripts and show that many undergo alternative splicing which leads to transcript isoforms with or without the GPI-anchor attachment signal.

Materials and methods

Sequences and Sequence Tools

Previous studies have identified 34 LTPGs from Arabidopsis and 27 from rice (Boutrot et al. 2008; Edstam et al. 2011). These sequences were initially used in this study. The Arabidopsis sequences were retrieved from The Arabidopsis Information Resource (TAIR, version 10) and the rice sequences from Rice Genome Annotation Project (TIGR RGAP, version 6.1) (Rhee et al. 2003; Ouyang et al. 2007). Basic Local Alignment Search Tool (BLAST v 2.2.18) was used locally to search for additional putative LTPGs among downloaded Arabidopsis and rice protein sequences (Altschul et al. 1990). All known sequences from each organism were used as bait and all settings were left as default (Matrix BLOSUM62, gap penalties: Existence 11 and extension 1). The cut off value was set to 0.0001. Results were manually investigated and false hits removed.

BLAST searches were also performed online at The Arabidopsis Information Resource (TAIR, database *TAIR10 transcripts*) and Rice Genome Annotation Project (RGAP, database *Rice full length cDNAs in Genbank*). Expressed Sequence Tag (EST) databases were searched using genomic sequences as bait, in order to find introns and reveal alternative splicing. PredGPI (available at <http://gpcr2.biocomp.unibo.it/gpipe/pred.htm>) was used to predict presence of sites for the post translational addition of a GPI-anchor (Pierleoni et al. 2008). PredGPI was used for all expressed isoforms of each sequence. TargetP 1.1 was used to predict subcellular targeting (Emanuelsson et al. 2007).

A phylogenetic tree was constructed to visualize relations between the LTPGs in Arabidopsis and rice. The tree is based on multiple alignments done with the ClustalW method (Thompson et al. 1994), using the program ClustalW2 v2.0.7 (Larkin et al. 2007). The alignments were

done using the slow/accurate method and the protein matrix Gonnet. The gap extension penalty and gap opening penalty were set to 0.01 and 10, respectively. Only the cores of the mature proteins including the conserved Cys residues were used in alignments; the GPI-anchor and the link to the anchor were excluded.

Manually refined alignments were used as input in ProtTest v2.4 (Abascal et al. 2005), which was run with all candidate models, a BIONJ tree and the slow optimization strategy. LG +I +G was predicted as the best model and thus used to construct a Maximum Likelihood phylogenetic tree using the program Phylml v3.0 (Guindon and Gascuel 2003). All other settings were left as default, but with 100 replicates for bootstrapping. Three LTPGs from *Physcomitrella patens* were used as outgroup (Edstam et al. 2011). Additionally, searches for overrepresented motifs in the promoter regions (0–2,000 bp upstream the start codon) of all genes in each module were performed. The web tool *elfinder* at Matt Hudson Lab was used for this purpose (<http://stan.cropsci.uiuc.edu/tools.php>).

Expression and coexpression networks

The eFP browser for Arabidopsis at The Bio-Array Resource for Plant Biology were used to retrieve expression data from different tissues and developmental stages, during stresses, hormone and chemical treatment (Schmid et al. 2005; Kilian et al. 2007; Winter et al. 2007; Goda et al. 2008). Further, coexpression networks of LTPGs from Arabidopsis and rice were constructed. Pairwise Pearson correlation coefficients between the LTPGs were obtained using the web tool Cornet (<https://cornet.psb.ugent.be/main/tool>) for Arabidopsis and the coexpression analysis tool from the Rice Oligonucleotide Array Database (<http://www.ricearray.org/>) for rice (Jung et al. 2008). The correlation coefficients were obtained from six predefined datasets in Arabidopsis (*All*, *Development*, *Whole Plant*, *Hormone treatment*, *Biotic Stress* and *Abiotic Stress*) and three in rice (*General*, *Biotic Stress* and *Abiotic Stress*). A correlation between two genes was considered present when the coefficient was higher than 0.7. A network consisting of three or more LTPGs genes with correlation coefficients above 0.7 was considered as a module. In Arabidopsis, each module was then used to make genome wide networks of genes coexpressed with the LTPGs genes. To expand the modules of LTPGs to genome wide coexpression networks, every gene with a correlation coefficient higher than 0.7 towards any of the LTPGs in the module was connected to that module. This was performed in the dataset *All*. The software Cytoscape was used for visualization of the resulting networks of coexpressed genes (Shannon et al. 2003).

Cluster analysis

As an additional method to identify groups of coexpressed genes in the data set we used a fuzzy clustering algorithm (Kaufman and Rousseeuw 2008) as opposed to hard clustering. For our analysis we used the “fanny” function in the R-package “cluster”. The parameter m defines the degree of fuzzification allowed between clusters. As m approaches 1, the fuzzy clusters become hard clusters, where each data point belongs to only one cluster. As m approaches infinity the clusters become completely fuzzy, and each point will belong to each cluster to the same degree ($1/K$, where K is the number of clusters analyzed) regardless of the data. Usually $m = 2$ is initially chosen, and this is also the value we use here. However, we evaluated our choice of value of m using an exhaustive grid-search varying m and K (Futschik and Kasabov, 2002). In order to find the number of clusters K that describes the best partitioning of the data set, one usually executes the clustering algorithm with different numbers of the expected number of clusters K ($K_m < K < K_M$). Then a quality index Q_K is computed for each value of K tested and the K giving the “best” value Q_K is chosen. We used the R-package “clValid” to validate the best number of clusters using all three validation measures included in the validation parameter “internal”.

Gene Ontology Enrichments

To search for gene ontology enrichments in the constructed genome wide networks for Arabidopsis, the plugin BiNGO in Cytoscape was used (Maere et al. 2005). As a statistical test the hypergeometric test was used, together with the Benjamin and Hochberg False Discovery Rate (FDR) correction. The significance level was set to 0.05 and whole Arabidopsis annotation was used as a reference set. Three different sets of ontology terms were used separately: *Biological Process*, *Molecular Function* and *Cellular Component* (Berardini et al. 2004).

RNA analysis

Four of the genes that were predicted to be alternative spliced were further investigated *in planta*. The *Arabidopsis thaliana* ecotype Col-0 was used for all experiments. Seeds were sown on agar plates containing 1/2 strength Murashige and Skoog medium supplemented with 1 % sucrose. After 14 days the seedlings were transferred to a mixture of soil and vermiculite. The plants were grown either under a light cycle of 16 h light and 8 h dark in a growth chamber, or under constant light. RNA from leaves, flowers, siliques and roots was extracted using the RNeasy Plant Minikit (Qiagen, Hilden, Germany) according to the

manufacturers' protocol. The extracted RNA was treated with DNase and then used as template in a cDNA synthesis. For each reaction 1 µg RNA was used. Oligo(dT)₁₈ primer was used to avoid amplification of immature mRNA. RevertAid Reverse Transcriptase (Fermentas, Vilnius, Lithuania) was used for first strand cDNA synthesis according to the manufacturers' protocol. To ensure that no traces of genomic DNA were contaminating the samples an additional cDNA synthesis was performed without the Reverse Transcriptase, as a negative control. The synthesized cDNAs were used for PCR with gene specific primers (Online Resource 1). DreamTaq DNA Polymerase (Fermentas) was used for the PCR, according to the manufacturers' protocol. The PCR was performed as follows: 3 min of initial denaturation at 95 °C, followed by 35 cycles of 30 s denaturation at 95 °C, 30 s annealing at 55 °C and 1 min elongation at 72 °C. After the cycling, there was a final elongation step at 72 °C for 7 min. The PCR products were run on an agarose gel (2 %), fragments excised and DNA recovered using a QIAquick gel extraction kit (Qiagen). Extracted DNA was sent to Eurofins MWG Operon (Ebersberg, Germany) for sequencing.

Results

LTPGs in Arabidopsis and rice

We initiated this study by identifying the complete set of LTPGs in Arabidopsis and rice (Tables 1, 2). Previously, 34 LTPGs genes in Arabidopsis and 25 LTPGs in rice have been identified in genome-wide analyses (Boutrot et al. 2008). During this study two additional rice genes were identified giving a current total number of 27 GPI-anchored LTPGs in rice. The occurrence of the GPI-anchors in nsLTPs is mainly based on predictions. However, GPI-anchors have been shown experimentally for AtLTPG1, AtLTPG11, AtLTPG12 and AtLTPG31 (Borner et al. 2003; Elortza et al. 2003). When transcriptome databases were searched for transcripts of LTPG genes, three of the Arabidopsis genes and two of the rice genes lacked a corresponding transcript in the databases (Tables 1, 2). The genes lacking matching transcripts were considered as putative pseudogenes and were removed from the remaining investigation of the expression profiles. However, we do not exclude the possibility that transcripts from all or some of these five genes could possibly be identified during conditions not yet investigated. The intracellular localizations of the identified LTPGs were investigated using the subcellular predictor TargetP (Emanuelsson et al. 2007). As expected, most of the proteins are predicted to be secreted. More surprisingly there are three Arabidopsis proteins and two rice proteins that by TargetP are assigned to other

localizations, such as chloroplast and mitochondria (Tables 1, 2). However, most of these predictions to chloroplast and mitochondria show low reliability scores and the localization of LTPGs to organelles should be experimentally verified.

Coexpression of LTPGs

We reasoned that there are probably functional groups of LTPG genes that are involved in related biological processes. Further, if we could identify functional groups it would be helpful for the rational design of experiments aiming at elucidating the biological role of these proteins. LTPG genes involved in the same process are likely to share correlated expression profiles. Therefore, to identify functional groups of LTPGs we turned our attention to Arabidopsis and also rice microarray datasets. For Arabidopsis, we used six different microarray datasets: *All*, *Whole Plant*, *Development*, *Hormone*, *Biotic* and *Abiotic* stresses (De Bodt et al. 2009). Expression data for 26 LTPG genes were available in these datasets. We treated the microarray datasets separately, to learn if identified coexpression patterns would be based on for instance stress responses or developmental programs.

The coexpression between the LTPGs was obtained as Pearson correlation coefficients (R). At first, coexpressed LTPG genes were identified using an arbitrary threshold of $R > 0.7$. This cut off was selected since $R > 0.7$ is generally considered a true correlation and used in various analysis (Lee et al. 2004; Ren et al. 2005; Zheng et al. 2008). In each dataset we could identify 3–5 groups of Arabidopsis LTPG genes, which according to the definition above, were coexpressed (see Online Resources 2–7). After identifying the groups of coexpressed genes in each dataset, we next placed the LTPGs in composite coexpression modules. To be included in a composite coexpression module a gene had to be a part of a specific coexpression group in at least four of the six investigated microarray datasets. The composite coexpression modules therefore reflect the stability of the coexpression groups in a larger number of samples. Using this arbitrary threshold approach, 14 of the 26 Arabidopsis LTPG genes could be distributed in three different composite coexpression modules (Table 3). The remaining 12 genes were not showing strong enough coexpression to other LTPGs to be placed in any of the composite modules. However, these unplaced genes could show an $R > 0.7$ to a gene within the modules in some datasets, as described below.

According to the arbitrary threshold approach, three genes, AtLTPG1, AtLTPG2 and AtLTPG6, are placed in the composite module AtI (Table 3). These three genes are coexpressed in all six datasets. The only exception is that AtLTPG6 is not reaching the cut off $R > 0.7$ in the *Biotic*

Stress dataset. However, in this dataset the highest correlation coefficient for AtLTPG6 is 0.68, and thus very close to the cut off. None of the other LTPG genes are found in this module in any of the datasets (see Online Resources 2–7). Module AtII is the largest composite module with seven genes; AtLTPG5, AtLTPG15, AtLTPG16, AtLTPG17, AtLTPG20, AtLTPG22 and AtLTPG30. Five of the genes are found in this module in every dataset (AtLTPG15, AtLTPG16, AtLTPG20, AtLTPG22 and

AtLTPG30). AtLTPG17 is below the cut off value in the *Whole Plant* dataset, with a highest correlation coefficient of 0.58 towards another gene in AtII. AtLTPG5 is missing the cut in both the *Whole Plant* and the *Biotic Stress* datasets, but the highest correlation coefficient is not far below in any of the cases (0.67 and 0.65, respectively). Three LTPG-genes (AtLTPG7, AtLTPG11 and AtLTPG33) outside the composite AtII-module are co-expressed with the module in two datasets each. However, in

Table 1 The LTPGs in Arabidopsis

Name	Locus	Number of introns	Position of first intron	Form of alternative splicing (predicted)	Anchor affected	Predicted localisation
AtLTPG1	At1g27950	1	4	Intron retention	Yes	Secreted (1)
AtLTPG2	At3g43720	2	4	Alternative acceptor site	No	Secreted (3)
AtLTPG3	At1g18280	2	13	–		Secreted (2)
AtLTPG4	At4g08670	2	1	–		Secreted (5)
AtLTPG5	At3g22600	2	7	–		Secreted (2)
AtLTPG6	At1g55260	2	4	Alternative acceptor site	No	Secreted (3)
AtLTPG7	At1g62790	2	4	–		Secreted (1)
AtLTPG8	At1g73550	2	13	Intron retention	Yes	Secreted (2)
AtLTPG9	At1g73560	2	13	–		Secreted (1)
AtLTPG10	At1g73890	2	16	–		Secreted (2)
AtLTPG11	At2g13820	2	4	Intron retention	Yes	Secreted (2)
AtLTPG12	At2g27130	2	4	–		Secreted (2)
AtLTPG13	At2g44290	1	4	–		Secreted (2)
AtLTPG14	At2g44300	1	4	–		Secreted (2)
AtLTPG15	At2g48130	2	4	–		Mitochondria (5)
AtLTPG16	At2g48140	3	4	Alternative acceptor site	Yes	Other (5)
AtLTPG17	At3g22570	0	–	–		Secreted (1)
AtLTPG18	At2g44290	NTI	–	–		Secreted (1)
AtLTPG19	At1g03103	2	4	–		Secreted (5)
AtLTPG20	At3g22620	2	4	–		Secreted (5)
AtLTPG21	At1g05450	4	4	–		Secreted (5)
AtLTPG22	At3g58550	2	4	–		Secreted (4)
AtLTPG23	At1g36150	2	1	–		Secreted (4)
AtLTPG24	At4g12360	2	13	–		Secreted (3)
AtLTPG25	At4g14805	NTI	–	–		Secreted (4)
AtLTPG26	At4g14815	2	4	–		Secreted (2)
AtLTPG27	At4g22630	NTI	–	–		Secreted (2)
AtLTPG28	At4g22666	1	16	Alternative acceptor site	No	Secreted (1)
AtLTPG29	At5g09370	2	7	Exon skipping	Yes	Secreted (2)
AtLTPG30	At5g13900	1	7	–		Secreted (1)
AtLTPG31	At5g64080	2	4	Alternative acceptor site	No	Secreted (1)
AtLTPG32	At1g70250	4	4	–		Secreted (2)
AtLTPG33	At4g22640	0	–	–		Secreted (2)
AtLTPG34	At4g22650	0 ^a	–	–		Chloroplast (4)

