

Co-expression of soybean Dicer-like genes in response to stress and development

Shaun J. Curtin · Michael B. Kantar · Han W. Yoon ·
Adam M. Whaley · Jessica A. Schlueter ·
Robert M. Stupar

Received: 15 November 2011 / Revised: 15 March 2012 / Accepted: 21 March 2012 / Published online: 15 April 2012
© Springer-Verlag 2012

Abstract Regulation of gene transcription and post-transcriptional processes is critical for proper development, genome integrity, and stress responses in plants. Many genes involved in the key processes of transcriptional and post-transcriptional regulation have been well studied in model diploid organisms. However, gene and genome duplication may alter the function of the genes involved in these processes. To address this question, we assayed the stress-induced transcription patterns of duplicated gene pairs involved in RNAi and DNA methylation processes in the paleopolyploid soybean. Real-time quantitative PCR and Sequenom MassARRAY expression assays were used to profile the relative expression ratios of eight gene pairs across eight different biotic and abiotic stress conditions. The transcriptional responses to stress for genes involved in DNA methylation, RNAi processing, and miRNA processing were compared. The strongest evidence for pairwise co-expression in response to stresses was exhibited by non-paralogous Dicer-like (DCL) genes GmDCL2a-GmDCL3a and GmDCL1b-GmDCL2b, most profoundly in root tissues. Among homoeologous or paralogous DCL genes, the Dicer-like 2 (DCL2) gene pair exhibited the strongest response to stress and most conserved co-expression pattern.

Electronic supplementary material The online version of this article (doi:10.1007/s10142-012-0278-z) contains supplementary material, which is available to authorized users.

S. J. Curtin · M. B. Kantar · H. W. Yoon · R. M. Stupar (✉)
Department of Agronomy and Plant Genetics,
University of Minnesota,
St. Paul, MN 55108, USA
e-mail: rstupar@umn.edu

A. M. Whaley · J. A. Schlueter
Department of Bioinformatics and Genomics,
University of North Carolina,
Charlotte, NC 28223, USA

This was surprising because the DCL2 duplication event is more ancient than the other DCL duplications. Possible mechanisms that may be driving the DCL2 co-expression are discussed.

Keywords Soybean · Paleopolyploid · RNAi · Dicer-like · Stress

Introduction

Polyploidy, the state of having more than two sets of chromosomes, is very common in plants and there is extensive evidence for whole-genome duplication events in basal angiosperm lineages (Soltis et al. 2009). Species experiencing relatively recent genome duplications are defined as polyploids per se, whereas species with more ancient duplications are typically defined as paleopolyploids or diploids because their chromosome sets have differentiated such that they no longer pair and/or resemble one another.

Polyploid studies frequently focus on gene redundancies and the divergence of duplicated gene copies. The fate of duplicated genes over evolutionary time is typically divided into three categories: non-functionalization, neo-functionalization, and sub-functionalization (Force et al. 1999; Prince and Pickett 2002). Increasingly, studies of gene duplication focus on gene expression data (Adams 2007; Jackson and Chen 2010) as divergence in gene expression profiles may indicate a divergence in duplicate gene function. A recent study in *Arabidopsis thaliana* supports this idea, reporting that highly co-expressed duplicate gene pairs shared more similar protein–protein interaction profiles than less co-expressed pairs (Arabidopsis Interactome Mapping Consortium 2011). Studies have focused on duplicate gene transcript partitioning as a consequence of plant developmental differentiation (Buggs et al. 2011; Chaudhary et

al. 2009; Hovav et al. 2008; Nomura et al. 2005) or stress conditions (Dong and Adams 2011; Liu and Adams 2007; Stamati et al. 2009) in a wide range of natural and/or synthetic plant polyploids. Recent investigations have incorporated structural aspects into the analysis by focusing on transcript partitioning between genes located within duplicated linkage blocks (Flagel et al. 2009; Lin et al. 2010).

Soybean is a sequenced paleopolyploid genome that maintains at least one gene duplicate for ~75 % of its genes (Schlueter et al. 2007; Schmutz et al. 2010). The most recent genome doubling event occurred approximately 9–14 million years ago (Blanc and Wolfe 2004; Schlueter et al. 2004; Schmutz et al. 2010). A high proportion of the duplicated soybean genes resulted from the most recent genome duplication event. These gene pairs are located within syntenic chromosomal regions and are termed homoeologous gene pairs. A smaller proportion of soybean gene duplicates are arranged in tandem or are located within non-syntenic regions; these would be considered non-homoeologous paralogs.

