

Evolution of non-specific lipid transfer protein (nsLTP) genes in the Poaceae family: their duplication and diversity

Cheol Seong Jang · Won Cheol Yim · Jun-Cheol Moon · Je Hyeong Jung ·
Tong Geon Lee · Sung Don Lim · Seon Hae Cho · Kwang Kook Lee ·
Wook Kim · Yong Weon Seo · Byung-Moo Lee

Received: 5 December 2007 / Accepted: 21 January 2008 / Published online: 13 February 2008
© Springer-Verlag 2008

Abstract Previously, the genes encoding non-specific lipid transfer proteins (nsLTPs) of the Poaceae family appear to evidence different genomic distribution and somewhat different shares of EST clones, which is suggestive of independent duplication(s) followed by functional diversity. To further evaluate the evolutionary fate of the Poaceae nsLTP genes, we have identified *Ka/Ks* values, conserved, mutated or lost *cis*-regulatory elements, responses to several elicitors, genome-wide expression profiles, and nsLTP gene-coexpression networks of both (or either) wheat and rice. The *Ka/Ks* values within each group and between groups appeared to be similar, but not identical, in both species. The conserved *cis*-regulatory elements, e.g. the RY repeat (CATGCA) element related to ABA regulation in group A, might be reflected in some degree of long-term conservation in transcriptional regulation post-

dating speciation. In group A, wheat nsLTP genes, with the exception of *TaLTP4*, evidenced responses similar to those of plant elicitors; however, the rice nsLTP genes evidenced differences in expression profiles, even though the genes of both species have undergone purifying selection, thereby suggesting their independent functional diversity. The expression profiles of rice nsLTP genes with a microarray dataset of 155 gene expression omnibus sample (GSM) plates suggest that subfunctionalization is not the sole mechanism inherent to the evolutionary history of nsLTP genes but may, rather, function in concert with other mechanism(s). As inferred by the nsLTP gene-coexpression networks, the functional diversity of nsLTP genes appears not to be randomized, but rather to be specialized in the direction of specific biological processes over evolutionary time.

Keywords Duplication · Evolutionary fate · Functional diversity · Gene family · nsLTP · Poaceae

Communicated by Y. Van de Peer.

C.S. Jang and W.C. Yim are equally contributed to this report.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-008-0327-4) contains supplementary material, which is available to authorized users.

C. S. Jang (✉)
Institute of Life Science and Natural Resources, Korea University,
Seoul 136-713, South Korea
e-mail: csjang@korea.ac.kr

J. H. Jung · T. G. Lee · S. H. Cho · K. K. Lee · W. Kim · Y. W. Seo
Division of Biotechnology, Korea University,
Seoul 136-713, South Korea

W. C. Yim · J.-C. Moon · S. D. Lim · B.-M. Lee (✉)
Department of Plant Biotechnology, Dongguk University,
Seoul 100-715, South Korea
e-mail: bmlee@dongguk.edu

Introduction

The complete genome sequencing of model plants including rice and *Arabidopsis*, together with the development of high-throughput methods for the expression profiling of genes, would facilitate the identification of the function(s) of a variety of genes. For example, rice microarray analysis using cDNA or oligonucleotide chips allows for the simultaneous quantification of tens of thousands of gene transcripts and their repository public databases, such as the gene expression omnibus (GEO) in NCBI, allow for a practical exploration of the expressional values of a huge amount of transcripts, thereby addressing a wide range of biological issues. In another case, namely wheat EST analysis, which has been completed up to a point of 1,050,926,

as of 17th August 2007, followed by physical mapping using aneuploid lines, would provide us with an understanding of its genomic content, even though it possesses an exceedingly complex hexaploid, with a large genome size of 16,000 Mbp (e.g. Qi et al. 2004). The rapid growth of genomic databases may facilitate the prediction and identification of the function(s) of all genes on a genome via in silico analysis, with the development of bioinformatic tools.

The complexity of plant genomes, as compared to those of micro- and animal organisms, may be the result of repetitive whole genome duplications, thereby further supporting the high frequency of members of a plant gene family (Lockton and Gaut 2005). In the genome of *Arabidopsis*, for instance, only 35% of genes are single-copy, whereas >37% are members of a gene family consisting of five or more copies, which may be formed via multi-time whole genome duplications (*Arabidopsis* Genome Initiative 2000). The ancient polyploidization event of the Poaceae family, which includes the edible crops that are utilized as major carbohydrate sources for humans, appears to have occurred approximately 70 million years ago (Mya), predating their divergence (Paterson et al. 2004). In addition, a segmental duplication of 3-Mbp has been shown to occur in the distal portions of rice chromosomes 11 and 12, approximately 7.7 Mya (The rice chromosomes 11 and 12 sequencing consortia 2005). Blanc and Wolfe (2004a) anticipated widespread paleopolyploidization events in flowering plants via the in silico analysis of 14 model plants.

Duplicate genes, members of a gene family, are formed via whole genome duplication and/or single gene amplification may perform a crucial role in driving both evolutionary and genetic diversity, as compared to single-copy genes. Papp et al. (2003) determined that the total number of regulatory motifs in each duplicate gene does not change with evolutionary time, although the number of shared regulated motifs in the duplicate genes is reduced, thereby suggesting that subfunctionalization is not the sole mechanism underlying expression diversity. Blanc and Wolfe (2004b) determined that significantly asymmetric duplicate pairs in protein sequence evolution tend toward more divergent expression. In general, the evolutionary fate of duplicate genes might arrive at a conclusion of functional diversity following mutations in the coding and/or regulatory regions (For a review, see Moore and Purugganan 2005).

The plant MSDS-box gene family, the members of which are key regulators of vegetative and reproductive development, is an excellent example of the evolutionary fate of duplicate genes. For example, phylogenetic analyses of the *AGMOUS* (*AG*) subfamily have shown that their ancient gene duplications were crucial to shaping their evolution, resulting in dissociation among genetic modules,

and thereby allowing for an increase in morphological diversity (Kramer et al. 2004). Recently, Zahn et al. (2006) reported that functional diversity among orthologues is evident, as is convergence among paralogues sampled from different species, and also proposed that the retention of multiple *AG* homologues in several paralogous lineages can be explained by the conservation of ancestral gene functions.

Plant lipid transfer proteins (LTPs) with a molecular mass of 9 kDa are present in abundant quantities, up to as much as 4% of total soluble proteins. In plants, genes encoding non-specific LTPs are members of a typical small multigene family, thereby suggesting that their different possible functions, including cutin synthesis, β -oxidation, somatic embryogenesis, allergenicity, plant signaling, and plant defense (for a review, see Carvalho Ade and Gomes 2007). However, the true physiological functions of the nsLTP genes remain to be clearly elucidated, although a host of studies have focused on the regulation and diversity of the expression of nsLTP genes against environmental stresses and at different developmental stages. An outstanding study using tDNA tagging mutant lines revealed that an *Arabidopsis* nsLTP gene corresponding to *defective in induced resistance 1-1* (*dirl-1*) is involved in systemic resistance signaling (Maldonado et al. 2002). However, considering that at least 15 nsLTP genes of 9 kDa exist within the *Arabidopsis* genome (Arondel et al. 2000), integrated investigations via genomics approaches will be required in order to clarify their functions, in addition to reverse genetic research.

In our previous report, the genes encoding non-specific lipid transfer proteins (nsLTPs) of the Poaceae family appear to evidence different genomic distribution and somewhat disproportionate shares of EST clones among the cereal nsLTP genes, thereby suggesting independent duplication(s) followed by functional diversity (Jang et al. 2007). For example, an increase in the numbers of rice nsLTP genes has been attributed to a segmental duplication on the distal portion of rice chromosomes 11 and 12 after independent tandem duplications. In the case of the wheat nsLTP genes, the growth in their dosage appears to be ascribed to polyploidizations, especially considering their chromosomal localizations. It remains questionable as to what the evolutionary fate of Poaceae nsLTP genes are, after independent gene duplications. In this study, we have attempted to evaluate the functional diversity of nsLTP genes in both wheat and rice, via RT-PCR and in silico analysis conducted using a public database. In addition, we have also evaluated the selection mode of the nsLTP genes via the calculation of non-synonymous and synonymous substitutions per site (K_a/K_s), in addition to conserved regulatory elements, via a phylogenetic fingerprinting method.