The position of the intron is given as number of bases after the eighth cysteine in the 8 cm. The number after the localisation indicates how reliable the prediction is, where 1 is the most reliable and 5 the least

NTI no transcripts identified

^a No transcript were identified in TAIR10 database, but the gene was expressed according to microarray data

Table 2 The LTPGs in rice

Name	Locus	Number of introns	Position of first intron	Form of alternative splicing	Anchor affected	Predicted localisation
OsLTPG1	Os01g59870	2	4	–		Secreted (2)
OsLTPG2	Os03g07100	2	4	–		Secreted (2)
OsLTPG3	Os03g09230	2	4	Intron retention	Yes	Secreted (4)
OsLTPG4	Os03g20760	1	58	–		Secreted (2)
OsLTPG5	Os03g26800	0	–	–		Secreted (1)
OsLTPG6	Os03g26820	1	4	–		Secreted (2)
OsLTPG7	Os03g57970	1	4	–		Secreted (2)
OsLTPG8	Os03g57980	3	4	–		Secreted (3)
OsLTPG9	Os03g57990	1	4	–		Secreted (4)
OsLTPG10	Os03g58940	2	4	–		Secreted (2)
OsLTPG11	Os04g38840	1	4	–		Secreted (2)
OsLTPG12	Os05g41030	2	4	–		Secreted (3)
OsLTPG13	Os06g47200	2	4	Intron retention	Yes	Secreted (2)
OsLTPG14	Os07g07790	1	4	–		Secreted (4)
OsLTPG15	Os07g07870	NTI	–	–		Mitochondria (5)
OsLTPG16	Os07g07860	1	4	–		Secreted (2)
OsLTPG17	Os07g07920	2	4	–		Secreted (1)
OsLTPG18	Os07g07930	1	4	–		Secreted (3)
OsLTPG19	Os07g09970	2	4	–		Secreted (3)
OsLTPG20	Os07g30590	1	4	–		Secreted (5)
OsLTPG21	Os07g43290	2	4	Intron retention	Yes	Secreted (2)
OsLTPG22	Os08g42040	2	4	Alternative acceptor site	No	Secreted (1)
OsLTPG23	Os11g37320	NTI	–	–		Secreted (1)
OsLTPG24	Os06g49770	2	46	Intron retention	Yes	Secreted (1)
OsLTPG25	Os11g37280	1	13	–		Secreted (1)
OsLTPG26	Os03g46150	1	4	Intron retention	No	Mitochondria (2)
OsLTPG27	Os03g46180	1	4	–		Secreted (2)

The position of the intron is given as number of bases after the eighth cysteine in the 8 cm. The number after the localisation indicates how reliable the prediction is, where 1 is the most reliable and 5 the least

NTI no transcripts identified

Table 3 Composite modules of co-expressed genes in Arabidopsis and rice

Module	Genes
<i>Arabidopsis</i>	
AtI	AtLTPG1, AtLTPG2, AtLTPG6
AtII	AtLTPG5, AtLTPG15, AtLTPG16, AtLTPG17, AtLTPG20, AtLTPG22, AtLTPG30
AtIII	AtLTPG3, AtLTPG4, AtLTPG23, AtLTPG26
<i>Rice</i>	
OsI	OsLTPG10, OsLTPG12, OsLTPG22
OsII	OsLTPG7, OsLTPG8, OsLTPG9, OsLTPG14, OsLTPG17, OsLTPG18, OsLTPG26, OsLTPG27
OsIII	OsLTPG1, OsLTPG2, OsLTPG24

the other datasets the correlation is weaker, although in some cases just below the threshold (see Online Resources 2–7). The composite module AtIII consists of four genes;

AtLTPG3, AtLTPG4, AtLTPG23 and AtLTPG26 (Table 3). Only AtLTPG4 is found in this module in all six datasets, the others are below the cut off in one or two datasets each. AtLTPG3 is just below in the *Whole Plant* and *Biotic Stress* datasets (highest coexpression coefficients 0.62 and 0.69). AtLTPG23 is excluded when using the dataset All and AtLTPG26 when using *Development* (highest coexpression coefficients 0.62 and 0.63 respectively). Four genes outside the module (AtLTPG9, AtLTPG24, AtLTPG29 and AtLTPG34) show significant coexpression with module AtIII in two or three of the datasets (see Online Resources 2–7). However, AtLTPG9, AtLTPG24 and AtLTPG34 show much weaker correlation to the AtIII-module in the other datasets. The same is true for AtLTPG29 although in the *Whole Plant* dataset, this gene has a correlation to module AtIII which is only just below the cut off.

In summary, the arbitrary cut off approach results in three modules of coexpressed LTPG genes in Arabidopsis. The module AtI is the most stable over all the tested datasets. In module AtII five of the genes are consistently coexpressed, according to the given definition, over the six datasets, whereas there are a few genes that show a correlated expression to the module only in some datasets. The stability of module AtIII is weaker with only one gene fitting to the module in all datasets.

In rice, 13 of the LTPGs could be placed in either of three composite coexpression modules (Table 3). The rice genes that were placed in expression modules showed a coexpression pattern in at least two of the three investigated microarray datasets: *General*, *Biotic* and *Abiotic* (Jung et al. 2008). Three coexpressed genes (OsLTPG10, OsLTPG12 and OsLTPG22) were grouped in module OsI, eight genes (OsLTPG7, OsLTPG8, OsLTPG9, OsLTPG14, OsLTPG17, OsLTPG18, OsLTPG26 and OsLTPG27) were assigned to module OsII and another three genes (OsLTPG1, OsLTPG2 and OsLTPG24) were placed in module OsIII. Eight of the rice genes did not fit in any module, and thus seem to lack significant coexpression to other genes encoding LTPGs, at least in more than one of the investigated datasets.

A potential problem of using an arbitrary threshold value is by setting this threshold too high, important relationships can be lost. For example, we have with the approach described above identified several genes which are above the threshold in some datasets but below in the majority of the datasets (see Online Resources 2–7). With the approach and cut off threshold used in this study, these genes could not be assigned to any of the expression modules. On the other hand, setting the threshold too low could result in connections that are very weak or possibly false positive results. Therefore, we also used a clustering algorithm to identify groups of coexpressed Arabidopsis LTPG genes in the datasets. Clustering techniques seek to partition a given data set into a set of disjoint groups so that objects within groups are more similar to each other than objects in separate groups (Kaufman and Rousseeuw 2008). The rationale is that many coexpressed genes are co-regulated and important groups can then be revealed with the usage of cluster analysis (Domany, 2003). We here used a fuzzy clustering algorithm (Kaufman and Rousseeuw 2008) as opposed to hard clustering. For hard clustering the clusters are mutually exclusive. Fuzzy clustering, on the other hand, allows data points to belong to several clusters simultaneously. The partial membership is presented as a probability of a data point i belonging to cluster k . In many data sets fuzzy clustering is more natural compared to hard clustering (Do and Choi 2008) since data points on the boundaries between several clusters are not forced to belong to one of them, but rather are assigned a

partial membership between 0 and 1. For a fixed observation the membership probabilities sum to 1.

The clustering was done on the all 26 AtLTPG genes that were available in the microarray datasets. In some datasets (*Abiotic Stress*, *Biotic Stress*) the analysis indicated three clusters, while in other datasets (*Hormone*, *All*) there were support for two clusters. Further, in the remaining datasets (*Development*, *Whole Plant*) the analysis did not indicate any particular number of clusters that best describe the partitioning of the datasets. In general, silhouette width scores over 0.6 is considered as significant. In the case of the AtLTPGs the average silhouette width scores were low for k 2–5 and did not give strong support for the partitioning in any of the datasets, as shown for datasets *All* and *Development* in Fig. 1a. Twelve genes were not placed in the coexpression modules with the threshold approach. Probably, these genes with a low expression correlation to other LTPGs reduce the probability to obtain well-defined clusters. Next, we followed the partitioning in clusters of the genes we previously assigned to expression modules AtI, AtII and AtIII with the arbitrary cutoff approach. When $k = 3$ was used in the analysis we noted that these genes were showing the highest silhouette width scores and therefore most strongly associated with each of the three clusters, as shown for dataset *All* and *Development* in Fig. 1b. This could also be concluded by visualizing the clustering in two dimensions shown in Fig. 2 for datasets *Development*, *Whole Plant* and *Abiotic Stress*, where the AtLTPGs designated previously to modules AtI, AtII and AtIII were found in three separate clusters.

The cluster partitioning was further examined by evaluation of the fuzzy C-Means plots (Fig. 1c and Online Resource 8). The AtI-genes (AtLTPG1, AtLTPG2 and AtLTPG6) are in all six datasets found in the same cluster. However, in the dataset *Hormone*, all three genes show an equal probability for membership in another cluster. The AtII-genes (AtLTPG5, AtLTPG15, AtLTPG16, AtLTPG17, AtLTPG20, AtLTPG22 and AtLTPG30) are also found together in the same cluster in all six datasets. In the dataset *Whole Plant* the AtII-genes AtLTPG5, AtLTPG17, AtLTPG22 and AtLTPG30 also show a lower probability (0.25–0.50) for a second cluster. Further, in the *Biotic Stress* dataset AtLTPG5 and AtLTPG20 have an equal probability for membership in three or two different clusters, respectively (Online Resource 8). As described previously, AtLTPG7, AtLTPG11 and AtLTPG33 showed an $R > 0.7$ to genes in the AtII-module in some datasets. In the fuzzy clustering, these genes and also AtLTPG14 are found in the same clusters as the AtII-genes in all datasets. The fuzzy clustering located the AtIII-genes (AtLTPG3, AtLTPG4, AtLTPG23, AtLTPG26) in all six datasets to a third separate cluster. In dataset *All* AtLTPG3, AtLTPG4, AtLTPG26

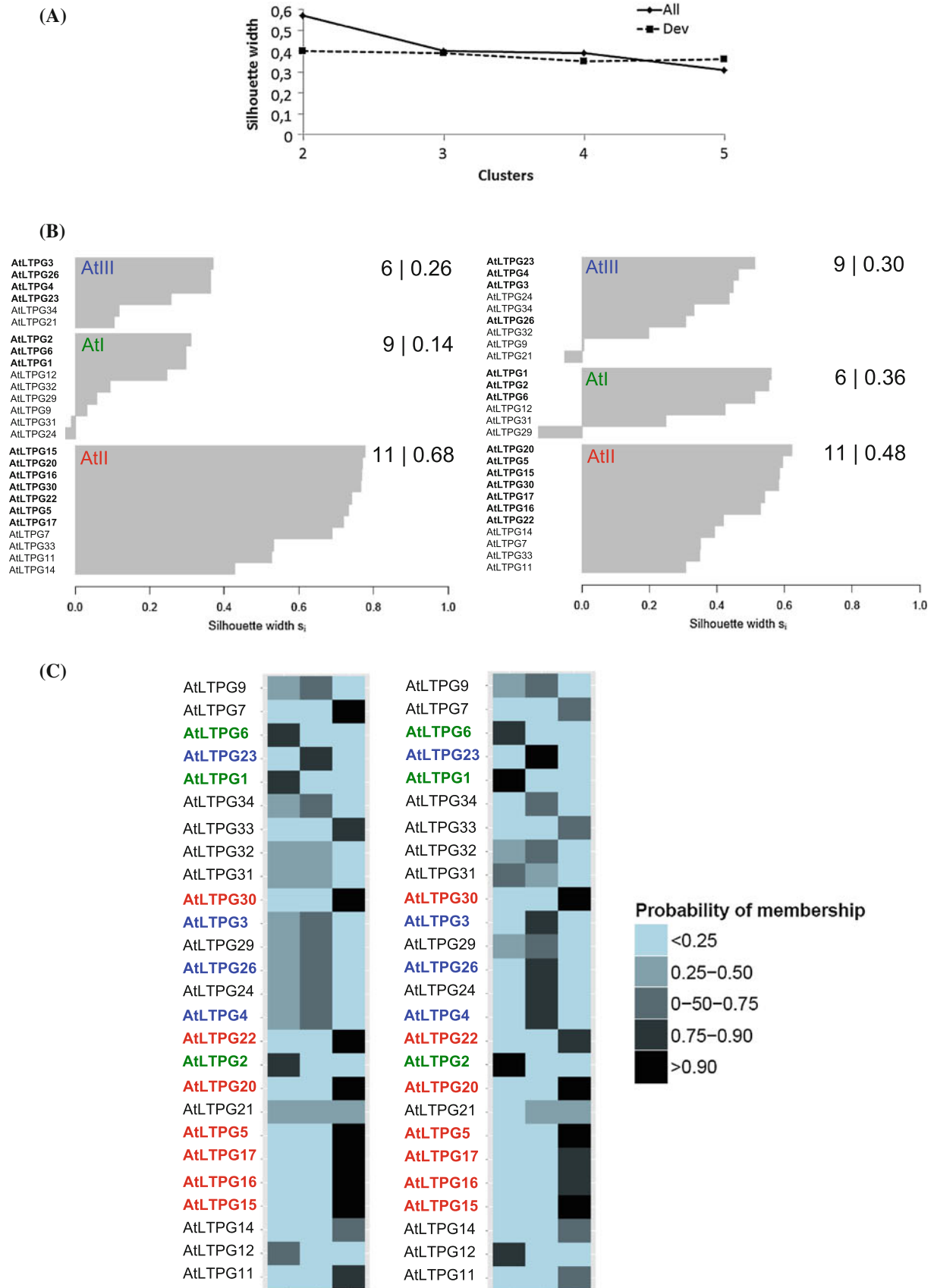


Fig. 1 Fuzzy C-means clustering of AtLTPGs. The graph in (a) shows the average silhouette widths (y-axis) for 2–5 clusters (x-axis) for microarray datasets *All* and *Development*. The graphs in (b) show the silhouette width for each AtLTPG with selection for three clusters. Dataset *All* is in the left panel and dataset *Development* is in the right panel. In both panels the numbers to the right indicate the number of genes in each cluster (left of the vertical line) and the average silhouette width for each cluster (right of the vertical line). In (c) is the fuzzy C-Means plots for three clusters. The plots illustrate to which probability (from 0 to 1) each AtLTPG belongs to each of three clusters. Dataset *All* is in the left panel and dataset *Development* is in the right panel. The genes in module AtI are green, genes in module AtII are red and genes placed in AtIII are blue

have a lower probability for memberships in another cluster, whereas in *Hormone* all four genes show this tendency (Fig. 1c and Online Resource 8). AtLTPG24 and AtLTPG34 showed $R > 0.7$ to the genes in module AtIII in several datasets. In the fuzzy clustering, these genes were found in the same cluster as the AtIII-genes in all datasets. Moreover, they showed a similar promiscuity in the *All* and *Hormone* datasets.