RNA interference (RNAi) and DNA cytosine methylation are epigenetic processes that regulate gene expression and silencing. RNAi processes are governed by the activity of paralogous RNaseIII Dicer-like (DCL) genes that encode endonuclease proteins that process double-stranded RNA (dsRNA) into small RNAs (sRNAs). Each DCL has a specialized function that has been well characterized in *Arabidopsis* and other model organisms (Bouche et al. 2006; Eamens et al. 2008b; Margis et al. 2006). The DCL family of proteins in *Arabidopsis* has four canonical DCLs (AtDcl1, AtDcl2, AtDcl3, and AtDcl4) which control the expression of developmentally regulated genes, repression of mobile DNA elements, and defense against viral infection by generating a variety of sRNAs, including micro (miRNAs), natural-anti-sense (nat-siRNAs), repeat-associated (rasiRNAs), trans-acting (tasiRNAs), and viral small (vsRNAs) (Margis et al. 2006). DCL1 is the enzyme responsible for the processing and maturation of miRNAs in *Arabidopsis*. miRNAs are single-stranded 21-nt RNA molecules derived from partially complementary stem loop precursor structures transcribed from host genes that control gene expression (Eamens et al. 2008b). DCL2 is required for the processing of nat-siRNAs generated from two overlapping RNA transcripts in *cis*-antisense orientation (Katiyar-Agarwal et al. 2006) and the transitive silencing of transgenes (Mlotshwa et al. 2008). DCL3 is one of several components of the RNA-directed DNA methylation pathway (RdDM) and is responsible for processing 24-nt rasiRNAs from endogenous repeat sequences and transposons. DNA methylation of repetitive and transposon sequences suppresses their aberrant expression, thereby maintaining genome stability (Chan et al. 2005). DCL4 sequentially processes tasiRNAs from specific miRNA-

targeted transcripts that convert into double-stranded RNA by RNA-directed RNA polymerase 6 (RDR6). These tasiRNAs negatively regulate various transcripts involved in organ development and vegetative phase changes in the plant (Allen et al. 2005).

RNAi and DNA methylation have a well-established association and have been reported to be influenced and altered by stress conditions in *Arabidopsis* (Ben Amor et al. 2009; Borsani et al. 2005; Boyko et al. 2010; Navarro et al. 2008), rice (Yan et al. 2011), and *Medicago truncatula* (Capitao et al. 2011). Specific roles in stress response have been identified for DCL2 and DCL3 in *Arabidopsis* (Borsani et al. 2005; Boyko et al. 2010; Brosnan et al. 2007; Eamens et al. 2008a; Yan et al. 2011). In the paleopolyploid soybean, widespread gene duplication adds an additional layer of complexity to defining the roles of the genes that govern these processes.

In this study, we profiled the transcriptional responses to stress of eight soybean duplicated gene pairs and one non-duplicated gene (DCL3) known to be involved in epigenetic processes, particularly RNAi and DNA methylation. The RNAi genes include seven canonical soybean DCL genes and homoeologous pairs of ARGONAUTE1 (AGO1), RDR6, and the double-stranded RNA binding protein DRB1 (Eamens et al. 2009; Margis et al. 2006; Vaucheret 2008; Wassenegger and Krczal 2006), respectively. The cytosine–DNA–methyltransferase genes include a homoeologous gene pair with homology to DNA METHYLTRANSFERASE1 (MET1) (Finnegan and Dennis 1993) and a paralogous pair with homology to DOMAINS REARRANGED METHYLASE1 (DRM1) (Cao and Jacobsen 2002). The transcript analysis was conducted across eight different stress treatments and three different soybean genotypes in an attempt to define co-expression patterns between duplicate genes and identify unique transcriptional responses to stress.

Materials and methods

Identification of predicted RNAi pathway and methylation genes in soybean

The predicted soybean homologs for several gene families, such as the DCL, AGO, and methyltransferase families, were obtained from The Arabidopsis Information Resource (<http://www.arabidopsis.org/>) and the Rice Genome Annotation Project (<http://rice.plantbiology.msu.edu/>). The resulting amino acid sequences were used to query the soybean sequenced genome databases (<http://www.phytozome.org/>). Most genes of interest had more than one copy. We chose to focus on a subset of genes that clearly showed two duplicate, intact copies: GmDCL1, GmDCL2, GmDCL4, GmAGO1, GmDRB1, GmDRM1, GmMET1, and GmRDR6. The

soybean homolog gene models are shown in Table 1, all renamed with “Gm” prefixes to denote the species *Glycine max* and generic “a/b” annotations to specify the two duplicate copies. GmDCL3, which displayed one intact copy, was also included in the downstream transcriptional analysis. Thus, all seven complete soybean DCL genes were included in this part of the study.