Materials and methods

Phylogenetic analysis

The sequences encoding the 9 kDa nsLTPs of rice, wheat, barley, maize, and sorghum were retrieved, as has been described elsewhere (Jang et al. 2007) with some modifications. For more mining sequences encoding the nsLTPs of other monocot plants, the 24,245 EST sequences for *Allium* and the 18,415 EST sequences for *Musa* were downloaded from the EST database of NCBI as of 24th June 2007, then clustered using the CAP3 program (Huang and Madan 1999), thereby resulting in tentative unigenes of 14,168 and 14,332, respectively. The cereal nsLTP genes were matched against the unigene sets of *Allium* and *Musa* using the tBLASTx program with a cut-off of e^{-10} . The outgroup used was comprised of a total of 15 *Arabidopsis* nsLTPs, as reported by Arondel et al. (2000) for phylogenetic analysis. The deduced amino acid sequences of all the plant taxa listed above were aligned with the ClustalW program with default parameters, and alignment was then edited manually using the BioEdit software (Hall 1999). The phylogenetic methodology was conducted with the MEGA software (version 3.1; Kumar et al. 2004) and the neighbor-joining algorithm (Saitou and Nei 1987), with 1,000 bootstrap replicates.

Ka/Ks calculation

The amino acid sequences of the nsLTPs of rice and wheat were aligned and then adjusted manually, as described above. Synonymous and non-synonymous substitutions per site (K_s and K_a) were calculated with the MEGA software via a modified Nei–Gojobori method (Nei and Kumar 2000). The K_a/K_s values were averaged within a group or between groups.

Isolation of 5' promoter regions of wheat nsLTP genes

Genomic DNA was extracted from a wheat-rye translocation line with 2RL, as described by Seo et al. (1997) with some modifications. For DNA walking onto the 5' upstream sequences adjacent to the wheat nsLTP genes, genomic pools were constructed using a GenomeWalker™ universal kit (Clontech Laboratories, Inc.), in accordance with the commercial protocols after digestion with each of the following eight enzymes: *Dra* I, *Eco*R V, *Pvu* II, *Stu* I, *Avi* II, *Hind* II, *Nru* I, and *Ssp* I. Two PCR amplifications per library were conducted with each of the gene-specific primers of the wheat nsLTP genes, the sequences of which are provided in Table S1, and each of the adapter primers at adequate anneal temperature for each primer pairs, as described elsewhere (Jang et al. 2004). The PCR products

were then cloned into pGEM-T easy vector (Promega, Madison, WI) and sequenced with an ABI3700 automated DNA analyzer (Applied Biosystems, Foster City, CA).

Phylogenetic footprinting analysis

In order to retrieve the putative promoter regions of the rice nsLTP genes, the rice unigene sequences encoding nsLTP genes were BLAST-searched against the rice pseudomolecules (version 4; <http://www.tigr.org/tdb/e2k1/osa1/>) in order to detect the rice pseudogenes, which were then best matched to the sequences. One kbp upstream sequences from the translation start site were downloaded manually from the TIGR rice genome annotation database and resource. The putative promoter sequences of wheat nsLTP genes with 1 kbp or less and rice putative promoter sequences of 1 kbp were analyzed using the FootPrinter program (Blanchette and Tompa 2003). The promoter sequences of both species were then divided into five different groups on the basis of phylogenetic analyses of the coding regions of their nsLTP genes, as described by Jang et al. (2007). In order to determine the conserved or lost motifs among the promoters of both species within the same group, we utilized default parameters, with motif sizes of 6, 8, and 10, and allowed for regulator element loss with a significant level spanned tree, and a motif loss of a cost of 1. The motif sequences obtained were compared with the known *cis*-elements via a PLACE web signal scan (Higo et al. 1999).

Hormone treatments and RT-PCR

The wheat-rye translocation line was permitted to germinate on moistened filter paper and then transferred into containers as previously described by Lee et al. (2007). After 7 days, the plants were grown in 1/1,000 diluted Hypo-nex™ solution until the hormone treatments were applied. Young seedlings grown for 2 weeks were transferred into each new container with 100 μ M of methyljasmonic acid (MeJA), 5 mM of salicylic acid (SA), 50 μ M of abscisic acid (ABA) and 10 mM of H_2O_2 . For ethylene treatment, young plants were exposed to a dose of 100 μ L L⁻¹ ethylene gas in transparent plastic, airtight containers. The shoots of the treated plants were harvested at 3, 9, and 24 h after each treatment, respectively.

RNAs were extracted with Trizol reagent (Invitrogen Corp., Carlsbad, CA), in accordance with the manufacturer's instructions. Two μ g of total RNA from each sample was utilized as a template for first-strand cDNA synthesis in a total reaction volume of 20 μ L containing 50 mM Tris–HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 50 μ M dNTPs, 20 U AMV reverse transcriptase (iNtRON Biotechnology, Seongnam, South Korea) and 50 pmol of the

oligo-(dT)₁₈ primer. The reactions were incubated for 60 min at 42°C, followed by a final 5-min denaturation at 95°C. One-fifth of the first-strand cDNAs were used as employed in a total reaction volume of 25 µL containing 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.1 mM of each dNTP, 0.5 µL of each gene-specific primers, and 0.6 U of *Taq* polymerase (Takara Kyoto, Japan). The PCR program was as follows: 5 min at 94°C, followed by cycles of 45 s at 94°C, 45 s at each annealing temperature, and 1 min at 72°C, with a final extension step at 72°C (see Table S1 for gene-specific primer sequences, PCR cycles, and annealing temp.).

Rice expression data

A set of 155 gene expression omnibus samples (GSM) of the Affymetrix GeneChip rice genome array (accession no. GPL2025) were downloaded from the Gene Expression Omnibus (GEO) database in NCBI, as of 1st July 2007. Affymetrix GeneChip spot IDs matched to the rice nsLTP genes were retrieved via TIGR Genome Browser (http://www.tigr.org/tigr-scripts/osal_web/gbrowse/rice/). The expression data of each of the nsLTP genes were gathered manually from the 155 GSM data of the rice genome array. The expression values from each plate were subsequently normalized by dividing each individual signal by the average of the plate. For the clustering analysis of individual rice nsLTP genes and of GSM arrays based on the Affymetrix GeneChip expression data, the average linkage hierarchical clustering of the data accomplished via the uncentered Pearson correlation was conducted using the C clustering library source code (de Hoon et al. 2004) through 64-bit arithmetic on an IBM p690 Unix platform at KISTI (Korea Institute of Science and Technology Information, Daejeon, South Korea).

In order to find clusters of co-expressing genes with each of the rice nsLTP genes above the dataset, we conducted analyses using the SOTAarray program (<http://gepas.bio-info.cipf.es/cgi-bin/sotarray>) with default parameters, with the exception of complete linkage using correlation distance with linear correlation coefficient (Herrero et al. 2001). The expression profile of co-expressing genes in each cluster was visualized using SigmaPlot version 9.0 (Systat Software, Richmond, CA, USA).

Results

Phylogenetic analysis of unigenes encoding 9 kDa nsLTPs

In the previous report, we divided the unigenes encoding the nsLTPs of the Poaceae family into five different groups on the basis of a phylogenetic tree, thereby suggesting that they had undergone independent duplication(s) followed by

functional diversity (Jang et al. 2007). In order to further define the evolutionary history of the nsLTP genes in monocots, nine unigenes encoding the nsLTPs—four for *Allium* (onion) and five for *Musa* (banana)—which diverged prior to the independent evolution of cereal lineages, were retrieved from the EST databases of both species. A total of 76 unigenes in monocots, including 62 nsLTP genes of the Poaceae family, in addition to 15 *Arabidopsis* nsLTPs (Arondel et al. 2000) as an outgroup were employed for phylogenetic analysis. As is shown in Fig. 1, seven *Arabidopsis* nsLTPs were clustered into a new group, whereas eight others were close to group D, which was clustered by Jang et al. (2007). Curiously, all five of the *Allium* nsLTPs were clustered with seven *Arabidopsis* nsLTPs into the new group, even if the *Allium* family was closer to those of *Musa* and Poaceae than to that of *Arabidopsis*, which is in fact a dicot. In *Musa*, three nsLTPs were clustered into group B, whereas one was in group E. In the previous phylogenetic analysis, group A was comprised of two other subgroups. However, group A appeared to encompass two different groups in the new tree. We concluded that the previous clustering was suitable for the nsLTPs of the Poaceae family, thereby suggesting that they have undergone relatively rapid evolution, including repetitive independent duplication events, since the ancient divergence of at least *Allium* and Poaceae.