In summary, the fuzzy clustering approach confirmed the coexpression modules identified with the arbitrary

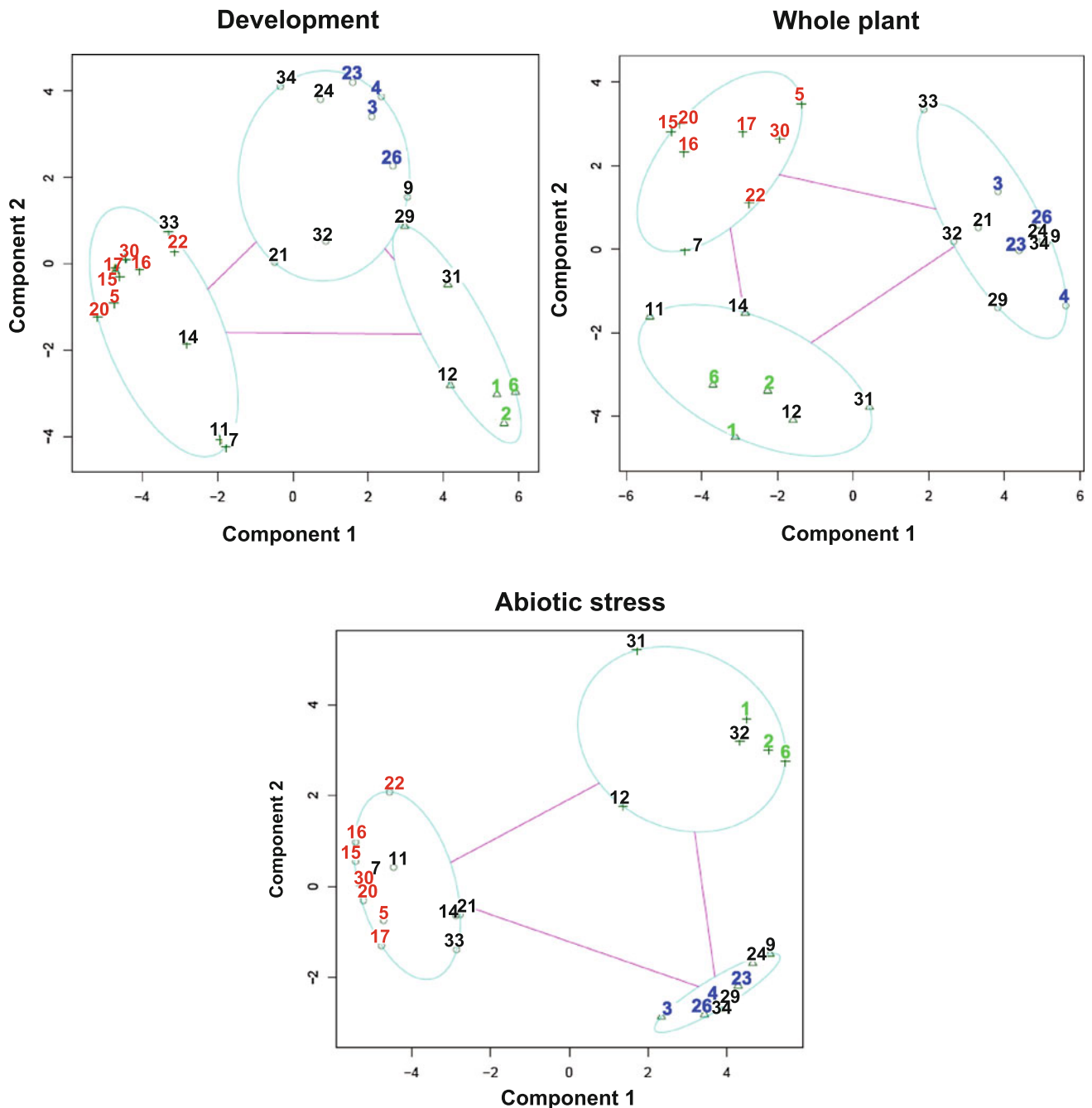


Fig. 2 Fuzzy clustering of AtLTPGs in the datasets *Development*, *Whole Plant* and *Abiotic Stress*. The genes in module AtI are green, genes in module AtII are red and genes placed in AtIII are blue. Clustering were done with $k = 3$

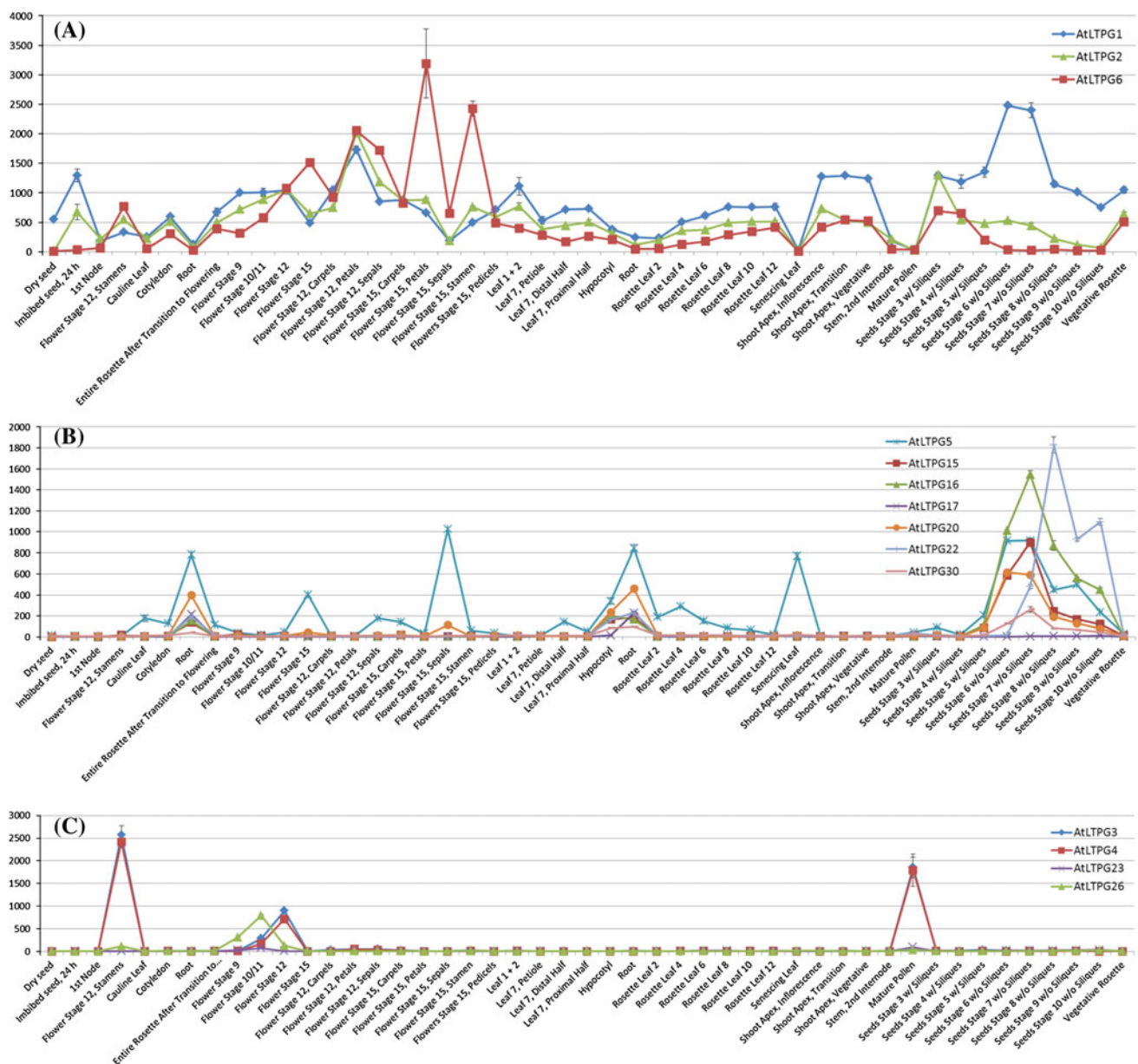


Fig. 3 The expression pattern of members in modules AtI (A), AtII (B) and AtIII (C) during different developmental stages of Arabidopsis. The genes are indicated in the top right corner of each panel.

The Y-axis shows the expression levels of LTPG transcripts. Standard deviation is shown as *error bars*

threshold approach. Moreover, the clustering revealed tendencies for larger networks consisting of more LTPG genes then could be identified with the chosen arbitrary threshold $R > 0.7$. It also clear from both the fuzzy clustering and the arbitrary threshold approach that there are a number of LTPGs that do not correlate strongly in terms of expression profiles with other LTPG genes. For instance, according to the fuzzy clustering AtLTPG9, AtLTPG12, AtLTPG21, AtLTPG29, AtLTPG31 and AtLTPG32 show a probability of at least 0.25 to associate with two or more clusters in most datasets (Online Resource 8).

Expression patterns of the modules

We continued this investigation by screening the microarray datasets for the detailed expression pattern of the Arabidopsis expression modules AtI, AtII and AtIII (Table 3). We focused on these genes since both the arbitrary cut-off method and the fuzzy clustering showed a connection between the genes within each module. In comparison with the other modules, AtI have a high expression baseline. In the *Developmental* dataset all three genes in AtI have their highest expression levels in flower and seed, and the lowest

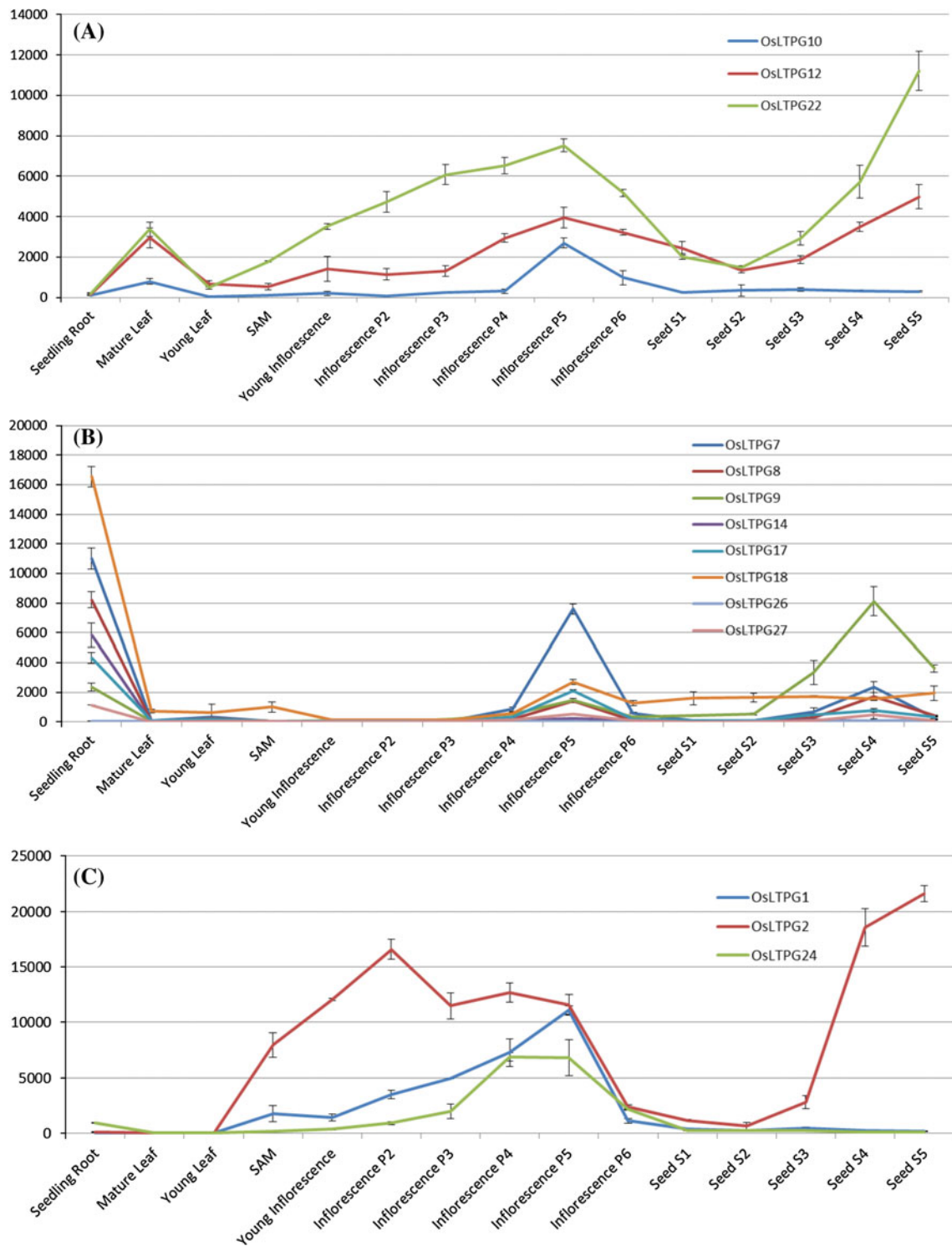


Fig. 4 The expression pattern of members in modules OsI (A), OsII (B) and OsIII (C) during different developmental stages of rice. The genes are indicated in the *top right* corner of each panel. The Y-axis

shows the expression levels of LTPG transcripts. Standard deviation is shown as *error bars*

in root, cauline leaf, senescing leaf and mature pollen (Fig. 3). In the *Abiotic Stress* dataset downregulation were shown for drought, heat and UV-B (Online Resource 9). Wounding causes an upregulation after 1 and 3 h, and then a

downregulation after 6 and 12 h. In the *Biotic Stress* dataset there are no significant upregulations, but several downregulations (Online Resource 10). Interestingly, treatment with the photosynthesis inhibitor *N*-octyl-3-nitro-2,4,6-

trihydroxybenzamide (PNO8) causes a large decrease in the expression of the genes in module AtI (Online Resource 11).