The positions of duplicated blocks for soybean were taken from published data (Schmutz et al. 2010) and the location of individual genes of interest within those blocks were visualized using Circos (Krzywinski et al. 2009). Evolutionary distances between duplicate genes were determined using the gene model nucleotide sequences. Duplicate gene coding regions were aligned using the Smith–Waterman pairwise alignments algorithm. Synonymous (Ks) and nonsynonymous (Ka) changes between the duplicated sequences were determined using PAML (Yang 1997; Yang 2007). To determine the age of the duplications, a molecular clock was assumed and dating was determined as previously described (Schlueter et al. 2004).

Plant materials, growth, and nucleic acid extraction

Soybean cultivars Williams 82, Archer, and Noir 1 seeds were obtained from Dr. James Orf at the University of Minnesota. The plants were grown in 50:50 soil and vermiculite mix and maintained under standard growth chamber conditions (22–25 °C, 16-h photoperiod at 150–200 $\mu\text{mol m}^{-2} \text{s}^{-1}$).

Abiotic stress experiments were performed after 14 days of growth. Seedlings were gently uprooted, with soil material removed and incubated in their respective treatments for 3 h, then flash-frozen for RNA extraction. Salt-stressed seedlings were incubated in 200 mM of sodium chloride (NaCl). Cold-stressed plants were incubated at 4 °C distilled water in a well-lit walk-in cold room. Drought stress was carried out by incubation of the plant root system between two pieces of 3-mm Whatman filter paper. Pathogen response stress was simulated by incubation in 1 mM salicylic acid (SA) solution. Mock control seedlings were incubated in dH₂O. All treatments were carried out in growth chamber conditions with the exception of the cold treatment (described above). Triplicate samples of mock and treated 14-day-old tissue encompassing uni-foliolate and trifoliolate leaves, stems, and roots were each harvested separately at approximately six h after light onset and immediately flash-frozen in liquid nitrogen.

An isolate of the oomycete soybean pathogen *Phytophthora sojae* Race28 (Ps28) (supplied by Dr. Dean Malvick, University of Minnesota) was maintained by weekly sub-culture on V8 agar. A 1-cm vertical incision was made on the hypocotyl 2–3 mm below the cotyledons on soybean cultivars cvs. Williams 82 (susceptible), Archer (resistant), and Noir 1 (susceptible). Agar infected with *P. sojae* was inserted into the incision and the wound site wrapped in parafilm to protect against desiccation and unrelated infection. Mock controls were carried out using sterile V8 agar (Kachroo et al. 2008). Mock and *P. sojae*-infected

Table 1 Location and conservation of soybean gene duplicates in this study

<i>Arabidopsis</i> gene	Soybean homologs	Gene model	Ka	Ks	Ka/ Ks	<i>P</i> -value (Fisher)	Conserved homoeologs	Percentage conserved ^a (%)
AtAgo1 (At1g48410)	GmAGO1a	Glyma09g29720	0.01	0.144	0.072	2.16E-43	45 genes	70.6
	GmAGO1b	Glyma16g34300						63.2
AtDcl1 (At1g01040)	GmDCL1a	Glyma03g42290	0.032	0.039	0.821	0.180833	736 genes	57.5
	GmDCL1b	Glyma19g45060						58.8
AtDcl2 (At3g03300)	GmDCL2a	Glyma09g02930	0.085	0.252	0.339	3.38E-29		
	GmDCL2b	Glyma09g02920						
AtDcl4 (At5g20320)	GmDCL4a	Glyma17g11240	0.027	0.091	0.301	1.17E-14	45 genes	69.2
	GmDCL4b	Glyma13g22450						64.3
AtDrb1 (At1g09700)	GmDRB1a	Glyma06g10200	0.027	0.09	0.303	7.36E-05	561 genes	61.9
	GmDRB1b	Glyma04g10230						60.6
AtDRM1 (At5g15380)	GmDRM1a	Glyma05g08740	0.012	0.116	0.104	5.77E-14		
	GmDRM1b	Glyma19g00250						
AtMet1 (At5g49160)	GmMET1a	Glyma06g18790	0.018	0.097	0.19	1.07E-24	471 genes	61.1
	GmMET1b	Glyma04g36150						56.6
AtRdr6 (At3g49500)	GmRDR6a	Glyma04g07150	0.01	0.109	0.091	9.15E-30	561 genes	60.6
	GmRDR6b	Glyma06g07250						61.9

^a The percentage of genes within each block that have a conserved gene in the homoeologous region

inoculated plants were harvested at 3- and 24-h time points. The soybean viral pathogen, soybean mosaic potyvirus (SMV), was obtained from Prof. Ben Lockhart, University of Minnesota. The cotyledons, stems, uni-foliolate, and trifoliolate leaves of 14-day-old Williams 82 (susceptible), Archer (susceptible), and Noir 1 (susceptible) were lightly dusted with carborundum and mechanically inoculated with ground SMV-infected plant material in 100 mM PO₄ (pH 7.5) 0.5 % mercapto-ethanol inoculation buffer. Mock treatments were carried out using the same buffer minus infected material. Mock and virus-inoculated plants were incubated in the growth chamber and harvested at 10 and 30 days past inoculation (dpi). All plants were harvested approximately 6 h after light onset and immediately flash-frozen in liquid nitrogen and stored at -80 °C until required. Virus infection was confirmed by electron microscopy and PCR using virus-specific primers.