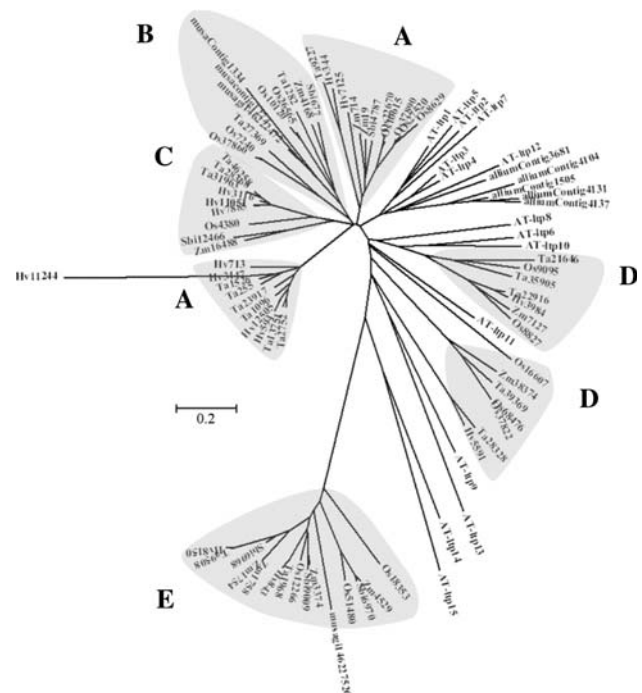


Fig. 1 A phylogenetic tree of nsLTP genes of monocot plants on the basis of sequence alignment of the amino acid sequences derived from the unigene CDSs. The phylogenetic analysis was conducted using the neighbor-joining algorithm with 1,000 bootstrap replicates. The scale bar indicated 0.2 substitutions per amino acid

Non-/synonymous substitutions (K_a/K_s) of nsLTPs in rice and wheat

In order to evaluate the selection mode of the plant nsLTP genes, those of rice and wheat were selected for further study. A total of 18 nsLTPs for rice and 19 for wheat were divided into five different groups on the bases of the phylogenetic analysis, as previously described by Jang et al. (2007). Eighteen unigenes encoding nsLTPs of rice were divided into five groups, resulting in 5 for group A, 4 for group B, 1 for group C, 5 for group D, and 3 for group E. By way of contrast, a total of 19 unigenes for wheat were clustered as follows: 7 for group A, 2 for group B, 3 for group C, 5 for group D, and 2 for group E, respectively. We calculated the K_a/K_s ratios over the mature coding regions between the nsLTPs of rice and wheat, respectively (Fig. 2). As calculated by pairwise comparisons within each group, group A for rice and groups A and C for wheat revealed K_a/K_s ratios of less than 1.000 (from 0.189 to 0.510), thereby indicating that they have undergone purifying selection. Other groups in both species with the exception of group B (≈ 1.000 , 1.040) and C (not determined) in rice evidenced K_a/K_s ratios of more than 1.000 (from 1.145 to 2.343), thereby reflecting positive selection. When pairwise comparisons were conducted between groups in each species, a slow evolutionary process appeared to have occurred between groups A and B (0.850) and between groups A and C (0.818) in rice, while for wheat this process appeared to have occurred between A and B (0.638) and between A and C (0.655). In rice, pairwise comparisons between other groups indicated a relatively rapid evolutionary process, with ranges of K_a/K_s ratios of 1.232–2.956. Similarly, pairwise comparisons between other groups were

reflective of a rapid evolutionary process with K_a/K_s ranges of 1.166–2.016, except between groups A and C (≈ 1.000 , 0.986) and between B and C (≈ 1.000 , 1.049).

Conserved, mutated or lost motifs on nsLTP promoters during speciation

We then attempted to discover conserved, mutated, or lost motifs on the upstream regions of nsLTP genes of rice and wheat. Among 12 wheat nsLTP genes, the upstream regions of seven genes were isolated from the wheat-rye translocation line, including 5 for group A (those of *TaLTP1*, 2, 4, 7, and 10), 1 for group B (*TaLTP3*), and 1 for group C (*TaLTP5*). The sizes of the isolated promoters ranged from 271 to 2,858 bp (Table 3), followed by 1 kb or less for further analysis. A total of 17 one kb fragments upstream of the nsLTP gene sequences were retrieved based on the BLASTN analysis of rice unigenes against the rice pseudomolecules. The utilized gene models lists are as follows: Os12g02310, Os12g02320, and Os11g24070 for group A, Os11g02330, Os11g02400, Os11g02424, Os12g02290, Os12g02330, and Os12g02340 for group B, Os11g02350 and Os12g02300 for group C, Os01g60740 for group D, and Os01g62980, Os04g33920, Os04g33930, and Os05g06780 for group E.

We attempted to employ the parameters of different motif sizes to detect conserved motifs among the nsLTP genes of rice and wheat included in the same group (Table 1). It is believed that there are adequate motif sizes of 6–10. In the case of group A, 8 promoter sequences—3 for rice and 5 for wheat—of different lengths were compared in order to discover common as well as lost regulatory elements, which are not detected in some nsLTP gene(s). As had been

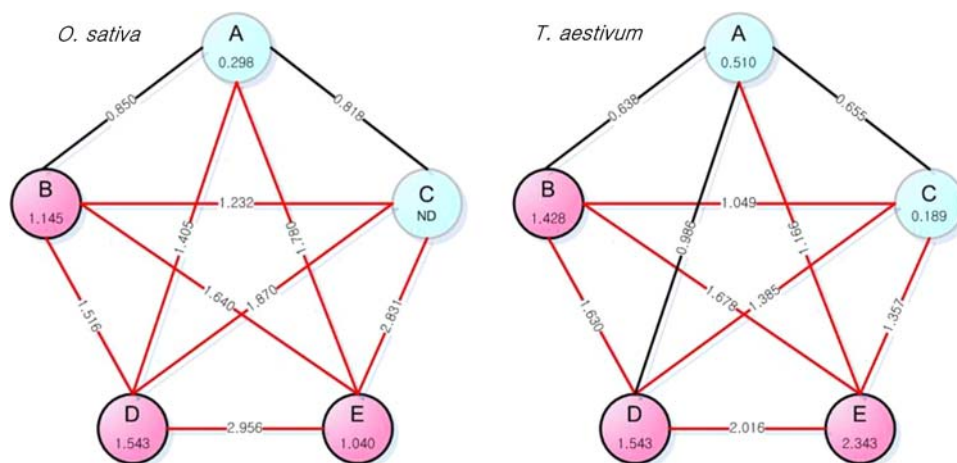


Fig. 2 Average K_a/K_s ratios within each group or between groups of rice (*Oryza sativa*) and wheat (*Triticum aestivum*) nsLTP genes, respectively. K_s and K_a were calculated using the MEGA software with the modified Nei–Gojobori method. The circles filled with pink color (with dark gray in the print version) and the red lines (gray lines

in the print version) indicate more than 1 of K_a/K_s values within each group and between groups, respectively. The circles filled with blue color (with light gray in the print version) and the black lines (black lines in the print version) represent K_a/K_s values of less than 1

Table 1 Numbers of conserved or loss motifs on nsLTP promoters between rice and wheat

Group	No. of tested promoters		Motif size		
	Rice	Wheat	6	8	10
A ^a	3	5	7/22 ^b	1/30	0/5
B	6	1	12/39	5/43	2/44
C	2	0	ND	ND	ND
D	2	1	ND	ND	0/7
E	4	0	51/0	28/26	1/5

ND not determined

^a Groups were divided according to the phylogenetic analysis reported by Jang et al. (2007)

^b Indicated numbers of conserved and loss motifs which are not contained in some nsLTP gene(s)

expected, phylogenetic footprinting analysis with small motif sizes could be utilized to detect more conserved motifs than those of large motif size. For example, seven conserved motifs were detected using a motif size of six, whereas only

one motif was detected for a motif size of eight. However, out of eight conserved motifs, one motif (ATTTAA) was conserved in all of the analyzed promoters, whereas the others were mutated in certain gene promoter(s) of 1–4 (Table 2). When scanning the conserved motif sequences against the PLACE database three known regulatory elements were matched, resulting in the RY repeat detected in the RY/G box (Ezcurra et al. 1999), the rbsS boxII found in the upstream region of the pea rbsS gene (Green et al. 1988), and the TATA box bound by OsTBP2 (Zhu et al. 2002). The lost regulatory elements detected in group A ranged from 5 to 30, according to changes in motif sizes.