In the *Developmental* dataset all genes in AtII have an expression peak in the roots of both adult plants and seedlings (Fig. 3). The expression of most of the genes in AtII also peak in hypocotyls and seeds, and a few of the genes are also upregulated in flowers. In the abiotic and biotic stress (Online Resources 12–13) there is only one condition that gives a significant change in all AtII-genes; 1 h of drought leads to a downregulation of gene expression. Moreover, all of them show an increased expression 3 h after addition of abscisic acid (ABA), although in different degrees and not in all cases significant (Online Resource 14). Module AtIII is highly expressed in flowers and mature pollen, but at much lower levels in other tissues (Fig. 3). In the *Hormone* dataset there are no big differences in the expression, since the experiments are conducted on seedlings and not flowers. The similar situation is found in the other datasets, resulting in no significant changes in the *Chemical* dataset or stress datasets (Online Resources 15–18). In summary, the most important points from the characterization of the expression patterns are that AtI transcripts are present in most aerial parts, AtII is found in roots, although is not restricted to underground tissues, and AtIII is restricted to reproductive tissues.

In rice, OsI has an expression peak in mature leaves, which distinguishes OsI from the other composite rice modules (Fig. 4). This module also has peaks during inflorescence stage P5 and seeds stage S5. Inflorescence stage P5 corresponds to the vacuolated pollen stage (15–22 cm height), and seed stage S5 corresponds to 21–29 days after pollination (dap); during development of dormancy and desiccation tolerance as previously defined (Itoh et al. 2005). OsII has a clear expression peak in roots (Fig. 4) which discriminates OsII from the other modules. OsII also reaches high levels in inflorescence stage P5 and seeds stage S4, corresponding to embryo maturation 11–20 dap. OsIII shows very low levels of expression in both roots and mature leaves. The genes in this module reach their highest levels in inflorescence, where OsLTPG1 and OsLTPG3 peaks at P5, while the OsLTPG2 transcript show higher levels at P2, corresponding the meiotic stage. OsLTPG2 is also abundantly expressed in seeds at stages S4 and S5. Thus, also in rice there are one module, OsI, with a broad expression pattern in aerial parts, another module, OsII, that is expressed in, but not restricted to roots, and a third module OsIII which is expressed in reproductive tissues.

Gene ontology enrichments

The three expression modules from Arabidopsis were used in genome wide searches for coexpression, leading to greatly expanded gene networks. These networks were then

checked for enrichments in gene ontology (GO) terms. Only results from the microarray dataset *All* for each module are presented here. The 20 terms with lowest *p* value for each ontology file are found in Tables 4, 5, 6. Extended lists restricted by *p* value <0.01 are given as supplementary information (Online Resources 19–21). In the *Biological Processes* ontology, the network for module AtI is most significantly enriched in the parent term *photosynthesis* with its children terms *light harvesting*, *chlorophyll biosynthetic process*, *nonphotochemical quenching* and several other photosynthesis related terms (Table 4). The enriched GO terms also include the parent term *response to abiotic stimulus* with enriched children terms *response to radiation*, *response to light stimulus* and also *response to cold*. *Cuticle development*, *wax biosynthesis* and *very long-chain fatty acid metabolism* are other enriched terms. The most significantly enriched term of the *Molecular Function* ontology is *chlorophyll binding*. The enriched terms in the *Cellular Component* ontology are mostly related to chloroplasts, such as *thylakoid*, but *apoplast* and *cell wall* are also represented.

The network based on module AtII is in the *Biological Processes* ontology most significantly enriched in the terms *cell wall organization or biogenesis*, *secondary metabolic process* and *response to chemical stimulus*. The parent term *secondary metabolic process* is followed by enriched children terms *phenylpropanoid metabolic process*, *phenylpropanoid biosynthetic process* and *suberin biosynthetic process* (Table 5). The ancestor term *root system development* with children terms *root development* and *root morphogenesis* are also significantly enriched in the AtII-network. In the *Molecular Function* ontology, some of the most significantly enriched terms in the AtII network are *oxidoreductase activity*, *heme binding*, *peroxidase activity* and *tetrapyrrole binding*. In the *Cellular Component* ontology the terms *cell wall*, *external encapsulating structure* and *extracellular region* are enriched in the AtII-network. In module AtIII, some of the most significantly enriched *Biological Process* terms are *pollen wall assembly*, *pollen exine formation* and *sporopollenin biosynthetic process* (Table 6). In the *Molecular Function* ontology the enriched terms include *hydrolase activity*, *hydrolyzing O-glycosyl compounds*, *lipase activity* and *nutrient reservoir activity*. To summarize, the GO-ontology enrichments give indications that the AtI-module could be involved in cuticle development, AtII in suberin biosynthesis and AtIII-in pollen exine formation.

Overrepresented promoter motifs

The occurrence of overrepresented motifs in the promoter regions of the Arabidopsis expression modules were examined in order to get further clues about the factors

involved in the transcriptional regulation. In module AtI, three of the found motifs are involved in light-regulated gene expression, four in ABA-signaling and stress responses and three related to different developmental stages (Table 7). The occurrence of motifs involved in light-regulation and leaf development fits well with the significant enrichment of many photosynthesis related GO-terms in the AtI-network. The finding of motifs related to ABA-signaling is not surprising either since there were also significant enrichment for several abiotic stress related terms, such as *response to radiation* and *response to cold*. However, there was no direct evidence for ABA-regulated expression of AtI in the microarray datasets. For module AtII there are two overrepresented motifs related to light-regulated gene expression, three involved in other stresses, one in ABA response and three related to different developmental stages. In addition to these, there is one overrepresented cis-element that is related to transcription of phenylpropanoid biosynthetic genes. This motif is particularly interesting since there was a significant enrichment of the GO-terms *phenylpropanoid metabolic process*, *phenylpropanoid biosynthetic process* and *suberin biosynthetic process* in the AtII-network. The results for module AtIII includes four stress related motifs, two related to developmental stages and one involved in light regulated gene expression. Further, there are one motif connected to regulation of histone genes and two CIRCADIAN CLOCK-ASSOCIATED 1 (CCA1) binding motifs. The CCA1 binding motifs are present in the promoters of many day-phased genes (Wang et al. 1997; Michael and McClung 2003). The occurrence of CCA1 binding motifs in AtIII-promoters suggests that the expression of these genes may have a circadian regulation. Interestingly, the promoters of all three expression modules are enriched for RAV1-A binding site motifs. RAV1 is a transcription factor that is considered to be a positive regulator of leaf senescence in Arabidopsis (Woo et al. 2010). The finding of RAV1-A binding site motifs in the LTPG promoters suggests that Arabidopsis LTPG may play a role in leaf senescence. The LTPGs could have an important role in remobilization of break-down products from lipid-containing cell components. As a part of the degradative process in leaf senescence, hydrolytic enzymes such as proteases are induced. Previously, it has been shown that some nsLTPs have a proteolytic activity. It is possible that this protease activity of the nsLTPs may be involved in leaf senescence.

Alternative splicing in Arabidopsis

When the RNA sequences were aligned to genomic sequences, it was revealed that 28 out of 31 expressed Arabidopsis LTPGs possess one or more introns. The in silico analysis of the transcripts further showed that some of these

intron-containing genes have several transcript forms. The differences between the various transcripts were found to be associated with the presence or absence of introns. Actually, the in silico analysis indicated that nine of the Arabidopsis genes are alternatively spliced (Table 1). When performing a similar in silico analysis of the rice transcriptome we found that at least six of the rice LTPG genes are undergoing alternative splicing (Table 2). The alternative splicing results in that five of the genes in Arabidopsis and four of the genes in rice have one transcript form encoding the GPI-anchor signal and another transcript form lacking the signal. To confirm or reject, the presence of alternative splicing in *planta*, the transcripts from AtLTPG1, AtLTPG8, AtLTPG11 and AtLTPG29 were amplified and analyzed. At least two primer combinations were used for each gene (Fig. 5). None of the primer combinations resulted in any amplicons for the negative control, where the reverse transcriptase had been omitted from the cDNA-synthesis step. Thus, there was no contamination of genomic DNA in the RNA-samples (Online Resource 22).

For AtLTPG1 the in silico analysis indicated two different isoforms, one with the intron removed and one with the intron retained. During growth in long day conditions only the isoform with the intron removed was found in leaf and root, while both isoforms were found in flower and none in silique (Fig. 6). In plants grown under constant light, both AtLTPG1-transcript forms were detected in flower and leaf. In siliques, only the AtLTPG1-isoform without intron was detectable (Fig. 6). Both isoforms of AtLTPG1 transcripts could be confirmed by sequencing of PCR products extracted from gels. In the isoform with the retained intron there is an in-frame stop codon upstream of the GPI-anchor signal. Due to this stop codon, proteins translated from this isoform would lack the GPI-anchor signal.

According to the in silico analysis of AtLTPG8-transcripts there are two isoforms present, one with both introns removed and one with intron 1 removed but intron 2 retained. We investigated the splicing patterns of both introns in this gene with three different primer combinations, At8.1, At8.2 and At8.3 (Fig. 5). During long day conditions there was no detectable expression of AtLTPG8 in leaves. In flower there were three isoforms present, one with both introns removed, one with both introns retained and one with only intron 2 retained (Fig. 6). In root, we detected the isoform with both introns retained, as well as the isoform with intron 1 retained. In silique, the isoform with both introns removed as well as the isoform with both introns retained were identified (Fig. 6). In plants grown under constant light, expression of AtLTPG8 was only detected in flower. The three isoforms that were found in long day conditions were also seen in the samples from constant light (Fig. 6). All three isoforms found was confirmed by sequencing of gel extracted PCR products. Both isoforms with retained introns would after translation yield

Table 4 Enriched gene ontology terms for module AtI

Description	<i>p</i> value	Adjusted <i>p</i> value	Number of genes in annotation file	Number of genes in network	Occurrence of term in annotation file	Occurrence of term in network
<i>Biological process</i>						
Photosynthesis	7.52E−30	2.66E−27	22,304	125	113	23
Photosynthesis, light reaction	5.27E−14	9.32E−12	22,304	125	63	11
Response to light stimulus	9.72E−12	1.15E−09	22,304	125	455	19
Response to radiation	1.77E−11	1.56E−09	22,304	125	471	19
Generation of precursor metabolites and energy	9.78E−11	6.93E−09	22,304	125	199	13
Cuticle development	2.99E−08	1.76E−06	22,304	125	17	5
Fatty acid metabolic process	4.47E−08	2.26E−06	22,304	125	171	10
Lipid transport	8.00E−08	3.54E−06	22,304	125	137	9
Photosynthesis, light harvesting	9.66E−08	3.80E−06	22,304	125	21	5
Response to abiotic stimulus	1.25E−07	4.03E−06	22,304	125	1168	23
Very long-chain fatty acid metabolic process	1.24E−07	4.03E−06	22,304	125	22	5
Lipid localization	2.06E−07	6.08E−06	22,304	125	153	9
Chlorophyll biosynthetic process	3.05E−07	8.32E−06	22,304	125	26	5
Response to cold	1.08E−06	2.72E−05	22,304	125	241	10
Wax biosynthetic process	1.22E−06	2.89E−05	22,304	125	15	4
Wax metabolic process	1.62E−06	3.59E−05	22,304	125	16	4
Fatty acid biosynthetic process	2.12E−06	4.41E−05	22,304	125	105	7
Chlorophyll metabolic process	2.52E−06	4.96E−05	22,304	125	39	5
Nonphotochemical quenching	3.39E−06	6.33E−05	22,304	125	6	3
Pigment biosynthetic process	4.36E−06	7.71E−05	22,304	125	76	6
<i>Molecular function</i>						
Chlorophyll binding	2.09E−11	3.77E−09	24,443	128	23	7
Transferase activity, transferring acyl groups other than amino-acyl groups	2.99E−08	2.70E−06	24,443	128	175	10
Transferase activity, transferring acyl groups	1.10E−07	6.63E−06	24,443	128	201	10
Carboxylesterase activity	2.00E−06	8.99E−05	24,443	128	343	11
Acyltransferase activity	2.74E−06	9.85E−05	24,443	128	166	8
Tetrapyrrole binding	6.32E−06	1.90E−04	24,443	128	314	10
Glyceraldehyde-3-phosphate dehydrogenase activity	7.71E−06	1.98E−04	24,443	128	8	3
Oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	3.31E−05	7.45E−04	24,443	128	35	4
Catalytic activity	5.11E−05	1.02E−03	24,443	128	7,553	61
Oxidoreductase activity, acting on the aldehyde or oxo group of donors	1.99E−04	3.58E−03	24,443	128	55	4
Glycine dehydrogenase (decarboxylating) activity	2.69E−04	4.04E−03	24,443	128	5	2
Oxidoreductase activity, acting on the CH-NH2 group of donors, disulfide as acceptor	2.69E−04	4.04E−03	24,443	128	5	2
Oxidoreductase activity	5.63E−04	7.80E−03	24,443	128	1326	17
Phosphoribulokinase activity	5.24E−03	2.81E−02	24,443	128	1	1
ADP binding	5.24E−03	2.81E−02	24,443	128	1	1
Lycopene epsilon cyclase activity	5.24E−03	2.81E−02	24,443	128	1	1
Lipid binding	2.20E−03	2.81E−02	24,443	128	173	5
Plastoquinol-plastocyanin reductase activity	5.24E−03	2.81E−02	24,443	128	1	1
3-chloroallyl aldehyde dehydrogenase activity	2.38E−03	2.81E−02	24,443	128	14	2
Hydrolase activity, acting on ester bonds	2.85E−03	2.81E−02	24,443	128	904	12
<i>Cellular component</i>						
Thylakoid	3.71E−32	1.23E−30	19,822	115	322	33

Table 4 continued

Description	<i>p</i> value	Adjusted <i>p</i> value	Number of genes in annotation file	Number of genes in network	Occurrence of term in annotation file	Occurrence of term in network
Plastid thylakoid membrane	5.26E–32	1.23E–30	19,822	115	211	29
Chloroplast thylakoid membrane	5.26E–32	1.23E–30	19,822	115	211	29
Photosynthetic membrane	1.38E–32	1.23E–30	19,822	115	227	30
Thylakoid part	6.14E–32	1.23E–30	19,822	115	266	31
Thylakoid membrane	3.19E–31	5.31E–30	19,822	115	224	29
Chloroplast thylakoid	4.48E–31	5.60E–30	19,822	115	254	30
Plastid thylakoid	4.48E–31	5.60E–30	19,822	115	254	30
Organelle subcompartment	5.70E–31	6.34E–30	19,822	115	256	30
Photosystem	7.24E–29	7.24E–28	19,822	115	38	17
Chloroplast part	1.47E–24	1.34E–23	19,822	115	755	37
Plastid part	5.01E–24	4.18E–23	19,822	115	782	37
Photosystem I	2.05E–22	1.58E–21	19,822	115	15	11
Plastoglobule	7.72E–20	5.51E–19	19,822	115	55	14
Photosystem I reaction center	1.46E–15	9.76E–15	19,822	115	8	7
Cell	9.63E–15	5.67E–14	19,822	115	11,708	105
Cell part	9.63E–15	5.67E–14	19,822	115	11,708	105
Chloroplast	1.73E–14	9.62E–14	19,822	115	20,70	43
Plastid	5.48E–14	2.88E–13	19,822	115	2,139	43
Plastid envelope	6.66E–14	3.33E–13	19,822	115	382	20

The 20 terms with lowest *p* value in the domains “Biological Process”, “Molecular Function” and “Cellular Component” are presented here, complete table is found as Internet Resource 1

proteins without the GPI-anchor, due to in-frame stop codons in the introns.