For all experiments, three biological replicate plants were sampled for each treatment × tissue type × genotype sample. Total RNA was extracted using TRIzol (Invitrogen), chloroform-treated twice to remove unwanted protein and precipitated with an equal volume of iso-propanol. After centrifugation, the pellet was re-suspended in nuclease-free water and immediately DNase-treated, followed by purification using an RNA cleanup kit as per manufacturer's instruction (Qiagen). The RNA concentration and purity was measured by a Nanodrop spectrophotometer (Nanodrop Technologies) and its integrity was validated by agarose gel electrophoresis. DNA samples were collected from Archer, Noir 1, and Williams 82 leaf samples using the Qiagen DNeasy kit (Qiagen).

Transcription analysis with quantitative real-time PCR

A 3-μg aliquot of DNase-treated RNA from each RNA sample was reverse-transcribed using Superscript III for first-strand synthesis according to the manufacturer's instructions (Invitrogen). Quantitative real-time PCR was performed to estimate the transcriptional responses of the Dicer-like genes GmDCL1a, GmDCL1b, GmDCL2a, GmDCL2b, GmDCL3a, GmDCL4a, and GmDCL4b (gene model names are shown in Table 1). Primers were designed using primer3 (<http://frodo.wi.mit.edu/>). Manual adjustments were made for some primers to assure paralog-specific amplification. Real-time PCR data were collected from amplification plots and measured relative to the calculated ΔC_t value of endogenous actin gene ACT2/7 (Glyma19g32990) (Jian et al. 2008) using the $2^{-\Delta\Delta C_t}$ method (Pfaffl 2001). The primer sequences for the real-time PCR experiments are shown in Table S1 of the “Electronic supplementary material”.

A linear model was created using the qPCR expression data. The model was a nested ANOVA with abiotic stress

nested within genotype nested within developmental tissue. The transcriptional responses to stress of the GmDCL genes were compared with the unstressed control samples and significance was determined by Fisher's least significant difference test. Principal component analysis (PCA) was conducted by using every genotype, tissue, and abiotic stress combination as explanatory variables to describe the relationship between the GmDCL genes. ANOVA and PCA were conducted using the statistical software package R. PCA was visualized using the R biplot Gui (la Grange et al. 2009).

Relative transcription of gene duplicates

We used the Sequenom MassARRAY technology to quantify the transcript ratios of duplicated soybean genes encoding GmDCL1, GmDCL2, GmDCL4, GmAGO1, GmDRB1, GmDRM, GmMET1, and GmRDR6 (gene model names are shown in Table 1). The procedure was nearly identical to the method used in a previous study to determine the transcript ratios of 29 homoeologous genes on soybean chromosomes 8 and 15 (Lin et al. 2010). Briefly, SNPs were identified between the coding regions of the gene duplicate pairs using Phytozome (<http://www.phytozome.net/>) and Align Sequences Nucleotide BLAST.

DNA from leaf tissues of Archer, Noir 1, and Williams 82 were used as controls for MassARRAY assay quality. The cDNA samples from the abiotic and biotic stress experiments on the three soybean genotypes were assayed with MassARRAY for the three biological replicates for each stress × tissue × genotype sample. To quantify the duplicate transcript ratios, PCR and extension PCR reactions on the cDNA and DNA control templates were performed according to the manufacturer's specifications (Sequenom). To increase the reliability of the measurements, four technical replications were performed for each sample. In downstream analyses, the value used for each biological replicate was the mean of the four technical replicates. Mass spectrometry quantification of duplicate transcript ratios was performed at the University of Minnesota Genotyping Facility. The resulting data were run through a quality control pipeline to remove unusable data and bad assays as described (Lin et al. 2010). Transcript ratios were standardized based on the DNA control data for each assay as described (Lin et al. 2010).

Each of the gene pairs were represented by multiple SNP assays (GmDCL1 = six assays, GmDCL2 = three assays, GmDCL4 = seven assays, GmAGO1 = three assays, GmDRB1 = three assays, GmDRM1 = seven assays, GmMET1 = four assays, GmRDR6 = ten assays). For graphing purposes, transcript data were averaged among assays for each gene pair. Microsoft Excel, PowerPoint, and Spotfire DecisionSite 9.1.1 software were used to

generate figures and tables of the duplicate gene transcription data.