In the phylogenetic analysis of group B, a total of 19 conserved motifs using motif sizes of 6–10 were retrieved from 7 promoter sequences, including 6 for rice and 1 for wheat. One ATTTATAC motif was conserved among all tested promoters, and was matched to two known *cis*-elements; one was detected in the upstream sequence of phosphate starvation-responsive genes (Rubio et al. 2001), and the other was identified as a type of the TATA box (Shirsat et al. 1989). In group C, the promoter sequences of

Table 2 Profiles of selected motifs, which contained conserved or mutated regulatory elements on nsLTP promoters of rice and wheat as predicted by FootPrinter (Blanchette and Tompa 2003)

Group	Conserved motif sequences	No. of genes conserved	No. of genes mutated	Regulatory element	Site number in place	Description
A	CATGCA	7	1	RYREPEAT	S000264	RY repeat found in RY/G box
	CAAATT	6	2	NF		
	CCACCC	7	1	NF		
	ATTTAA	8	0	NF		
	CCACAC	4	4	RBCBOX2	S000094	rbcS box II, 5' upstream region of pea rbcS gene
	TAGCTA	6	2	NF		
	CACCAC	7	1	NF		
	TATTTAAG	6	2	TATABOX	S000400	Binding site for OsTBP2, found in the promoter of rice pal gene
B	CCCTCCAC	7	0	NF		
	ATTTATAC	7	0	P1BS	S000459	Sequences found in the upstream regions of phosphate starvation responsive genes
				TATABOX2	S000109	TATA Box
	ACCCTCCA	7	0	NF		
	CCTATTTA	5	2	NF		
	ACCCATCG	7	0	NF		
	ACCCTCCACC	5	2	NF		
	CCTATTTATA	5	2	TATABOX2	S000109	TATA Box
E	GTTTTTCTGT	2	2	D1GMAUX28	S000328	Sequences found in the soybean auxin responsive gene
				GT1	S000198	Consensus GT-1 binding site in many light-regulated genes
				GT1	S000453	GT-1 motif; inducible by pathogen and NaCl

Regulatory elements were predicted by the signal scan of the PLACE database (Higo et al. 1999)

Os11g02350 and Os12g02300 were analyzed in order to detect the conserved motifs. However, we were unable to detect any conserved motifs, even when using a motif size of 12, due to the high level of conservation of both sequences derived from their segmental duplication, which occurred approximately 11 Mya. Similarly, the promoter sequences of three genes, including group D, evidenced no conserved motif sequences. Four rice promoter sequences, including group E, harbored highly conserved motifs with relatively small motif sizes, e.g. six and eight. When utilizing a motif size of ten, however, only one motif of GTTTTCTGT with mutated sequences was detected, and was matched to three *cis*-elements, e.g. sequences found in the auxin-responsive gene (Nagao et al. 1993), the consensus GT-1 binding site in many light-regulated genes (Villain et al. 1996), and the GT-1 motif inducible by pathogen and NaCl (Park et al. 2004).

Expression of wheat nsLTP genes

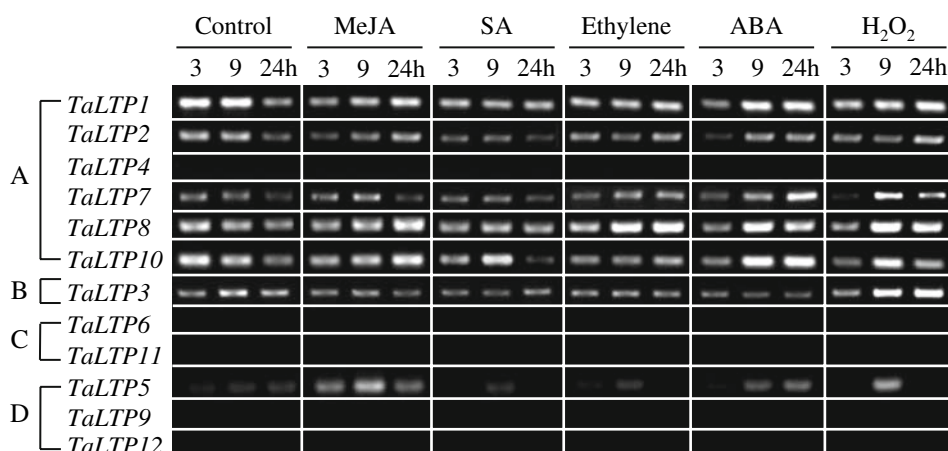
To evaluate the transcriptional diversity of wheat nsLTP genes, young seedlings of the wheat-rye translocation line, the *TaLTP* genes of which were isolated previously (Jang et al. 2002, 2005, 2007), were treated with each of five plant signal molecules, MeJA, SA, ethylene, ABA, and H₂O₂ (Fig. 3). The responses of *TaLTP* genes to plant signal molecules evidenced conspicuous differences among groups. The *TaLTP* genes in group A, with the exception of *TaLTP4*, evidenced responses similar, but not identical, to the five elicitors. In a previous report, we determined that the *TaLTP* genes in group A, with the exception of *TaLTP4*, were located on three homoeologous chromosomes, 3B and 3D, which suggests that they harbor features of the relationship of homoeologous genes formed via polyploidization. For example, MeJA treatment caused an increase in transcripts of *TaLTP 1, 2, 8, and 10* until 24 h, but not in those of *TaLTP7*, whereas the control treatment reduced the transcripts of the *TaLTP* genes at the same

times. With SA treatment, *TaLTP10* evidenced a transcript peak at 9 h, but this was not observed in the others. Ethylene gas treatment induced a significant increase in *TaLTP8* transcripts, but not in any other genes. In cases of ABA and H₂O₂ treatments, all transcripts of the group A *TaLTP* genes, except for *TaLTP4*, were increased at 9 and/or 24 h after both treatments when compared at 3 h. The gene-specific primer pair of *TaLTP4* generated no amplicons using the cDNAs acquired from the hormone-treated and control plants; however, they were able to isolate PCR products encoding *TaLTP4* using genomic DNA as a template, as has been reported previously (Jang et al. 2007). By way of contrast, the transcripts of the *TaLTP3* gene included in group B were increased at 9–24 h as the result of H₂O₂ treatment, whereas it was steadily expressed as the result of treatment with other hormones. *TaLTP6* and *11*, all of which are included in group C, were not amplified in any of the shoots of healthy young plants or hormone-treated plants. In group D, the *TaLTP5* gene was induced by treatment with MeJA, ethylene, ABA, and H₂O₂, but not by SA treatment. In the case of *TaLTP9*, we were unable to amplify any products from the genomic DNA and cDNAs using two pairs of gene-specific primers designed from the unigene sequences. It is believed that the *TaLTP9* gene was lost in the wheat-rye translocation line, or that the primer design may have been wrong. In the previous report, we discovered a mutation of the *TaLTP12* gene with a 24 bp deletion in its coding region in the wheat-rye translocation line (Jang et al. 2007), further supporting the absence of *TaLTP12* transcripts in the shoots of young plants.

Expression of rice nsLTP genes

A total of 155 GSM samples of the Affymetrix GeneChip rice genome array were downloaded from the public database and clustered via average linkage hierarchical clustering. Four array groups could be generally divided on the basis of the clustering (data not shown). The array group-

Fig. 3 Expression profiles of *TaLTP* genes in 2-week-old plant shoots. The material samples were harvested from the plants treated with MeJA (100 mM), SA (5 mM), ABA (50 μ M), H₂O₂ (10 mM) and ethylene gas (100 μ L L⁻¹) at 3, 9 or 24 h after the treatment



It is believed that the co-expression of genes across a given series of experiments is most probably explicable by virtue of some common functional role. In an effort to discover co-expressing genes with each of the nsLTP genes, clustering analyses were conducted using the dataset of 155 GSM samples with the Self Organising Tree Algorithm (SOTA), resulting in a total of 1,660 clusters. In order to verify that the clustering was performed properly, we selected two genes encoding phosphoribulokinase and sedoheptulose 1,7-bisphosphatase, which are included in the Calvin cycle, and then detected a cluster in which both genes were included. Interestingly, cluster 1621 with a set of 77 spots that encode some Calvin cycle enzymes, e.g. ribulose biphosphate carboxylase small chain A, ribulose biphosphate carboxylase/oxygenase activase, glyceraldehyde-3-phosphate

appeared to be expressed more abundantly in the leaves than in other tissues, reflecting some of the features of the Calvin cycle (Fig. 5).