In silico analysis of AtLTPG11 transcripts revealed the similar isoforms as in AtLTPG8; one with both introns removed and one with intron 1 removed but intron 2 retained. To investigate the splicing patterns of both introns three primer combinations were used; At11:1, At11:2 and At11:3 (Fig. 5). During long-day conditions AtLTPG11 was found to be expressed in flower and root, but not in leaf or silique (Fig. 6). The amplified At11.1 fragment was slightly larger than expected (149 bp) if both introns would have been removed, but smaller than expected for a fragment with a retained intron 1. In plants grown under constant light two products were detected with At11.1. One that corresponded well to a fragment with both introns removed (122 bp), whereas the other was similar to the 149 bp fragment detected during long day conditions. Sequencing of the larger 11.1 fragment revealed a partial tandem duplication of 27 bases in exon 2. Further investigations are needed to reveal if this is an artifact or an actual modification of the mRNA.

In AtLTPG29 the *in silico* analysis predicted an exon skipping event. In this case exon 3, containing a stop codon, is skipped and the alternative exon 4 is reached. Only the isoform without exon 3 contains the GPI-anchor signal, due to a stop codon in exon 3 leading to much

shorter polypeptide. Two primer combinations, At29:1 and At29:2, were used to investigate the splicing pattern *in planta*. Expression of AtLTPG29 was detected in flowers during long-day conditions and in flowers and siliques during constant light (Fig. 6). In all cases where expression was detected both isoforms of AtLTPG29-transcripts were found. Thus, also for AtLTPG29 there are transcripts encoding the GPI-anchor attachment signal, but also transcripts lacking the in frame GPI-anchor signal.

To conclude, it was verified *in planta* that there are alternative splicing of several LTPG transcripts in Arabidopsis. The occurrence of alternative splicing in the LTPG genes varies between different tissues and we could also note that the splicing patterns sometimes differ between plants grown under long-day and in constant light. Interestingly, there are in three of these four tested genes one transcript form that encode a protein with the GPI attachment signal and another transcript that should not yield a protein with a GPI anchor. Thus, it seems that alternative splicing could play a role in regulating the cellular localization of LTPGs.

Phylogeny of Arabidopsis and rice LTPGs

A phylogenetic analysis of the Arabidopsis and rice sequences are shown in Fig. 7. Members of the same

Table 5 Enriched gene ontology terms for module AtII

Description	<i>p</i> value	Adjusted <i>p</i> value	Number of genes in annotation file	Number of genes in network	Occurrence of term in annotation file	Occurrence of term in network
<i>Biological process</i>						
Response to stimulus	2.86E−13	3.02E−10	22,304	1,042	3,207	235
Secondary metabolic process	2.62E−12	1.38E−09	22,304	1,042	330	48
Response to chemical stimulus	6.46E−11	2.28E−08	22,304	1,042	1710	139
Cell wall organization or biogenesis	4.34E−10	9.91E−08	22,304	1,042	260	38
Phenylpropanoid metabolic process	4.68E−10	9.91E−08	22,304	1,042	133	26
Response to oxidative stress	1.35E−09	2.37E−07	22,304	1,042	247	36
Cellular amino acid derivative metabolic process	2.98E−09	4.50E−07	22,304	1,042	231	34
Response to stress	3.99E−09	5.27E−07	22,304	1,042	1,853	141
Phenylpropanoid biosynthetic process	1.19E−08	1.40E−06	22,304	1,042	104	21
Cell wall organization	5.21E−08	5.51E−06	22,304	1,042	165	26
Root development	1.16E−07	8.77E−06	22,304	1,042	230	31
Root system development	1.16E−07	8.77E−06	22,304	1,042	230	31
Toxin catabolic process	1.13E−07	8.77E−06	22,304	1,042	46	13
Toxin metabolic process	1.13E−07	8.77E−06	22,304	1,042	46	13
Suberin biosynthetic process	2.21E−07	1.56E−05	22,304	1,042	5	5
Intracellular signaling pathway	3.83E−07	2.53E−05	22,304	1,042	182	26
Cellular amino acid derivative biosynthetic process	4.09E−07	2.55E−05	22,304	1,042	171	25
Cellular aromatic compound metabolic process	1.32E−06	7.78E−05	22,304	1,042	296	34
Aromatic compound biosynthetic process	2.73E−06	1.52E−04	22,304	1,042	177	24
Root morphogenesis	3.13E−06	1.65E−04	22,304	1,042	99	17
<i>Molecular function</i>						
Oxidoreductase activity	7.26E−27	4.14E−24	24,443	1,173	1,326	158
Catalytic activity	9.05E−26	2.58E−23	24,443	1,173	7,553	529
Heme binding	2.23E−24	4.25E−22	24,443	1,173	290	63
Iron ion binding	3.96E−23	5.43E−21	24,443	1,173	352	68
Peroxidase activity	5.71E−23	5.43E−21	24,443	1,173	103	37
Oxidoreductase activity, acting on peroxide as acceptor	5.71E−23	5.43E−21	24,443	1,173	103	37
Tetrapyrrole binding	2.01E−22	1.64E−20	24,443	1,173	314	63
Electron carrier activity	3.37E−20	2.40E−18	24,443	1,173	427	71
Antioxidant activity	5.98E−20	3.80E−18	24,443	1,173	130	38
Metal ion binding	1.61E−17	8.80E−16	24,443	1,173	2,020	183
Cation binding	1.85E−17	8.80E−16	24,443	1,173	2,147	191
Ion binding	1.85E−17	8.80E−16	24,443	1,173	2,147	191
Transition metal ion binding	2.64E−15	1.16E−13	24,443	1,173	1,691	155
Structural constituent of cell wall	3.55E−12	1.45E−10	24,443	1,173	37	16
Oxygen binding	2.37E−09	9.01E−08	24,443	1,173	223	34
Transferase activity	8.09E−09	2.89E−07	24,443	1,173	2,429	177
Transferase activity, transferring glycosyl groups	1.91E−08	6.42E−07	24,443	1,173	416	48
Monooxygenase activity	5.07E−08	1.61E−06	24,443	1,173	300	38
Glutathione transferase activity	2.69E−07	8.08E−06	24,443	1,173	48	13
Oxidoreductase activity, acting on NADH or NADPH, with oxygen as acceptor	8.94E−07	2.55E−05	24,443	1,173	9	6
<i>Cellular component</i>						
Endomembrane system	1.87E−20	2.71E−18	19,822	919	2,482	214
Cell part	9.74E−13	4.71E−11	19,822	919	11,708	644

Table 5 continued

Description	<i>p</i> value	Adjusted <i>p</i> value	Number of genes in annotation file	Number of genes in network	Occurrence of term in annotation file	Occurrence of term in network
Cell	9.74E−13	4.71E−11	19,822	919	11,708	644
Intrinsic to membrane	1.41E−08	5.13E−07	19,822	919	714	68
Cell wall	1.79E−06	4.34E−05	19,822	919	458	45
Membrane	1.59E−06	4.34E−05	19,822	919	3,727	229
External encapsulating structure	2.27E−06	4.69E−05	19,822	919	462	45
Extracellular region	8.28E−05	1.50E−03	19,822	919	393	36
Anchored to membrane	9.31E−05	1.50E−03	19,822	919	193	22
Membrane part	1.11E−04	1.61E−03	19,822	919	1,098	78
Integral to membrane	9.08E−04	1.20E−02	19,822	919	397	33

The 20 terms with lowest *p* value in the domains “Biological Process”, “Molecular Function” and “Cellular Component” are presented here, complete table is found as Internet Resource 2

modules are distributed all over the phylogenetic tree. However, within all modules there are some putative paralogs that are clustered together, such as OsLTPG26 and OsLTPG27 and AtLTPG4 and AtLTPG23, suggesting that the modules have expanded through duplications after the separation of monocotyledons and dicotyledons. A more striking finding is that some rice LTPGs and Arabidopsis LTPGs from equivalent modules are putative orthologs found on the same branch of the tree. This is shown for AtI and OsI, such as between AtLTPG1 and OsLTPG22, between AtLTPG6 and OsLTPG10 and between AtLTPG2 and OsLTPG12 and also for AtII and OsII in the case of the cluster OsLTPG8, OsLTPG17, AtLTPG16 and AtLTPG20. The phylogenetic tree therefore indicates that the gene expression patterns, manifested in the expression modules, were established before the separation of rice and Arabidopsis. The genes that undergo alternative splicing are not located to specific branches of the phylogenetic tree (Fig. 7). However, the alternative splicing of the conserved rice and Arabidopsis genes OsLTPG22 and AtLTPG1 shows that there is at least one example where the evolution of the alternatively splicing event possibly pre-dates the separation of monocots and dicots.

Discussion

The aim with this study was to find groups of LTPG genes that are involved in related biological processes. We reasoned that the identification of such functional groups is important for further systematic investigations into the biological roles of this enigmatic family of proteins. Here, we have identified coexpressed LTPG genes in both rice and Arabidopsis. Among the coexpressed genes we could identify three different expression profiles. The coexpressed genes were therefore placed into three separate groups or modules. The

Arabidopsis module AtI is built from the three genes AtLTPG1, AtLTPG2 and AtLTPG6. The GO analysis of the AtI expression network resulted in many significantly enriched terms related to photosynthesis. Further, the search for regulatory elements identified that three promoter motifs, GATA, Ibox and SORLREP3, associated with light-regulated gene expression (Hudson and Quail 2003; Reyes et al. 2004) are overrepresented in the promoters of the AtI-module genes. Light is one of the factors that have been demonstrated to increase the wax deposition, as revealed from comparisons of light- and dark grown plants (reviewed in Shepherd and Wynne Griffiths 2006). The light regulated expression and the coexpression with photosynthesis genes therefore support that the genes in the AtI module have their main function in the deposition and biosynthesis of the cuticular waxes or cutin.

Our results are further supported by functional reports of the genes in module AtI (DeBono et al. 2009; Lee et al. 2009; Kim et al. 2012). Decreased AtLTPG1 expression in Arabidopsis resulted in that less wax was loaded on the stem surface (DeBono et al. 2009). However, when AtLTPG1 was disrupted in another study there were no significant alterations found for the wax load (Lee et al. 2009). Rather, Lee et al. demonstrated a 10 % reduction of the C29 alkane (nonacosane) which is the major component of cuticular waxes in the stems and siliques. Although, the data from these studies show some contradictions, the results indicate that AtLTPG1 is involved in cuticular lipid accumulation. More recently, it was shown that AtLTPG2 is functionally redundant or overlapping with AtLTPG1 since the wax load in stems and siliques was reduced with about 10 % also in an AtLTPG2 insertion mutant (Kim et al. 2012). Our data suggest that AtLTPG6 is functionally overlapping with AtLTPG1 and AtLTPG2. Possibly, the wax load and C29 alkane-levels would be further reduced in a triple mutant knocked out for AtLTPG1, AtLTPG2 and AtLTPG6.