Results

Identification and structure of duplicated soybean genes involved in epigenetic processes

Amino acid sequences from RNAi pathway genes previously characterized in *Arabidopsis* were used to query the soybean genome sequence database (www.phytozome.net) to identify soybean homologs. Searches were performed for genes known to be involved in the RNAi pathway, including the four canonical DCL genes, AGO1, RDR6, and DRB1. Additionally, searches were performed to identify soybean homologs to the cytosine–DNA–methyltransferase genes, MET1 and DRM1. Two duplicate soybean gene homologs were found for all but one of these genes. A duplicate gene was not identified in soybean for the homolog of DCL3.

The soybean GmRDR6, GmDRB1, and GmMET1 gene duplicates all reside within duplicated homoeologous blocks between chromosomes 4 and 6 with an average of ~60 % duplicate gene conservation between blocks (Fig. 1; Table 1). Similarly, the two GmAGO1 genes reside in a small but highly conserved block between chromosomes 9 and 16. The two GmDRM1 copies reside on chromosomes 5 and 19, which do not appear to belong to any conserved homoeologous block.

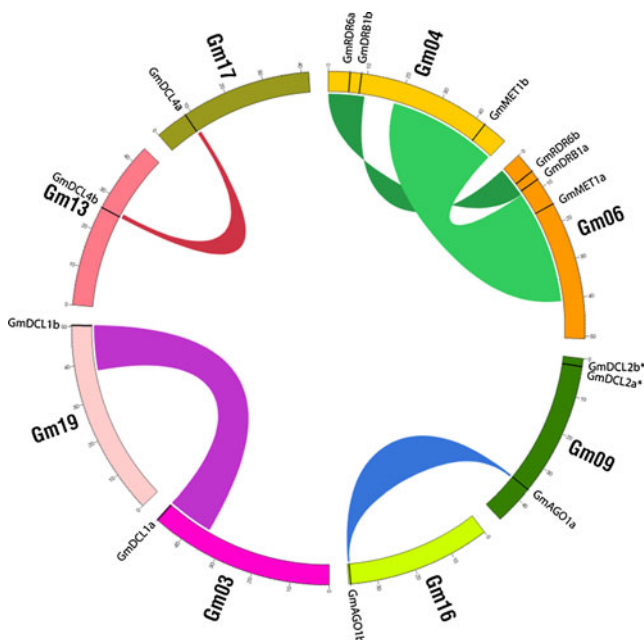


Fig. 1 Chromosomal positions of six homoeologous gene pairs in this study. The position of the GmDCL2 tandem duplication on chromosome 9 is also shown (indicated with asterisks)

The chromosomal locations of GmDCL1a and GmDCL1b were clearly defined and located in a large homoeologous block on the distal ends of chromosomes 3 and 19, respectively (Fig. 1). GmDCL2a and GmDCL2b both reside on chromosome 9 as a uni-directional tandem repeat separated by 5 kb. GmDCL2a and GmDCL2b share homology with two additional copies respectively located on chromosomes 8 and 15 (tentatively named GmDCL2c and GmDCL2d). However, analysis of predicted amino acid sequences of both GmDCL2c and GmDCL2d revealed several in-frame stop codons throughout the coding region, and both gene models have incomplete structural domains, leading to the conclusion that these “c” and “d” copies are likely pseudogenes. GmDCL4a and GmDCL4b are located on chromosomes 17 and 13, respectively, and are components of a small homoeologous block spanning several megabases between the two chromosomes. Both chromosomes 17 and 13 are highly rearranged, with many small homoeologous blocks matching several different chromosomes. The GmDCL3a locus situated on chromosome 4 resides within a robust homoeologous block between chromosomes 4 and 6, similar to the GmRDR6, GmDRB1, and GmMET1 gene duplicates. However, the GmDCL3b homoeologous candidate appears to be a pseudogene; GmDCL3b shares high DNA sequence homology (92–98 %) with GmDCL3a across several small regions of the locus but has an in-frame stop codon 42 amino acid residues down-stream of the start codon and lacks several critical domains.

Collectively, the genes analyzed in this study include six homoeologous pairs imbedded within homoeologous blocks, one unlinked paralogous pair (GmDRM1), one tandem repeat paralogous pair (GmDCL2), and a single gene copy with no intact duplicate (GmDCL3). The synonymous substitution rate (Ks values; Table 1) was calculated between each duplicate pair to estimate the age of the duplications. The age of duplication for the six homoeologous pairs ranged from 3 to 11.7 mya (data not shown). This finding, along with the estimated age of duplication for other gene pairs within these blocks, suggests that these duplications were potentially derived from the whole genome duplication event 9–14 mya. The age of duplication for the GmDCL2a tandem repeat was 19.4 mya, indicating that this duplication predated the whole-genome duplication event of soybean.