We visualized the expression profiles of each of the nsLTP gene(s) grouped into the same cluster, on the basis of their normalized expression values (Fig. 6). A total of

Thick lines and thin lines indicate the average and individual gene expression profiles, respectively. The *dot and dash line* represents the average expression value ± 2 fold values



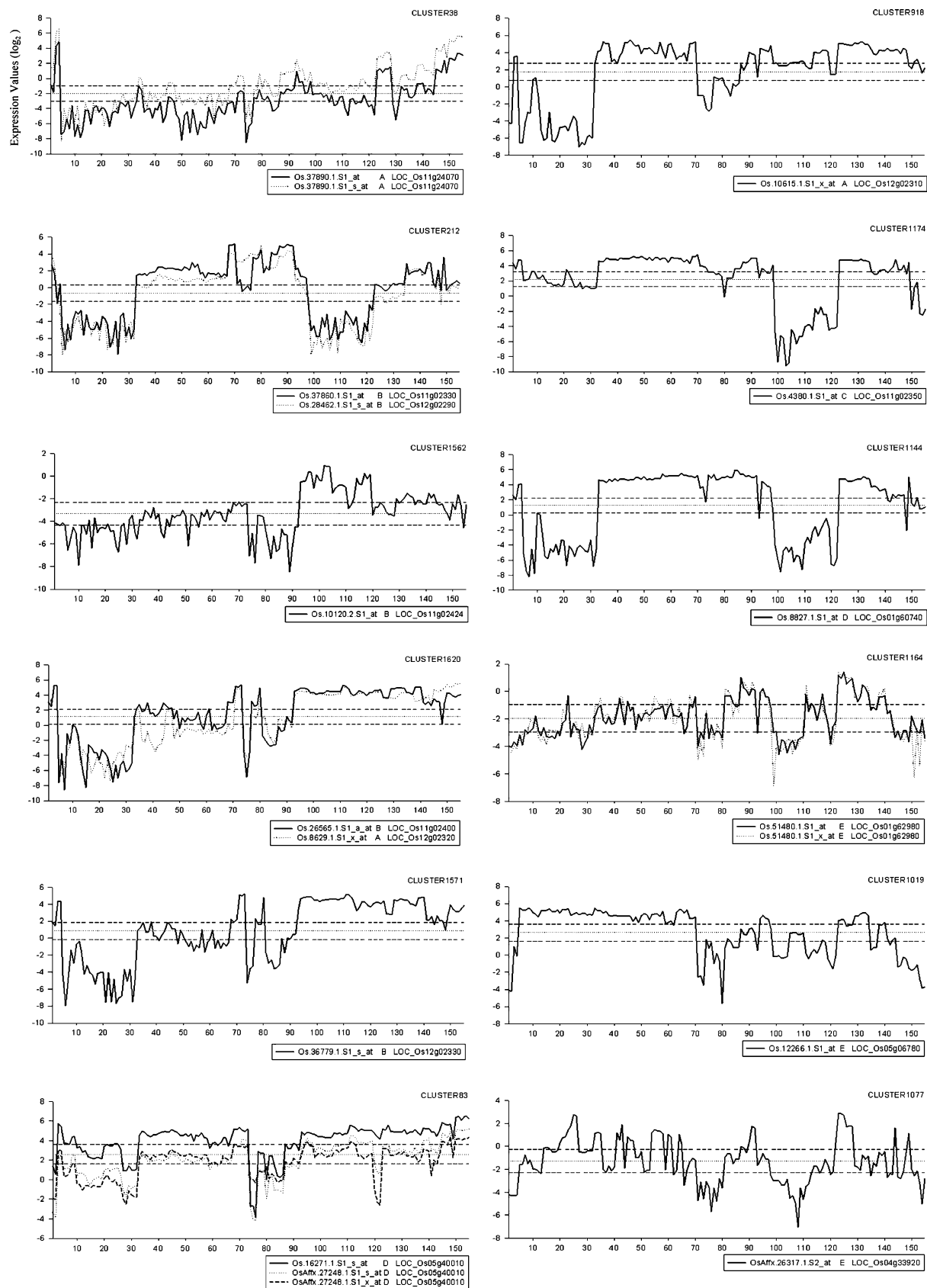


Fig. 6 Expression profiles of each of rice nsLTP genes. For descriptions in detail, see the legend for Fig. 5

12 clusters were retrieved, including at least one spot corresponding to the nsLTP gene. Two clusters, 212 and 1620, consisted of two different nsLTP genes, whereas the other clusters included only one nsLTP gene. For example, cluster 212 harbored two nsLTP genes, which were inferred from the segmental duplication of distal portions of chromosome 11 and 12, thereby reflecting their functional conservation. By way of contrast, two genes of cluster 1620, Os12g02320 and Os11g02400, evidenced similar expression profiles, although they were divided into different groups, respectively, on the basis of their sequence similarity. The expression profiles of clusters 1620 and 1571 closely resembled one another, and this was reflected in the relationship of the segmental duplicate genes of Os11g02400 (cluster 1620) and Os12g02330 (cluster 1571). Both of the nsLTP genes of Os11g02350 and Os01g60740 evidenced relatively similar expression profiles, but significant differences in root tissues were observed, particularly between the stress-treated tissues and those left untreated. Next, we retrieved the co-expressing genes of 12 different clusters, including the nsLTP spots 1–3; each cluster was involved in 9 spots for the cluster 1164, and 460 for cluster 1571 (Table 3). The average expression profile of genes in each cluster, coupled with the individual expression profiles of each gene, was visualized via SOTA. In order to evaluate the microarray target samples in which the genes in each cluster were abundantly or minimally co-expressed, expression values that were increased or reduced by twofold as compared to the average expression levels were determined as the cut-off values, respectively. As is shown in Fig. 7, all clusters evidenced high or low peaks representing twofold the average expression values, indicating the different expression patterns of co-expressing genes included in each cluster by specific tissue and/or stress-response. For example, the co-expressing genes in cluster 38 evidenced high expression peaks during the seed development stages at 3–29 days after pollination. By way of contrast, 22 genes grouped into cluster 1019 are likely to be highly expressed in the roots, crowns, and growing points. In clusters 918 and 1174, similar expression patterns were detected in the roots, crown, and growing point tissues, reproductive organs, and seed development stages, but different in the leaves; cluster 1174 evidenced a relatively low expression pattern in the leaves as compared to those observed with cluster 918. Similarly, slight differences in stem apical meristems and mature ovary tissues were detected between clusters 1174 and 1144.

Curiously, the functional categories of the co-expressing genes with each of the nsLTP genes are likely to be somewhat specialized in the direction of specific biological processes. For example, cluster 38 evidenced a conspicuous expression profile peak in the target samples of seed developmental stages, probably indicating the roles of numbers

of genes together with an nsLTP gene (Os11g24070) during seed development. As had been expected, the functional categories of the numbers of genes included in this cluster appeared to be associated with seed development, e.g. glutelin type-A 1 precursor, protease inhibitor/seed storage/LTP family protein, alpha-amylase/subtilisin inhibitor, trypsin/factor XIIA inhibitor precursor, prolamine, etc. (data not shown).

Discussion

The Poaceae family, which includes the cereal species, appears to have diverged from an ancient ancestor approximately 50–70 Mya following whole genome duplication (Paterson et al. 2004). Afterward, differential gene losses and subsequent independent whole or segmental duplication resulted in the variety of genome shapes of the cereal species. For example, the complexity of the wheat genome originated from two instances of allopolyploidization, resulting in a hexaploid possessing 42 chromosomes ($2n = 6X = 42$, AABBDD). Similarly, the maize genome appears to have undergone recent genome duplication after tetraploidization, about 4.8 Mya (Swigonova et al. 2004). In another case, segmental duplication occurred in the distal portions of three-Mbps of rice chromosomes 11 and 12 approximately 7.7 Mya (The rice chromosomes 11 and 12 Sequencing Consortia, 2005). These independent duplications occurring during individual speciation may result in some increase in copies of the Poaceae nsLTP genes. In our previous report, we detected 7 of 12 wheat nsLTP genes, which existed in a homoeologous relationship, in addition to the rice 12 genes on the nsLTP-rich regions, which originated from a segmental duplication occurring after tandem duplications (Jang et al. 2007). As shown in Fig. 1, phylogenetic analysis may have provided some evidence that the Poaceae nsLTP genes may have evolved rapidly with independent duplications after at least Poaceae-*Allium* divergence. Although we recognized certain hazards associated with a lack of *Allium* ESTs, the closer relationship between the nsLTPs of *Allium* and *Arabidopsis* as compared to those of *Allium* and Poaceae may support our hypothesis.