Table 6 Enriched gene ontology terms for module AtIII

Description	<i>p</i> value	Adjusted <i>p</i> value	Number of genes in annotation file	Number of genes in network	Occurrence of term in annotation file	Occurrence of term in network
<i>Biological process</i>						
Pollen wall assembly	5.89E–20	1.22E–17	22,304	216	20	12
Cellular component assembly involved in morphogenesis	5.89E–20	1.22E–17	22,304	216	20	12
External encapsulating structure organization	1.33E–17	1.83E–15	22,304	216	28	12
Pollen exine formation	4.48E–17	4.63E–15	22,304	216	16	10
Pollen development	6.32E–15	5.22E–13	22,304	216	149	18
Sexual reproduction	8.66E–13	5.96E–11	22,304	216	63	12
Gametophyte development	2.94E–12	1.73E–10	22,304	216	212	18
Anatomical structure formation involved in morphogenesis	2.28E–11	1.18E–09	22,304	216	103	13
Lipid localization	2.59E–11	1.19E–09	22,304	216	153	15
Cellular component morphogenesis	7.27E–10	3.00E–08	22,304	216	260	17
Lipid storage	5.68E–09	2.13E–07	22,304	216	16	6
Reproduction	2.96E–08	1.02E–06	22,304	216	931	29
Cellular developmental process	5.35E–08	1.70E–06	22,304	216	435	19
Sporopollenin biosynthetic process	8.96E–07	2.64E–05	22,304	216	3	3
Localization	1.09E–06	3.00E–05	22,304	216	1566	36
Anatomical structure morphogenesis	3.02E–06	7.80E–05	22,304	216	515	18
Multicellular organismal development	3.88E–06	9.42E–05	22,304	216	1655	36
Multicellular organismal process	4.22E–06	9.68E–05	22,304	216	1732	37
Developmental process	5.22E–06	1.14E–04	22,304	216	1820	38
Lipid transport	7.78E–06	1.61E–04	22,304	216	137	9
<i>Molecular Function</i>						
Lipid binding	1.91E–09	4.93E–07	24,443	240	173	14
Carboxylesterase activity	3.24E–07	4.18E–05	24,443	240	343	16
Hydrolase activity, acting on glycosyl bonds	5.43E–06	2.96E–04	24,443	240	426	16
Hydrolase activity	3.50E–06	2.96E–04	24,443	240	2632	50
Lipase activity	5.73E–06	2.96E–04	24,443	240	130	9
Hydrolase activity, hydrolyzing O-glycosyl compounds	9.40E–06	4.04E–04	24,443	240	395	15
GDP-dissociation inhibitor activity	3.18E–05	9.15E–04	24,443	240	7	3
Hydrolase activity, acting on ester bonds	3.19E–05	9.15E–04	24,443	240	904	23
Nutrient reservoir activity	2.57E–05	9.15E–04	24,443	240	36	5
Pectinesterase activity	1.01E–04	2.61E–03	24,443	240	146	8
Rho GDP-dissociation inhibitor activity	2.86E–04	6.71E–03	24,443	240	3	2
Enzyme regulator activity	4.27E–04	9.18E–03	24,443	240	328	11
Beta-galactosidase activity	6.28E–04	1.24E–02	24,443	240	40	4
Polygalacturonase activity	6.73E–04	1.24E–02	24,443	240	71	5
Galactosidase activity	7.57E–04	1.30E–02	24,443	240	42	4
Urea transmembrane transporter activity	1.40E–03	1.97E–02	24,443	240	6	2
Amide transmembrane transporter activity	1.40E–03	1.97E–02	24,443	240	6	2
Active transmembrane transporter activity	1.45E–03	1.97E–02	24,443	240	501	13
Acytransferase activity	1.30E–03	1.97E–02	24,443	240	166	7
Oxygen binding	1.69E–03	2.02E–02	24,443	240	223	8
<i>Cellular component</i>						
Endomembrane system	4.34E–09	4.12E–07	19,822	181	2482	52
Cell	8.33E–06	2.11E–04	19,822	181	11708	135
Cell part	8.33E–06	2.11E–04	19,822	181	11708	135
Intrinsic to membrane	8.87E–06	2.11E–04	19,822	181	714	20
Extracellular region	6.88E–05	1.31E–03	19,822	181	393	13

Table 6 continued

Description	<i>p</i> value	Adjusted <i>p</i> value	Number of genes in annotation file	Number of genes in network	Occurrence of term in annotation file	Occurrence of term in network
Beta-galactosidase complex	1.53E–04	2.42E–03	19,822	181	30	4
Integral to membrane	3.01E–04	4.08E–03	19,822	181	397	12
Membrane part	4.46E–04	4.71E–03	19,822	181	1098	22
Anchored to membrane	4.07E–04	4.71E–03	19,822	181	193	8
Endoplasmic reticulum lumen	2.86E–03	2.72E–02	19,822	181	9	2
Actin cytoskeleton	5.00E–03	4.31E–02	19,822	181	38	3

The 20 terms with lowest *p* value in the domains “Biological Process”, “Molecular Function” and “Cellular Component” are presented here, complete table is found as Internet Resource 3

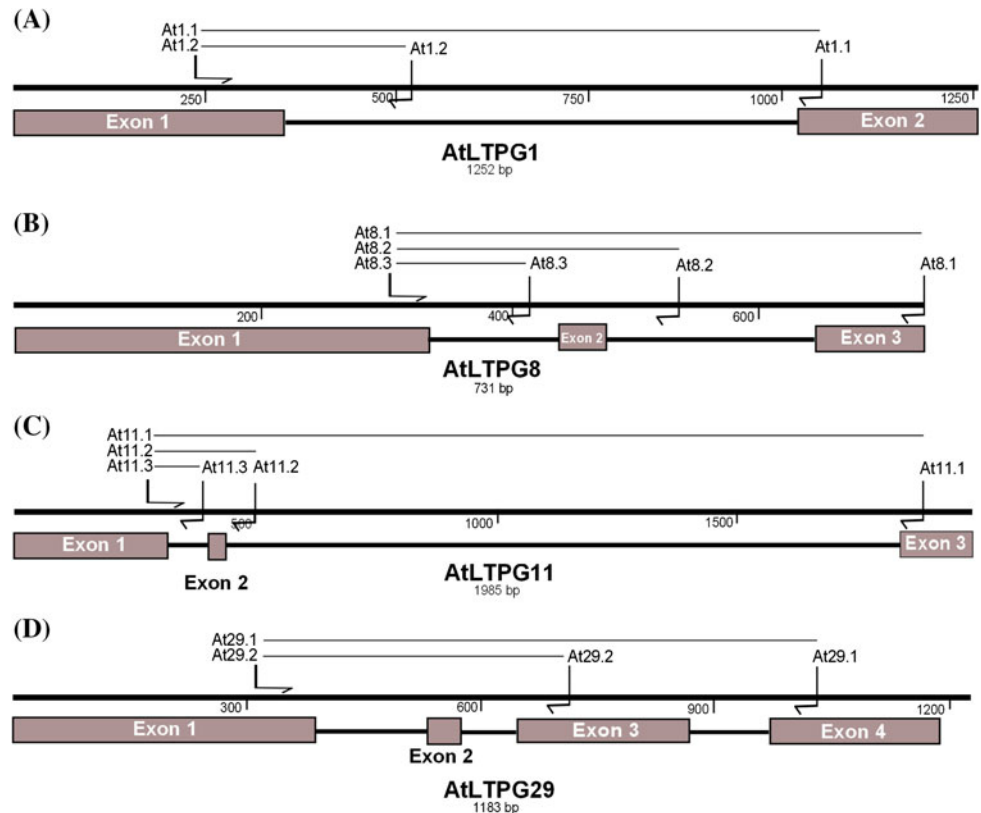
Table 7 Overrepresented promoter motifs in the genes of the LTPGs from arabidopsis

Promoter element	e-value	Description	References
<i>AtI</i>			
AtMYC2 BS in RD22	7.24e–04	Dehydration- and ABA-induced gene expression	Abe et al. (1997)
ATHB1 binding site motif	4.80e–03	Regulation of cotyledon and leaf development	Aoyama et al. (1995)
RY-repeat promoter motif	1.15e–02	Seed-specific gene regulation	Ezcurra et al. (1999)
ATHB5 binding site motif	1.17e–02	ABA signaling in seedlings	Johannesson et al. (2003)
Ibox promoter motif	5.18e–02	Light-regulated gene expression	Hiratsuka and Chua (1997)
RAV1-A binding site motif	5.66e–02	Leaf senescing, negative regulation of growth	Hu et al. (2004) and Woo et al. (2010)
SORLREP3	9.26e–02	Light-regulated gene expression	Hudson and Quail (2003)
GATA promoter motif [LRE]	1.31e–01	Light-regulated gene expression	Grob and Stüber (1987)
W-box promoter motif	2.50e–01	Defense response	Chen et al.(2002) and Maleck et al. (2000)
LTRE promoter motif	2.55e–01	Low-temperature-induced	Dunn et al. (1998)
<i>AtII</i>			
RAV1-A binding site motif	3.30e–07	Leaf senescing, negative regulation of growth	Hu et al. (2004) and Woo et al. (2010)
MYB4 binding site motif	1.13e–06	Response to environmental stresses	Chen et al. (2002)
BoxII promoter motif	6.85e–06	Light-regulated gene expression, defense response	Lam and Chua (1990), Buchel et al. (1996)
MYB binding site promoter	2.19e–04	Transcription of phenylpropanoid biosynthetic genes	Sablowski et al. (1994)
DPBF1&2 binding site motif	5.07e–04	ABA-responsive and embryo-specification elements	Kim et al. (1997)
RY-repeat promoter motif	7.04e–03	Seed-specific gene regulation	Ezcurra et al. (1999)
SORLIP1	1.36e–02	Light-regulated gene expression	Hudson and Quail (2003)
W-box promoter motif	2.41e–02	Response to wounding and fungal elicitors	Chen et al. (2002) and Maleck et al. (2000)
MYB3 binding site motif	2.43e–02	Response to environmental stresses	Chen et al. (2002)
CArG promoter motif	5.20e–02	Flower development	Tilly et al. (1998)
<i>AtIII</i>			
MYB4 binding site motif	1.25e–03	Response to environmental stresses	Chen et al. (2002)
RAV1-A binding site motif	3.93e–02	Leaf senescing, negative regulation of growth	Hu et al. (2004) and Woo et al.(2010)
Octamer promoter motif	4.22e–02	Cell-cycle-regulated expression of histone genes	Nakayama et al. (1992)
MYB3 binding site motif	5.05e–02	Response to environmental stresses	Chen et al. (2002)
CCA1 binding site motif	5.61e–02	Related to circadian clock	Alabadi et al. (2001) and Andronis et al. (2008)
CCA1 motif1 BS in CAB1	6.41e–02	Related to circadian clock	Alabadi et al. (2001) and Andronis et al. (2008)
CArG promoter motif	8.64e–02	Flower development	Tilly et al. (1998)
CBF2 binding site motif	9.82e–02	Response to abiotic stress, especially cold	Gilmour et al. (1998)and Novillo et al. (2012)
GBF1/2/3 BS in ADH1	9.82e–02	Induced by hypoxia	de Vetten and Ferl (1995)
T-box promoter motif	1.05e–01	Light-regulated gene expression	Chan et al. (2001)

Module AtII is the largest expression module with seven genes. The GO analysis of AtII revealed significant enrichment of the term biosynthesis of phenylpropanoid and its daughter term biosynthesis of suberin. Suberin

consists of an aliphatic cutin-like and an aromatic lignin-like domain (Bernards 2002) and is deposited for example in the endodermis and hypodermis of roots, the bundle sheaths of leaves, in seed coats and in the periderm of

Fig. 5 Maps showing the gene structures of AtLTPg1 (a), AtLTPg8 (b), AtLTPg11 (c) and AtLTPg29 (d). The primer combinations used for PCR analysis of alternative splicing are shown



shoots and roots. Suberin is deposited as a lamella on the inner surface of the cell wall, thus separating the cell wall from the plasma membrane (Pollard et al. 2008; Schreiber 2010). The GO analysis suggest that module AtII may be involved in suberin biosynthesis and deposition in roots. This is also supported by the expression pattern where several genes of AtII reach their highest transcript levels in roots. Furthermore, the MYB binding site motif that is significantly enriched in the AtII promoters is known to enhance the transcription of phenylpropanoid biosynthetic genes (Sablowski et al. 1994). The phenylpropanoid biosynthetic pathway provides precursors for the synthesis of suberin. These results open up for directed investigations aiming at elucidating the role of the LTPGs in suberin accumulation. So far, there are to our knowledge, not yet any experimental evidence published that link the function of nsLTPs to suberin deposition (Ranathunge et al. 2011).

Module AtIII is highly expressed in flowers and seeds and show GO enrichments that suggest a role for this module in sporopollenin biosynthesis or deposition. Sporopollenin is a major component of exine walls of pollen grains and contributes to the remarkable resistance of the pollen wall to abiotic and biotic stresses, such as dehydration, UV irradiation, and pathogen attack. The chemical composition of sporopollenin is not exactly known, due to its unusual chemical stability. Recent investigations show that sporopollenin is not a homogeneous macromolecule

but is instead made up of complex biopolymers derived mainly from saturated precursors such as long-chain fatty acids or long aliphatic chains. It has been suggested recently that nsLTPs may have a role in sporopollenin synthesis (Ariizumi and Toriyama 2011), although our study are, to our knowledge, the first to provide data pointing in such directions. Two CCA1 binding motifs are present in the AtIII-promoters, which indicate that the genes are regulated according to the circadian clock. The circadian clock is known to regulate the development of reproductive organs, the flower opening required for efficient pollination and the production of volatile compounds giving the signature scent of the plant (Yakir et al. 2007; Troncoso-Ponce and Mas 2012). One would assume that the maturation of pollen would coincide with these events and subsequently also be controlled by the circadian clock.