We calculated Ka/Ks ratios for each duplicate pair to examine whether any of the gene copies show evidence for current positive selection (Ka/Ks >1; Table 1). While none of these gene pairs show significant evidence of positive selection, GmDCL1a and b have a very high ratio relative to the other duplicates. This indicates that there may have been some positive selective pressure acting on one or both of these gene copies following duplication to allow for divergence in function, and over time those changes have become fixed and maintained under negative selection.

Transcriptional responses of DCL genes to abiotic stresses

Quantitative real-time PCR was used to measure the transcriptional responses of the seven GmDCL genes to abiotic stresses. Three genotypes (William 82, Archer, and Noir 1) were tested for four different stresses (cold, drought, SA, and high NaCl) along with an unstressed control. Each genotype \times treatment was performed on three biological replications. PCA was used to group genes based on expression profile similarity among the stress treatments, tissue types, and genotypes. The first principal component explained 52.1 % of the variance and the first two principal components combined explained 74.3 % of the variance. The data can be interpreted in a PCA biplot (Chapman et al. 2002; Park et al. 2008) of these two principal components (Fig. 2). GmDCL2a and GmDCL3a exhibited a strong co-upregulation in response to abiotic stress, as seen by the cluster on the right side of the biplot (Fig. 2). This co-upregulation of GmDCL2a and GmDCL3a is observed in nearly all stresses and tissue types, particularly roots (Fig. S1 of the “Electronic supplementary material”). Furthermore, GmDCL1b and GmDCL2b exhibited a strong co-downregulation in response to abiotic stress as seen by the cluster on the left (Fig. 2). This relationship between GmDCL1b and GmDCL2b was driven largely by the stress-induced co-downregulation in stem tissues (particularly the SA treatment; Fig. 1 of the “Electronic supplementary material”).

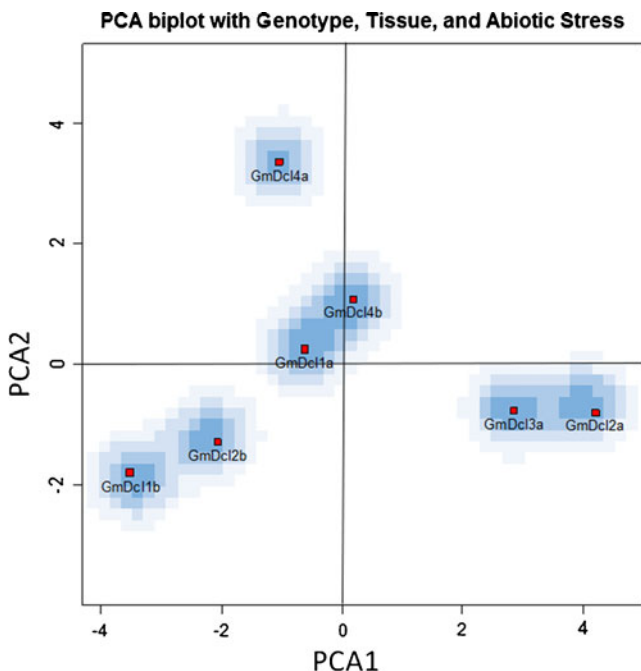


Fig. 2 Principal component analysis of the seven soybean Dicer-like genes in response to abiotic stresses. Real-time qPCR was performed on Williams 82, Archer, and Noir 1 cDNA in three different tissues under four abiotic stresses (and unstressed controls). This biplot shows clustering of soybean Dicer-like genes based on the first two principal components

GmDCL1a and GmDCL4b both plotted near the center of the biplot, indicating that they exhibited limited responses to the treatments. GmDCL4a did not cluster with any other gene, perhaps due to greater expression variation among genotypes.

Pairwise Pearson’s R correlations among the seven GmDCL genes were examined to further analyze co-expression patterns in response to stress (Table 2). The two major PCA clusters described above also exhibited the highest pairwise R values (GmDCL2a–GmDCL3a, 0.800; GmDCL1b–GmDCL2b, 0.821). The next highest R values were displayed by GmDCL2a–GmDCL2b ($R=0.656$) and GmDCL3a–GmDCL2b ($R=0.656$). Figure 3 shows scatterplots of the transcriptional responses to abiotic stresses for these four gene pair comparisons. Some trends were clearly observable for a given stress across the three genotypes. The most obvious was the effect of SA treatments on the GmDCL1b and GmDCL2b transcripts in stems; all three genotypes displayed a strong transcriptional down-regulation of these two gene copies (the yellow triangles in Fig. 3a). This treatment did not show down-regulation in the other five gene copies (note the positions of the yellow triangles in Fig. 3c, d). Figure S2 of the “Electronic supplementary material” shows all of the pairwise comparisons between the seven genes.