Polyploidy is believed to cause extensive changes in gene expression. A predominant example of rapid subfunctionalization inferred from polyploidy was proposed by Adams et al. (2003), who reported that at least 25% homoeologous genes evidenced silencing or unequal expression in the ovules of *Gossypium hirsutum*. The divergence and functional diversification of *TaLTP* genes may have occurred after genome polyploidization. Considering the *TaLTP* genes' sequence similarities and chromosomal locations, some genes appear to have been inferred from polyploidization events, e.g. *TaLTP1* and 2, *TaLTP* 7, 8 and 10,

Table 3 Numbers of co-expressing genes with each nsLTP genes and microarray target samples with high and low expression

Cluster ID	nsLTP spot ID	Group	Clade	No. co-expressing genes	Microarray target samples with high expression	Microarray target samples with low expression
38	Os.37890.1.S1_at_Os11g24070 Os.37890.1.S1_s_at_Os11g24070	A	I	36	Mature anther, seed 3–5, 5–10, 11–20, and 21–29 days, seed 5 day, endosperm 10 days, embryo 10 days	
212	Os.37860.1.S1_at_Os11g02330 Os.28462.1.S1_s_at_Os12g02290	B	I	60	Leaf coleoptiles_an aerobic stress, Mature ovary, suspension cell, inflorescences ~3, 5–10, and 10–15 cm, seed 0–2 days, seed 5 days, embryo 10 days	Root 7 day old, third leaves 2 week_tZ, mature rice plant
1562	Os.10120.2.S1_at_Os11g02424	B	I	61	Leaf_T0 empty vector control 6 weeks, leaf_T0 OsRR6 vector overexpressor 6 weeks, shoot, third leaves 2 weeks_tZ, third leaves 2 weeks_DMSO, leaf 11 days, mature rice plant, leaf 11 days, leaf 11 days_OsSRT1 RNAi transgenic line, young leaf mature rice plants, leaf 7 day_salt stress, seed 0, 2 days	Crown and growing point, crown and growing point_salt stress, mature stigmas, SAM_upto 0.7 mm, mature ovary, suspension cell, inflorescence 3, 5 cm
1620	Os.8629.1.S1_x_at_Os12g02320 Os.26565.1.S1_a_at_Os11g02400	A B	I	56	Mature stigma, leaf_T0 empty vector control 6 weeks, leaf_T0 OsRR6 vector overexpressor 6 weeks, shoot, third leaves 2 weeks_tZ, third leaves 2 weeks_DMSO, leaf 11 days, leaf 11 days_OsSRT1 RNAi transgenic line	Hydroponic roots 1 week_Arsenate inflorescences 3–5 cm
1571	Os.36779.1.S1_s_at_Os12g02330	B	I	460	Leaf_T0 empty vector control 6 weeks, leaf_T0 OsRR6 vector overexpressor 6 weeks, shoot, third leaves 2 weeks_tZ, third leaves 2 weeks_DMSO, leaf mature rice plant, leaf 11 days, leaf 11 days_OsSRT1 RNAi transgenic line, seed 5–10 days after pollination	Root 7 day old, inflorescences 3–5 cm
83	Os.16271.1.S1_s_at_Os05g40010 OsAffx.27248.1.S1_s_at_Os05g40010 OsAffx.27248.1.S1_x_at_Os05g40010	D	I	13	Leaf_T0 empty vector control 6 weeks, leaf_T0 OsRR6 vector overexpressor 6 weeks, shoot, leaf 7 days_salt stress, endosperm 10 days, seed 11–20 days	Leaf coleoptiles_an aerobic stress, root_DMSO, root_tZ, SAM 0.5–0.7, suspension cell
918	Os.10615.1.S1_x_at_Os12g02310	A	II	54	Crown and growing point_salt stressed, crown and growing point, shoot, leaf 11 days, leaf 11 days_OsSRT1 RNAi transgenic line, leaf 7 days	Leaf coleoptiles, leaf coleoptiles_an aerobic stress, hydroponic roots 1 week, hydroponic roots 1 week_Arsenate, root 2 weeks_DMSO, root 2 weeks_tZ, mature stigmas, endosperm 10 day old, embryo 10 day old
1174	Os.4380.1.S1_at_Os11g02350	C	II	76	Crown and growing point_salt stressed, crown and growing point, inflorescence 5–10, 10–15, and 15–22 cm, mature ovary, leaf_T0 OsRR6 vector overexpressor 6 weeks, leaf 7 days_cold, leaf 7 days	Hydroponic roots 1 week_Arsenate, root 2 weeks_DMSO, root 2 weeks_tZ, mature stigmas, suspension cell, third leaves 2 weeks_tZ, third leaves 2 weeks_DMSO, leaf mature rice plants, leaf 11 days, leaf 11 days_OsSRT1 RNAi transgenic line, young leaf mature rice plants, endosperm 10 day old, seed 11–20 and 21–29 days

Table 3 continued

Cluster ID	nsLTP spot ID	Group	Clade	No. co-expressing genes	Microarray target samples with high expression	Microarray target samples with low expression
1144	Os.8827.1.S1_at_Os01g60740	D	II	225	Crown and growing point_salt stressed, crown and growing point, SAM, up to 0.7 mm, mature ovary, suspension cell, young inflorescence up to 3 cm, embryo 10 day old	Shoot, third leaves 2 weeks_tZ, third leaves 2 weeks_DMSO, leaf mature rice plant, leaf 11 days_OsSRT1 RNAi transgenic line, young mature rice plants
1164	Os.51480.1.S1_at_Os01g62980 Os.51480.1.S1_x_at_Os01g62980	E	II	9	Leaf coleoptiles 4 days Crown and growing point_salt stressed, crown and growing point Leaf_T0 empty vector control 6 weeks, Leaf_T0 OsRR6 vector overexpressor 6 weeks, leaf 7 days_cold, drought, salt and unstressed	Hydroponic roots 1 week, root 2 weeks_DMSO, hydroponic roots 1 week_Arsenate, root 2 weeks_tZ, mature stigmas, mature ovary, suspension cell, third leaves 2 weeks_tZ, third leaves 2 weeks_DMSO, leaf mature rice plants, young mature rice plants, seed 5, 5–10 days, embryo 10 days, endosperm 10 days, seed 11–20 days
1019	Os.12266.1.S1_at_Os05g06780	E	II	22	Hydroponic roots 1 week, root 7 day old, root 2 weeks_DMSO, hydroponic roots 1 week_Arsenate, root 2 weeks_tZ, crown and growing point_salt stressed, crown and growing point, inflorescence 15–22 cm	Mature stigmas, mature ovary, suspension cell, young inflorescence upto 3 cm, inflorescence 3–5 cm, third leaves 2 weeks_tZ, third leaves 2 weeks_DMSO, leaf mature rice plants, leaf 11 days, leaf 11 days_OsSRT1 RNAi transgenic line, seed 5 days, embryo 10 days, endosperm 10 days, seed 11–20 and 21–29 days
1077	OsAffx.26317.1.S2_at_Os04g33920	E	II	56	Embryo 10 day old	Third leaves 2 weeks_tZ, third leaves 2 weeks_DMSO, leaf mature rice plants, leaf 11 days, leaf 11 days_OsSRT1 RNAi transgenic line

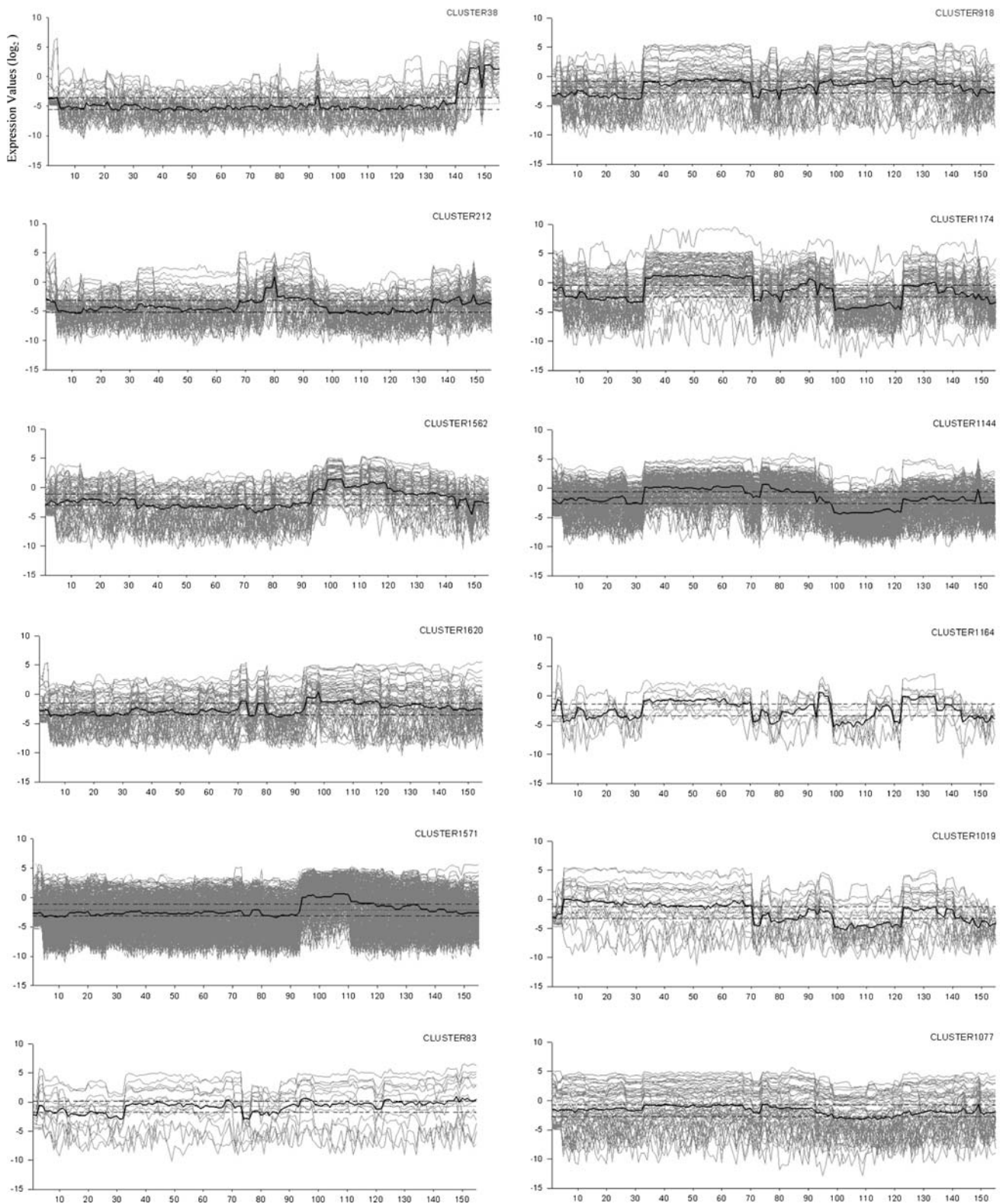


Fig. 7 Expression profiles of each of the co-expressing genes with individual nsLTP gene. For descriptions in detail see the legend for Fig. 5