To summarize, we suggest that module AtI is involved in light regulated deposition or synthesis of cutin or cuticle waxes, that module AtII may have a role in the synthesis and deposition of suberin in roots and seed coats, while module AtIII could be involved in sporopollenin biosynthesis and deposition in pollen grains. The cuticular waxes, suberin and sporopollenin are all polymers built from long-chain fatty acids or long aliphatic chains. Their synthesis requires at least four steps: (1) the *de novo* synthesis of polymer precursors (2) secretion from the lipid bilayer to the apoplastic compartment (3) transfer of the precursors through the apoplastic

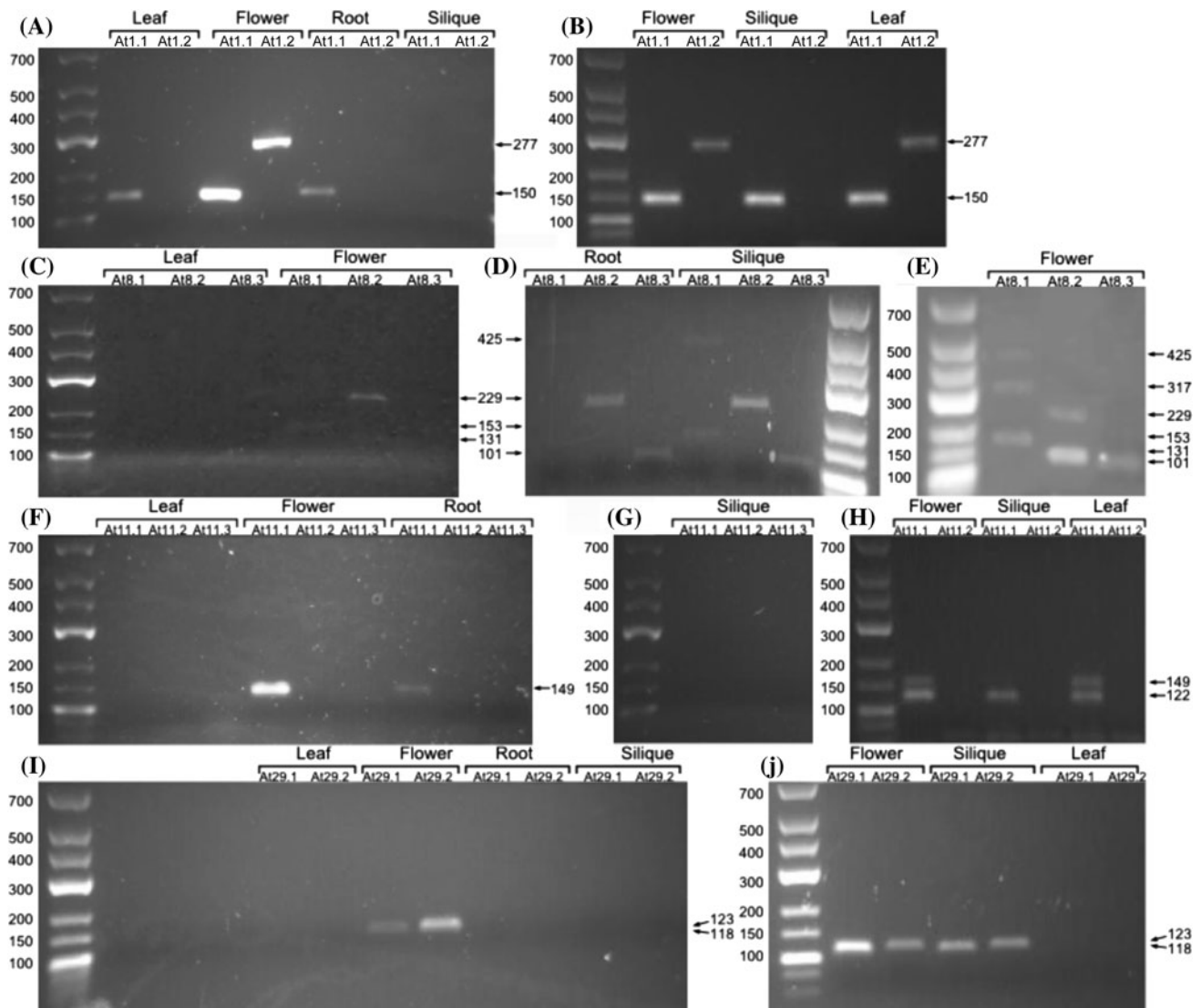


Fig. 6 Agarose gels showing PCR products from four genes that are putatively alternative spliced: AtLTPg1 (a, b), AtLTPg8 (c, d, e), AtLTPg11 (f, g, h) and AtLTPg29 (i, j). The cDNA used for PCR

were synthesized from RNA isolated from plants grown either under long day conditions (a, c, d, f, g, i) or constant light (b, e, h, j)

compartment or the cell wall and (4) polymerization (Arizumi and Toriyama 2011; Ranathunge et al. 2011). Thus in step (3) above, once the hydrophobic lipid polymer compounds are exported, they have to pass through a highly hydrophilic environment, such as the cell wall, on their way to the polymerization site. How this transport is achieved is still unknown, but it is not unlikely that the LTPGs are involved in the delivery of the polymer precursors.

We included rice in our investigation to see if our findings from Arabidopsis could be relevant also in monocots. Interestingly, we could note that in both rice and Arabidopsis there are one expression module which are predominant in aerial parts (AtI and OsI), another in roots (AtII and OsII), and a third module with an expression pattern restricted to reproductive tissues (AtIII and OsIII). Further,

in both rice and Arabidopsis the root abundant modules (AtII and OsII) contain the largest number of LTPG genes, with 7 genes in Arabidopsis and 8 genes in rice. In conclusion, according to the expression patterns, the number of members and the distribution in the phylogenetic tree, the modules found in Arabidopsis and rice appear to be functionally equivalent. The identification of equivalent expression modules in dicots and monocots indicates that the LTPG expression profiles were established before the separation of monocots and dicots. This evolutionary conservation renders further support that our approach is useful for deducing the function of LTPGs in flowering plants.

This study of the LTPGs is to our knowledge, one of the first cases suggesting alternative splicing as a potential regulator of the GPI-anchoring process in plants. However,

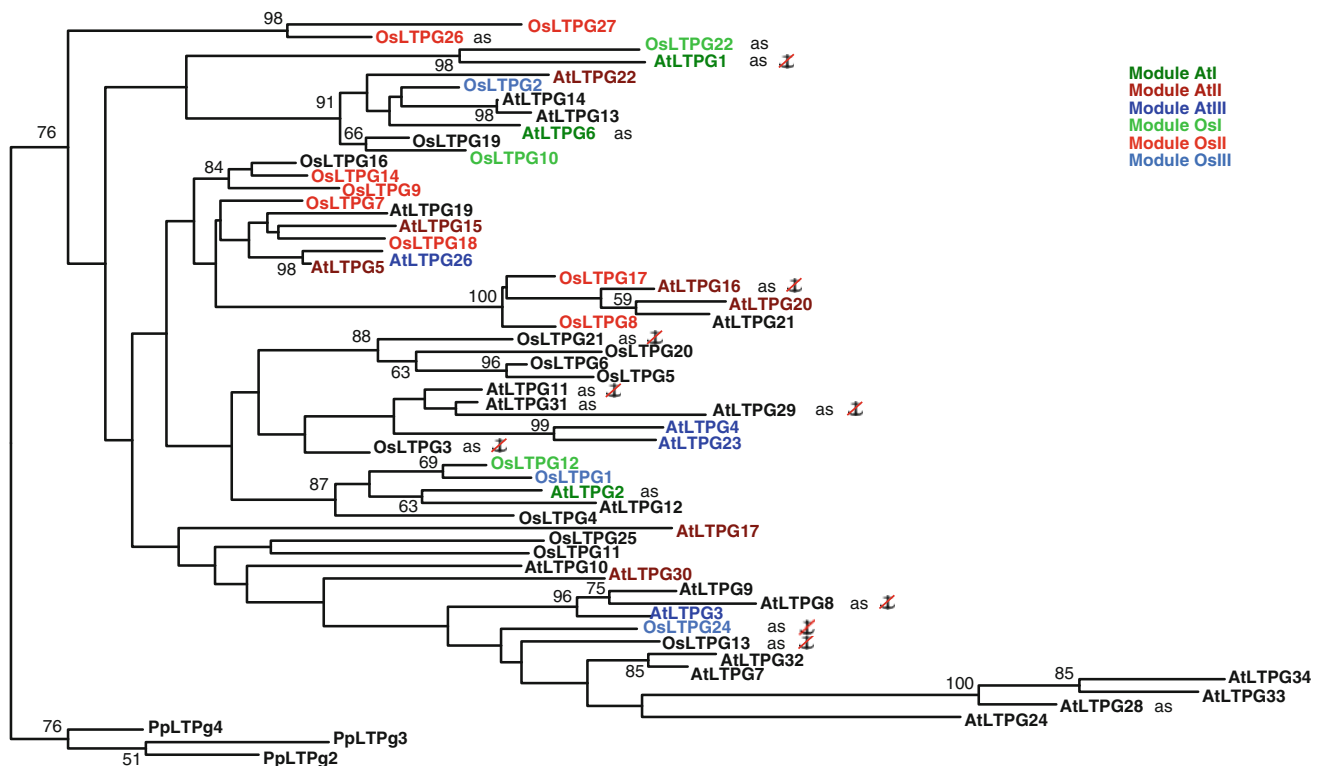


Fig. 7 A maximum likelihood phylogenetic tree of the LTPGs from Arabidopsis and rice, with LTPGs from the moss *P. patens* (PpLTPg2, PpLTPg3 and PpLTPg4) as outgroup. Pseudogenes are not present. Only bootstrap values above 50 are shown. Members of each expression module are shown as uniformly colored protein names,

while proteins not considered as members of a module are written in black. Genes with two or more transcript isoforms are indicated by the label AS behind the name. If the GPI-anchor signal is lost in one of the transcript forms, a crossed out anchor is shown

there are several similar mammalian examples where alternative splicing generates transcript isoforms with the anchoring signal and other isoforms lacking the signal (Patel et al. 2000; Grahnert et al. 2005; Kikuchi et al. 2008). For several LTPG genes the alternative splicing results in one transcript form with the GPI-anchoring signal and another form that is without the GPI-anchoring signal. This indicates that protein isoforms both with and without a GPI-anchor are produced from these genes. These isoforms may have different properties, for instance such as that the isoform without anchor is not functional, or perhaps more likely, that the isoforms have different localizations in the cell or the organism. In case of the LTPGs it is possible that the versions lacking the GPI-anchor is unattached to the plasma membrane and located to the apoplastic space, where they could be involved in the downstream transportation of lipids from GPI-anchored LTPGs to the plant surface. It seems plausible that alternative splicing has evolved as a mechanism to control the activity of at least some of the LTPGs. The observed alternative splicing further brings an evolutionary and functionally explanation to the conservation of an intron at a position between the last, most C-terminal, of the conserved cys and the

GPI-anchor signal (Edstam et al. 2011). If the alternative splicing is a regulatory mechanism it is likely that each isoform are predominant during certain conditions. Now it will be of special interest to obtain knowledge about when and where the different transcript and protein isoforms are accumulating. If we succeed in determining the localization of the LTPG isoforms we may get further important clues to the function of these proteins.

We have previously identified that the genes encoding the LTPGs likely evolved in plants soon after the colonization of land, since the genes are present in early diverging land plants, such as liverworts, but not identified in streptophyte algae (Edstam et al. 2011). The first land plants faced numerous challenges that included increased exposure to UV radiation, desiccation, and temperature stress when they adapted to a life on land approximately 470 million years ago. Sporopollenin and cuticular waxes are present in liverworts and mosses as well as in highly diverged plants like Arabidopsis (Neinhuis and Jetter 1995; Cook and Graham 1998). We speculate that the LTPGs may have been selected for during land plant evolution due to the fact that their gene products are involved in the defense against radiation and desiccation.

Acknowledgments The authors are grateful for the assistance from Muneeswaran Jayachandra Pandiyan during the initial phase of the study. This work was supported by Carl Tryggers Stiftelse.

References

- Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105
- Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, Shinozaki K (1997) Role of *Arabidopsis* MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell* 9:1859–1868
- Ahmad SR, Lidington EA, Ohta R, Okada N, Robson MG, Davies KA, Leitges M, Harris CL, Haskard DO, Mason JC (2003) Decay-accelerating factor induction by tumour necrosis factor- α through a phosphatidylinositol-3 kinase and protein kinase C-dependent pathway protects murine vascular endothelial cells against complement deposition. *Immunology* 110:258–268
- Alabadi D, Oyama T, Yanovsky MJ, Harmon FG, Mas P, Kay SA (2001) Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. *Science* 293:880–883
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Andronis C, Barak S, Knowles SM, Sugano S, Tobin EM (2008) The clock protein CCA1 and the bZIP transcription factor HY5 physically interact to regulate gene expression in *Arabidopsis*. *Mol Plant* 1:58–67
- Aoyama T, Dong CH, Wu Y, Carabelli M, Sessa G, Ruberti I, Morelli G, Chua NH (1995) Ectopic expression of the *Arabidopsis* transcriptional activator Athb-1 alters leaf cell fate in tobacco. *Plant Cell* 7:1773–1785
- Ariizumi T, Toriyama K (2011) Genetic regulation of sporopollenin synthesis and pollen exine development. *Ann Rev Plant Biol* 62:437–460
- Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G, Moseyko N, Yoo D, Xu I, Zoeckler B, Montoya M, Miller N, Weems D, Rhee SY (2004) Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiol* 135:745–755
- Bernards MA (2002) Demystifying suberin. *Can J Plant Sci* 80: 227–240
- Borner GH, Lilley KS, Stevens TJ, Dupree P (2003) Identification of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A proteomic and genomic analysis. *Plant Physiol* 132:568–577
- Boutrot F, Chantret N, Gautier MF (2008) Genome-wide analysis of the rice and *Arabidopsis* non-specific lipid transfer protein (nsLtp) gene families and identification of wheat nsLtp genes by EST data mining. *BMC Genomics* 9:86
- Buchel AS, Molenkamp R, Bol JF, Linthorst HJ (1996) The PR-1a promoter contains a number of elements that bind GT-1-like nuclear factors with different affinity. *Plant Mol Biol* 30:493–504
- Cameron KD, Teece MA, Smart LB (2006) Increased accumulation of cuticular wax and expression of lipid transfer protein in response to periodic drying events in leaves of tree tobacco. *Plant Phys* 140:176–183
- Chan CS, Guo L, Shih MC (2001) Promoter analysis of the nuclear gene encoding the chloroplast glyceraldehyde-3-phosphate dehydrogenase B subunit of *Arabidopsis* thaliana. *Plant Mol Biol* 46:131–141
- Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan S, Zou G, Whitham SA, Budworth PR, Tao Y, Xie Z, Chen X, Lam S, Kreps JA, Harper JF, Si-Ammour A, Mauch-Mani B, Heinlein M, Kobayashi K, Hohn T, Dangl JL, Wang X, Zhu T (2002) Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell* 14:559–574
- Cook ME, Graham LE (1998) Structural similarities between surface layers of selected Charophycean Algae and Bryophytes and the cuticles of vascular Plants. *Int J Plant Sci* 159:780–787
- De Bodt S, Proost S, Vandepoele K, Rouze P, Van de Peer Y (2009) Predicting protein–protein interactions in *Arabidopsis* thaliana through integration of orthology, gene ontology and co-expression. *BMC Genomics* 10:288
- de Vetten NC, Ferl RJ (1995) Characterization of a maize G-box binding factor that is induced by hypoxia. *Plant J* 7:589–601
- DeBono A, Yeats TH, Rose JK, Bird D, Jetter R, Kunst L, Samuels L (2009) *Arabidopsis* LTPG is a glycosylphosphatidylinositol-anchored lipid transfer protein required for export of lipids to the plant surface. *Plant Cell* 21:1230–1238
- Do JH, Choi DK (2008) Clustering approaches to identifying gene expression patterns from DNA microarray data. *Mol Cells* 25: 279–288
- Domany E (2003) Cluster analysis of gene expression data. *J Stat Phys* 110:1117–1139
- Dunn MA, White AJ, Vural S, Hughes MA (1998) Identification of promoter elements in a low-temperature-responsive gene (blt49) from barley (*Hordeum vulgare* L). *Plant Mol Biol* 38:551–564
- Edqvist J, Farbos I (2002) Characterization of germination-specific lipid transfer proteins from *Euphorbia lagascae*. *Planta* 215:41–50
- Edstam MM, Viitanen L, Salminen TA, Edqvist J (2011) Evolutionary history of the non-specific lipid transfer proteins. *Mol Plant* 4:947–964
- Eklund DM, Edqvist J (2003) Localization of nonspecific lipid transfer proteins correlate with programmed cell death responses during endosperm degradation in *Euphorbia lagascae* seedlings. *Plant Physiol* 132:1249–1259
- Elortza F, Nühse TS, Foster LJ, Stensballe A, Peck SC, Jensen ON (2003) Proteomic analysis of glycosylphosphatidylinositol-anchored membrane proteins. *Mol Cell Proteomics* 2:1261–1270
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2:953–971
- Ezcurra I, Ellerstrom M, Wycliffe P, Stalberg K, Rask L (1999) Interaction between composite elements in the napA promoter: both the B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. *Plant Mol Biol* 40:699–709
- Futschik ME, Kasabov NK (2002) Fuzzy clustering of gene expression data. In: proceedings of world congress of computational intelligence WCCI 2002, Hawaii, IEEE Press
- Ghiran I, Klickstein LB, Nicholson-Weller A (2003) Calreticulin is at the surface of circulating neutrophils and uses CD59 as an adaptor molecule. *J Biol Chem* 278:21024–21031
- Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF (1998) Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. *Plant J* 16:433–442
- Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li W, Ogawa M, Yamauchi Y, Preston J, Aoki K, Kiba T, Takatsuto S, Fujioka S, Asami T, Nakano T, Kato H, Mizuno T, Sakakibara H, Yamaguchi S, Nambara E, Kamiya Y, Takahashi H, Hirai MY, Sakurai T, Shinozaki K, Saito K, Yoshida S, Shimada Y (2008) The AtGenExpress hormone and chemical treatment data set: experimental design data evaluation, model data analysis and data access. *Plant J* 55:526–542
- Grahner M, Friedrich M, Engeland K, Hauschildt S (2005) Analysis of mono-ADP-ribosyltransferase 4 gene expression in human monocytes: splicing pattern and potential regulatory elements. *Biochim Biophys Acta* 1730:173–186