GmDCL2 was the only duplicate gene pair to display a high co-expression value ($R=0.656$; R values for the GmDCL1 and GmDCL4 paralogs were 0.195 and 0.268, respectively). This is surprising considering that these tandem-arranged copies are estimated to have a divergence time that is much more ancient (19.4 mya) than the 9–14 mya divergence estimate of the homoeologs (i.e., the presumed divergence time of the GmDCL1 and GmDCL4 duplicates). Therefore, the GmDCL2a and GmDCL2b were further analyzed for their transcriptional responses to stresses relative to one another. Figure 4 shows the relative transcriptional responses of GmDCL2a versus GmDCL2b as measured by both qRT-PCR and Sequenom MassARRAY quantitative SNP assays. These data indicate that, despite their co-expression, GmDCL2a is more transcriptionally responsive to stresses than GmDCL2b across nearly all genotypes, tissue types, and stresses tested. The qRT-PCR data in Fig. 3c support this trend as the distribution of data points is relatively flat but extends much further along the right side of the GmDCL2a axis, indicating that GmDCL2a transcripts are frequently up-regulated under stress, while GmDCL2b transcript levels show less response.

Another important finding in the qRT-PCR versus MassARRAY comparison was the relative cross-validation of the two platforms (Fig. 4). Sequenom MassARRAY is a multiplex PCR assay that allows for the automated quantification of several different SNPs in a single reaction. The SNPs can be quantified between paralogous genes for 384

Table 2 Real-time qPCR correlation matrix of transcript response to abiotic stress for the Dicer-like gene family members^a

	GmDCL1b	GmDCL2a	GmDCL2b	GmDCL3a	GmDCL4a	GmDCL4b
GmDCL1a	0.195	0.312	0.356	0.374	-0.118	0.149
GmDCL1b		0.355	0.821	0.361	-0.157	0.167
GmDCL2a			0.656	0.800	0.285	0.533
GmDCL2b				0.656	0.107	0.471
GmDCL3a					0.129	0.438
GmDCL4a						0.268

^a Data are represented as Pearson's *R* values based on log₂ transformations of the transcriptional fold-change relative to the unstressed control across 36 treatment × tissue × genotype samples

templates for approximately 30 SNPs per sample. Therefore, this technology has much higher throughput than standard quantitative PCR. We chose to use the MassARRAY technology to screen relative transcriptional responses for a larger set of paralogous genes and a larger panel of biotic and abiotic stresses.

Assessing transcription of eight paralogous gene pairs in response to biotic and abiotic stresses

Along with the three GmDCL gene pairs, MassARRAY SNP assays were designed for five additional paralogous gene pairs involved in RNAi pathways and other epigenetic