and *TaLTP 6* and *11*. Relative slight differences in transcriptomes were detected between homoeologous *TaLTP* genes via plant elicitor treatments. However, expressions

between genes, including those of the other groups, evidenced significant differences when comparing the transcriptional patterns of *TaLTP1* for group A, *TaLTP3* for

group B, and *TaLTP5* for group D. Boutrot et al. (2005) elucidated the complexity of the wheat nsLTP gene family, particularly, in the stages of seed development. Recently, the same group further assessed the comparative promoter activity of certain *TaLTP* genes in transgenic rice, thereby suggesting that the TaLTPs evidence distinct functions in planta (Boutrot et al. 2007). Our present results supported the notion of the transcriptional diversity of wheat nsLTP genes inferred from polyploidization after gene duplication.

The majority of duplicate genes resulting from genome duplication and/or gene amplification tend to lose one of either copy to avoid functional redundancy. However, in the case of some retained duplicate genes, different changes in their protein sequences together with *cis*-regulatory elements may result in a diversification of their functions or alterations in the timing, location, and relative number of their transcripts (Moore and Purugganan 2005). Considering the copy numbers of nsLTP genes in plant taxa, they may have undergone independent dynamic evolutionary processes during speciation. Curiously, we detected similar selection modes in the protein sequences of paralogous gene pairs in both wheat and rice species, although they appear to have undergone independent duplications and evolutionary processes after diverging. For example, the nsLTP genes included in group A of both species evidenced relatively slow evolution, whereas the genes within other groups, e.g. groups B and D of both species, are likely to have undergone somewhat more rapid evolution. The overlapping of expression patterns in the *TaLTP* genes of group A, with the exception of *TaLTP4*, may be reflected in the relatively low *Ka/Ks* value of 0.510. We detected no *TaLTP4* transcripts in the leaves, although they could be amplified from genomic DNA (Jang et al. 2007). Considering the seed-specific expression profile of the unigene corresponding to *TaLTP4*, suggested by the analysis of EST counts in the NCBI database (data not shown) together with the relatively high sequence substitutions of *TaLTP4* among the group A genes, these features appear to be an example of the rapid evolution of duplicate genes.

Degenerative mutations in the *cis*-regulatory elements may influence the divergence in their expression, facilitating the preservation of both duplicates (Force et al. 1999). Papp et al. (2003) previously determined via genomic analyses of the yeasts that the number of shared regulatory motifs in duplicates decreases with evolutionary time. In plants, conserved noncoding sequences, which are putative regulatory elements between orthologues, are smaller and far less frequent than those identified in mammalian genes, thereby suggesting that they should prove useful as monitors of the regulatory complexity of genes (Kaplinisky et al. 2002). Our phylogenetic footprinting analysis indicated functional redundancy as well as diversity among the nsLTP genes over evolutionary time. For example, the putative

eight promoters of the nsLTP genes of wheat and rice included in group A harbored one conserved CATGCA motif (but this was mutated in *TaLTP4*), corresponding to the RY repeat detected in the RY/G box of the *napA* gene in *Brassica napus* (Ezcurra et al. 1999); This box is known to be required for the seed-specific expression associated with the ABA response. As shown in Fig. 3, the *TaLTP* genes of group A, with the exception of *TaLTP4*, were activated by ABA treatment. These findings indicated some degree of long-term conservation in transcriptional regulation after nsLTP gene duplication followed by speciation. Conversely, many lost or mutated regulatory motifs detected might be reflected in ongoing functional diversity via evolutionary mechanisms, including subfunctionalization.

An enormous amount of information, including public microarray databases and large-scale EST analyses, together with complete genome sequencing in rice, enables us to conduct more detailed functional diversity analyses of post-duplication rice nsLTP genes. Rice genes on the nsLTP-rich regions, which were inferred from a recent segmental duplication, may have undergone evolutionary dynamics in order to avoid functional redundancy. We detected no evidence of pseudogenization of the nsLTP genes on the nsLTP-rich regions, on the basis of a set of unigenes corresponding to all nsLTP genes (data not shown), thereby suggesting that segmental duplication may result in a dynamic evolutionary fate, such as subfunctionalization or neofunctionalization (Moore and Purugganan 2005). Subfunctionalization is generally explained in that the preservation of duplicate genes can be ascribed to the partitioning of the functions of an ancestral gene (Force et al. 1999). A question has been raised as to whether subfunctionalization is a primary force in the evolutionary fate of nsLTP genes. If the evolution of nsLTP genes is dependent on the mode of subfunctionalization, the expression profiles of the nsLTP genes should evidence greater similarity between those in the same group than between groups. A case of two duplicate genes grouped into cluster 212 pointed to redundant (but not identical) expression between duplicate genes inferred from segmental duplication. In other cases, each of the genes from segmental duplication followed by tandem duplication were grouped into the other's cluster, thereby indicating the ongoing nature of the process of their expression divergence, e.g. Os11g02400 versus Os12g02330 and Os12g02310 versus Os12g02320. However, some cases evidenced expression profiles relatively similar to those of other groups. For example, the expression profile of Os12g02320 of group A closely resembled that of Os11g02400 of group B, but featured significant differences from that of Os11g24070 of group A. Three expression profiles—those of Os12g02310, Os11g11g02350, and Os01g60740—appeared to be

extremely similar to each other, even if they were clustered into three other groups (A, C, and D, respectively). These findings may support the notion that subfunctionalization is not the sole mechanism inherent to the evolutionary history of nsLTP genes, which was suggested by Papp et al. (2003), but rather that this occurs in concert with other mechanism(s). Recently, Wang et al. (2007) suggested an attractive hypothesis, namely that the extensive concerted evolution of rice paralogues and the road to regaining independence have been attributed by whole- and partial-gene conversion and several reciprocal exchanges of chromosomal segments, particularly in a single recent segmental duplication on chromosomes 11 and 12. The concerted evolution contributed by the partial-gene conversion between paralogues might explain, at least in part, our finding that high similarities exist in expression patterns among nsLTP genes including different groups, although further detailed analyses will be required to further confirm this.

It has been recently proposed that plant nsLTP genes evidence partitioned organ-specific and developmental expression (Kader 1996). In a host of nsLTP studies, their expressions have been demonstrated to be predominantly exhibited in the aerial portions, including the leaves, stems, shoot meristems, and reproductive tissues, but are rare in the roots. This may suggest a common feature of plant nsLTP genes (Gausling 1994; Vignols et al. 1994, 1997; Arondel et al. 2000; Jang et al. 2002). However, Os05g06780 evidenced relatively high expression profiles in unstressed- and stressed roots, whereas Os11g24070 evidenced a gradual increase in its transcripts during the stages of seed development. Significant differences in expression patterns in the leaves between nsLTP genes were uncovered, of which some, including Os11g02330 and Os12g02290, were high, whereas others, including Os11g02400 and Os12g02330, were low. However, these genes were clustered into the same group (B) as the result of phylogenetic analysis. Therefore, the different expression profiles of 14 rice nsLTP genes over a wide range of tissues and stresses supported the notion that the degree of specialization of individual nsLTPs is higher than was previously expected, although they may still share overlapping redundant functions, further supporting the notion that nsLTP genes do not harbor single consensus in association with their physiological roles (Carvalho Ade and Gomes 2007).