- Grob U, Stüber K (1987) Discrimination of phytochrome dependent light inducible from non-light inducible plant genes. Prediction of a common light-responsive element (LRE) in phytochrome dependent light inducible plant genes. *Nucleic Acids Res* 15:9957–9973
- Guerbette F, Grosbois M, Jolliot-Croquin A, Kader JC, Zachowski A (1999) Comparison of lipid binding and transfer properties of two lipid transfer proteins from plants. *Biochemistry* 38:14131–14137
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704
- Hiratsuka K, Chua NH (1997) Light regulated transcription in higher plants. *J Plant Res* 110:131–139
- Hu YX, Wang YX, Liu XF, Li JY (2004) *Arabidopsis* RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. *Cell Res* 14:8–15
- Hudson ME, Quail PH (2003) Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. *Plant Phys* 133:1605–1616
- Itoh J, Nonomura K, Ikeda K, Yamaki S, Inukai Y, Yamagishi H, Kitano H, Nagato Y (2005) Rice plant development: from zygote to spikelet. *Plant Cell Physiol* 46:23–47
- Johannesson H, Wang Y, Hanson J, Engstrom P (2003) The *Arabidopsis thaliana* homeobox gene ATHB5 is a potential regulator of abscisic acid responsiveness in developing seedlings. *Plant Mol Biol* 51:719–729
- José-Estanyol M, Gomis-Ruth FX, Puigdomenech P (2004) The eight-cysteine motif, a versatile structure in plant proteins. *Plant Physiol Biochem* 42:355–365
- Jung KH, Dardick C, Bartley LE, Cao P, Phetsom J, Canlas P, Seo YS, Shultz M, Ouyang S, Yuan Q, Frank BC, Ly E, Zheng L, Jia Y, Hsia AP, An K, Chou HH, Rocke D, Lee GC, Schnable PS, An G, Buell CR, Ronald PC (2008) Refinement of light-responsive transcript lists using rice oligonucleotide arrays: evaluation of gene-redundancy. *PLoS ONE* 3:3337
- Kader JC, Julienne M, Vergnolle C (1984) Purification and characterization of a spinach-leaf protein capable of transferring phospholipids from liposomes to mitochondria or chloroplasts. *Eur J Biochem* 139:411–416
- Kaufman L, Rousseeuw PJ (2008) Finding groups in data: an introduction to cluster analysis. John Wiley & Sons, Inc., Hoboken
- Kikuchi Y, Takeya T, Nakajima O, Sakai A, Ikeda K, Yamaguchi N, Yamazaki T, Tanamoto K, Matsuda H, Sawada J, Takatori K (2008) Hypoxia induces expression of a GPI-anchorless splice variant of the prion protein. *FEBS J* 275:2965–2976
- Kilian J, Whitehead D, Horak J, Wanke D, Weindl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J* 50:347–363
- Kim SY, Chung HJ, Thomas TL (1997) Isolation of a novel class of bZIP transcription factors that interact with ABA-responsive and embryo-specification elements in the Dc3 promoter using a modified yeast one-hybrid system. *Plant J* 11:1237–1251
- Kim H, Lee SB, Kim HJ, Min MK, Hwang I, Suh MC (2012) Characterization of glycosylphosphatidylinositol-anchored lipid transfer protein 2 (LTPG2) and overlapping function between LTPG/LTPG1 and LTPG2 in cuticular wax export or accumulation in *Arabidopsis thaliana*. *Plant Cell Physiol* 53:1391–1403
- Kirubakaran SI, Begum SM, Ulaganathan K, Sakthivel N (2008) Characterization of a new antifungal lipid transfer protein from wheat. *Plant Physiol Biochem* 46:918–927
- Lalanne E, Honys D, Johnson A, Borner GH, Lilley KS, Dupree P, Grossniklaus U, Twell D (2004) SETH1 and SETH2, two components of the glycosylphosphatidylinositol anchor biosynthetic pathway, are required for pollen germination and tube growth in *Arabidopsis*. *Plant Cell* 16:229–240
- Lam E, Chua NH (1990) GT-1 binding site confers light responsive expression in transgenic tobacco. *Science* 248:471–474
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948
- Lascombe MB, Bakan B, Buhot N, Marion D, Blein JP, Larue V, Lamb C, Prange T (2008) The structure of defective in induced resistance protein of *Arabidopsis thaliana*, DIR1, reveals a new type of lipid transfer protein. *Protein Sci* 17:1522–1530
- Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P (2004) Coexpression analysis of human genes across many microarray data sets. *Genome Res* 14:1085–1094
- Lee SB, Go YS, Bae HJ, Park JH, Cho SH, Cho HJ, Lee DS, Park OK, Hwang I, Suh MC (2009) Disruption of glycosylphosphatidylinositol-anchored lipid transfer protein gene altered cuticular lipid composition, increased plastoglobules, and enhanced susceptibility to infection by the fungal pathogen *Alternaria brassicicola*. *Plant Phys* 150:42–54
- Lindorff-Larsen K, Winther JR (2001) Surprisingly high stability of barley lipid transfer protein, LTP1, towards denaturant, heat and proteases. *FEBS Lett* 488:145–148
- Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21:3448–3449
- Maldonado AM, Doerner P, Dixon RA, Lamb CJ, Cameron RK (2002) A putative lipid transfer protein involved in systemic resistance signalling in *Arabidopsis*. *Nature* 419:399–403
- Maleck K, Levine A, Eulgem T, Morgan A, Schmid J, Lawton KA, Dangel JL, Dietrich RA (2000) The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nat Genet* 26:403–410
- Nakayama T, Sakamoto A, Yang P, Minami M, Fujimoto Y, Ito T, Iwabuchi M (1992) Highly conserved hexamer, octamer and nonamer motifs are positive cis-regulatory elements of the wheat histone H3 gene. *FEBS Lett* 300:167–170
- Neinhuis C, Jetter R (1995) Ultrastructure and chemistry of epicuticular wax crystals in Polytrichales sporophytes. *J Bryol* 18:399–406
- Nielsen KK, Nielsen JE, Madrid SM, Mikkelsen JD (1996) New antifungal proteins from sugar beet (*Beta vulgaris* L) showing homology to non-specific lipid transfer proteins. *Plant Mol Biol* 31:539–552
- Novillo F, Medina J, Rodriguez-Franco M, Neuhaus G, Salinas J (2012) Genetic analysis reveals a complex regulatory network modulating CBF gene expression and *Arabidopsis* response to abiotic stress. *J Exp Bot* 63:293–304
- Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR (2007) The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res* 35:D883–D887
- Park SY, Jauh GY, Mollet JC, Eckard KJ, Nothnagel EA, Walling LL, Lord EM (2000) A lipid transfer-like protein is necessary for lily pollen tube adhesion to an in vitro stylar matrix. *Plant Cell* 12:151–164
- Patel BN, Dunn RJ, David S (2000) Alternative RNA splicing generates a glycosylphosphatidylinositol-anchored form of ceruloplasmin in mammalian brain. *J Biol Chem* 275:4305–4310
- Pierleoni A, Martelli PL, Casadio R (2008) PredGPI: a GPI-anchor predictor. *BMC Bioinformatics* 9:392
- Pollard M, Beisson F, Li Y, Ohlrogge JB (2008) Building lipid barriers: biosynthesis of cutin and suberin. *Trends Plant Sci* 13:236–246

- Ranathunge K, Schreiber L, Franke R (2011) Suberin research in the genomics era—new interest for an old polymer. *Plant Sci* 180:399–413
- Ren XY, Fiers MW, Stiekema WJ, Nap JP (2005) Local coexpression domains of two to four genes in the genome of *Arabidopsis*. *Plant Physiol* 138:923–934
- Reyes JC, Muro-Pastor MI, Florencio FJ (2004) The GATA family of transcription factors in *Arabidopsis* and rice. *Plant Physiol* 134:1718–1732
- Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, Miller N, Mueller LA, Mundodi S, Reiser L, Tacklind J, Weems DC, Wu Y, Xu I, Yoo D, Yoon J, Zhang P (2003) The *Arabidopsis* information resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res* 31:224–228
- Rinne PL, Welling A, Vahala J, Ripel L, Ruonala R, Kangasjarvi J, van der Schoot C (2011) Chilling of dormant buds hyperinduces FLOWERING LOCUS T and recruits GA-inducible 1,3- β -glucanases to reopen signal conduits and release dormancy in *Populus*. *Plant Cell* 23:130–146
- Sablowski RW, Moyano E, Cuihanz-Macia FA, Schuch W, Martin C, Bevan M (1994) A flower-specific Myb protein activates transcription of phenylpropanoid biosynthetic genes. *EMBO J* 13:128–137
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37:501–506
- Schreiber L (2010) Transport barriers made of cutin, suberin and associated waxes. *Trends Plant Sci* 15:546–553
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504
- Shepherd T, Wynne Griffiths D (2006) The effects of stress on plant cuticular waxes. *New Phytol* 171:469–499
- Shin DH, Lee JY, Hwang KY, Kim KK, Suh SW (1995) High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. *Structure* 3:189–199
- Simpson C, Thomas C, Findlay K, Bayer E, Maule AJ (2009) An *Arabidopsis* GPI-anchor plasmodesmal neck protein with callose binding activity and potential to regulate cell-to-cell trafficking. *Plant Cell* 21:581–594
- Sodano P, Caille A, Sy D, de Person G, Marion D, Ptak M (1997) 1H NMR and fluorescence studies of the complexation of DMPG by wheat non-specific lipid transfer protein global fold of the complex. *FEBS Lett* 416:130–134
- Sterk P, Booi H, Schellekens GA, Van Kammen A, De Vries SC (1991) Cell-specific expression of the carrot EP2 lipid transfer protein gene. *Plant Cell* 3:907–921
- Thoma S, Hecht U, Kippers A, Botella J, De Vries S, Somerville C (1994) Tissue-specific expression of a gene encoding a cell wall-localized lipid transfer protein from *Arabidopsis*. *Plant Phys* 105:35–45
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Tilly JJ, Allen DW, Jack T (1998) The CARG boxes in the promoter of the *Arabidopsis* floral organ identity gene APETALA3 mediate diverse regulatory effects. *Development* 125:1647–1657
- Troncoso-Ponce MA, Mas P (2012) Newly described components and regulatory mechanisms of circadian clock function in *Arabidopsis thaliana*. *Mol Plant* 5:545–553
- Wang J, Gunning W, Kelley KM, Ratnam M (2002) Evidence for segregation of heterologous GPI-anchored proteins into separate lipid rafts within the plasma membrane. *J Membrane Biol* 189:35–43
- Wang SY, Wu JH, Ng TB, Ye XY, Rao PF (2004) A non-specific lipid transfer protein with antifungal and antibacterial activities from the mung bean. *Peptides* 25:1235–1242
- Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ (2007) An electronic fluorescent pictograph browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* 2:e718
- Woo HR, Kim JH, Kim J, Kim J, Lee U, Song IJ, Kim JH, Lee HY, Nam HG, Lim PO (2010) The RAV1 transcription factor positively regulates leaf senescence in *Arabidopsis*. *J Exp Bot* 61:3947–3957
- Yakir E, Hilman D, Harir Y, Green RM (2007) Regulation of output from the plant circadian clock. *FEBS J* 274:335–345
- Zachowski A, Guerbette F, Grosbois M, Jolliot-Croquin A, Kader JC (1998) Characterisation of acyl binding by a plant lipid-transfer protein. *Eur J Biochem* 257:443–448
- Zhang DS, Liang WQ, Yuan Z, Li N, Shi J, Wang J, Liu YM, Yu WJ, Zhang DB (2008) Tapetum degeneration retardation is critical for aliphatic metabolism and gene regulation during rice pollen development. *Mol Plant* 1:599–610
- Zheng BS, Rönberg E, Viitanen L, Salminen TA, Lundgren K, Moritz T, Edqvist J (2008) *Arabidopsis* sterol carrier protein-2 is required for normal development of seeds and seedlings. *J Exp Bot* 59:3485–3499