Insight into gene-coexpression networks may provide some clues into biological functions that require the coordinated expression of a large number of genes (Stuart et al. 2003). Interestingly, Blanc and Wolfe (2004b) reported several cases in which groups of *Arabidopsis* duplicate gene pairs diverged in concert, forming two parallel networks, thereby indicating that this process involves the concerted divergence of gene expression. Our clustering of genes that are co-expressed with individual nsLTP genes

(e.g. cluster 38) revealed some degree of specialization of a large fraction of genes. It appears likely that the functional diversity of nsLTP genes is not randomized, but specialized in the direction of specific biological processes occurring over evolutionary time.

In the previous report, we suggested that nsLTP genes have undergone independent duplication(s) followed by functional diversity in each species, which occurred during speciation (Jang et al. 2007). Also, the findings described herein might further extend our knowledge regarding the evolutionary fate of nsLTP genes following duplication and speciation in higher plants. Collectively, our results, including our nsLTP gene-coexpression networks together with pending genome sequencing and microarray data of other species, e.g. wheat and barley, underline the necessity for further studies into functional analyses of the nsLTP genes.

Acknowledgments This work was supported by a Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2005-050-F00001) to CSJ.

References

- Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Natl Acad Sci USA* 100:4649–4654
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Arondel VV, Vergnolle C, Cantrel C, Kader J (2000) Lipid transfer proteins are encoded by a small multigene family in *Arabidopsis thaliana*. *Plant Sci* 157:1–12
- Blanc G, Wolfe KH (2004a) Wide spread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16:1667–1678
- Blanc G, Wolfe KH (2004b) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16:1679–1691
- Blanchette M, Tompa M (2003) FootPrinter: a program designed for phylogenetic footprinting. *Nucleic Acids Res* 31:3840–3842
- Boutrot F, Guirao A, Alary R, Joudrier P, Gautier MF (2005) Wheat non-specific lipid transfer protein genes display a complex pattern of expression in developing seeds. *Biochim Biophys Acta* 1730:114–125
- Boutrot F, Meynard D, Guiderdoni E, Joudrier P, Gautier MF (2007) The *Triticum aestivum* non-specific lipid transfer protein (*TaLtp*) gene family: comparative promoter activity of six *TaLtp* genes in transgenic rice. *Planta* 225:843–862
- Carvalho Ade O, Gomes VM (2007) Role of plant lipid transfer proteins in plant cell physiology-a concise review. *Peptides* 28:1144–1153
- de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* 20:1453–1454
- 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
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545

- Gausling K (1994) Lipid transfer protein genes specifically expressed in barley leaves and coleoptiles. *Planta* 192:574–580
- Green PJ, Yong MH, Cuozzo M, Kano-Murakami Y, Silverstein P, Chua NH (1988) Binding site requirements for pea nuclear protein factor GT-1 correlate with sequences required for light-dependent transcriptional activation of the *rbcS-3A* gene. *EMBO J* 7:4035–4044
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
- Herrero J, Valencia A, Dopazo J (2001) A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics* 17:126–136
- Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant *cis*-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27:297–300
- Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9:868–877
- Jang CS, Kim DS, Bu SY, Kim JB, Lee SS, Kim JY, Johnson JW, Seo YW (2002) Isolation and characterization of lipid transfer protein (LTP) genes from a wheat-rye translocation line. *Plant Cell Rep* 20:961–966
- Jang CS, Lee HJ, Chang SJ, Seo YW (2004) Expression and promoter analysis of the *TaLTP1* gene induced by drought and salt stress in wheat (*Triticum aestivum* L.). *Plant Sci* 167:995–1001
- Jang CS, Johnson JW, Seo YW (2005) Differential expression of *TaLTP3* and *TaCOMT1* induced by Hessian fly larval infestation in a wheat line possessing *H21* resistance gene. *Plant Sci* 168:1319–1326
- Jang CS, Jung JH, Yim WC, Lee B-M, Seo YW, Kim W (2007) Divergence of genes encoding non-specific lipid transfer proteins in the Poaceae family. *Mol Cells* 24:215–223
- Kader JC (1996) Lipid-transfer proteins in plants. *Annu Rev Plant Physiol Plant Mol Biol* 47:627–654
- Kaplinsky NJ, Braun DM, Penterman J, Goff SA, Freeling M (2002) Utility and distribution of conserved noncoding sequences in the grasses. *Proc Natl Acad Sci USA* 99:6147–6151
- Kramer EM, Jaramillo MA, Di Stilio VS (2004) Patterns of gene duplication and functional evolution during the diversification of the *AGAMOUS* subfamily of MADS box genes in angiosperms. *Genetics* 166:1011–1023
- Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163
- Lee TG, Jang CS, Kim JY, Kim DS, Park JH, Kim DY, Seo YW (2007) A Myb transcription factor (*TaMyb1*) from wheat roots is expressed during hypoxia: roles in response to the oxygen concentration in root environment and abiotic stresses. *Physiol Plantarum* 129:375–385
- Lockton S, Gaut BS (2005) Plant conserved non-coding sequences and paralogue evolution. *Trends Genet* 21:60–65
- 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
- Moore RC, Purugganan MD (2005) The evolutionary dynamics of plant duplicate genes. *Curr Opin Plant Biol* 8:122–128
- Nagao RT, Goekjian VH, Hong JC, Key JL (1993) Identification of protein-binding DNA sequences in an auxin-regulated gene of soybean. *Plant Mol Biol* 21:1147–1162
- Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York, p 52
- Papp B, Pal C, Hurst LD (2003) Evolution of *cis*-regulatory elements in duplicated genes of yeast. *Trends Genet* 19:417–422
- Park HC, Kim ML, Kang YH, Jeon JM, Yoo JH, Kim MC, Park CY, Jeong JC, Moon BC, Lee JH (2004) Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor 1. *Plant Physiol* 135:2150–2161
- Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci USA* 101:9903–9908
- Qi LL, Echalié B, Chao S, Lazo GR, Butler GE, Anderson OD, Akhunov ED, Dvorak J, Linkiewicz AM, Ratnasiri A, Dubcovsky J, Bermudez-Kandianis CE, Greene RA, Kantety R, La Rota CM, Munkvold JD, Sorrells SF, Sorrells ME, Dilbirligi M, Sidhu D, Erayman M, Randhawa HS, Sandhu D, Bondareva SN, Gill KS, Mahmoud AA, Ma XF, Miftahudin, Gustafson JP, Conley EJ, Nduati V, Gonzalez-Hernandez JL, Anderson JA, Peng JH, Lapitan NL, Hossain KG, Kalavacharla V, Kianian SF, Pathan MS, Zhang DS, Nguyen HT, Choi DW, Fenton RD, Close TJ, McGuire PE, Qualset CO, Gill BS (2004) A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. *Genetics* 168:701–712
- Rice chromosomes 11, 12 sequencing consortia (2005) The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. *BMC Biol* 3:20. doi:10.1186/1741-7007-3-20
- Rubio V, Linhares F, Solano R, Martin AC, Iglesias J, Leyva A, Paz-Ares J (2001) A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. *Gene Dev* 15:2122–2133
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Seo YW, Johnson JW, Jarret RL (1997) A molecular marker associated with the *H21* Hessian fly resistance gene in wheat. *Mol Breed* 3:177–181
- Shirsat A, Wilford N, Croy R, Boulter D (1989) Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. *Mol Gen Genet* 215:326–331
- Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302:249–255
- Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004) Close split of sorghum and maize genome progenitors. *Genome Res* 14:1916–1923
- Vignols F, Lund G, Pammi S, Tremousaygue D, Grellet F, Kader JC, Puigdomenech P, Delseny M (1994) Characterization of a rice gene coding for a lipid transfer protein. *Gene* 142:265–270
- Vignols F, Wigger M, Garcia-Garrido JM, Grellet F, Kader JC, Delseny M (1997) Rice lipid transfer protein (LTP) genes belong to a complex multigene family and are differentially regulated. *Gene* 195:177–186
- Villain P, Mache R, Zhou DX (1996) The mechanism of GT element-mediated cell type-specific transcriptional control. *J Biol Chem* 271:32593–32598
- Wang X, Tang H, Bowers JE, Feltus FA, Paterson AH (2007) Extensive concerted evolution of rice paralogs and the road to regaining independence. *Genetics* 177:1753–1763
- Zahn LM, Leebens-Mack JH, Arrington JM, Hu Y, Landherr LL, dePamphilis CW, Becker A, Theissen G, Ma H (2006) Conservation and divergence in the *AGAMOUS* subfamily of MADS-box genes: evidence of independent sub- and neofunctionalization events. *Evol Dev* 8:30–45
- Zhu Q, Ordiz MI, Dabi T, Beachy RN, Lamb C (2002) Rice TATA binding protein interacts functionally with transcription factor IIB and the RF2a bZIP transcriptional activator in an enhanced plant in vitro transcription system. *Plant Cell* 14:795–803