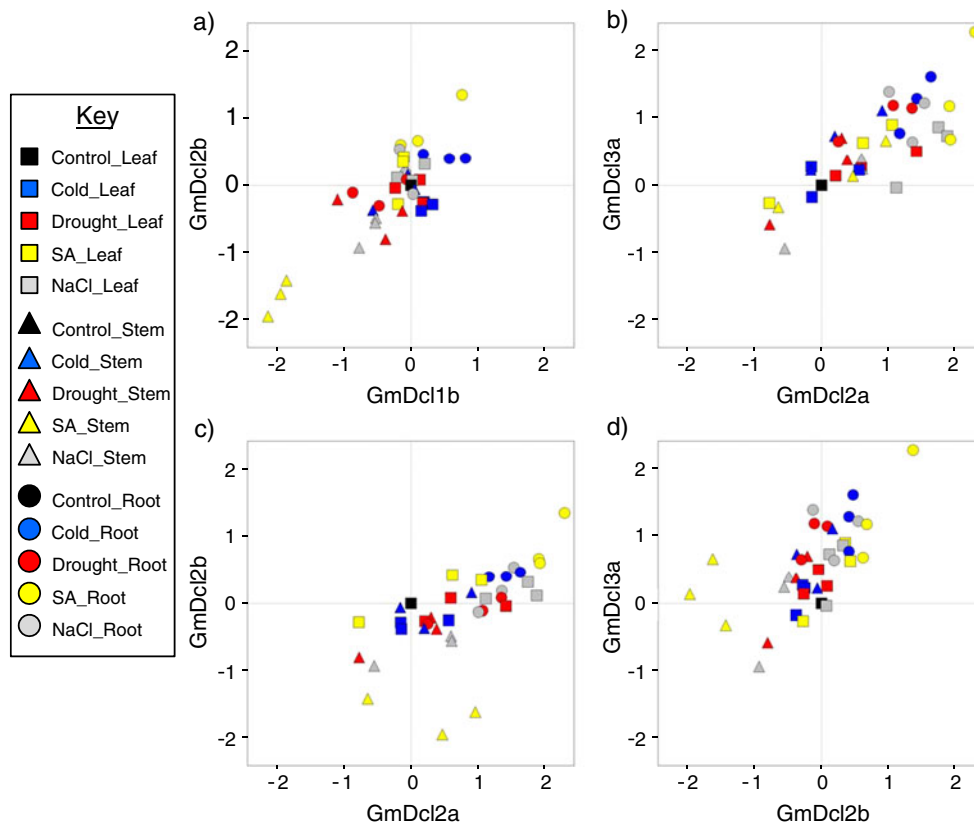
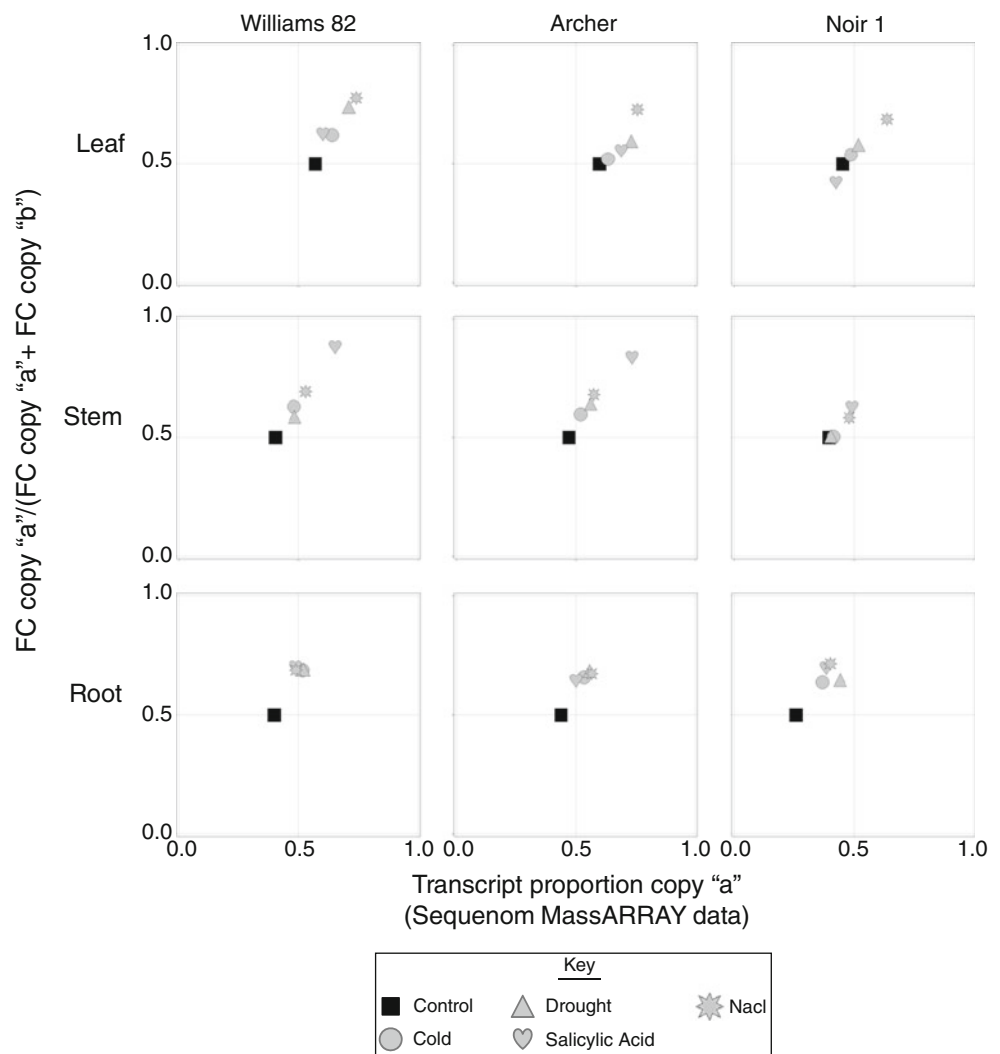


Fig. 3 a–d Gene × gene pairwise plots of transcriptional changes for the Dicer-like gene family members in response to abiotic stresses. Real-time qPCR was performed on Williams 82, Archer, and Noir 1 cDNA (genotypes are not distinguished in the plots) for 15 treatment × tissue type groups (color and shape coded according to the key; SA salicylic acid). Data are represented as log₂-transformed values of the fold change relative to the unstressed control; each data point is the mean of three biological replications. The controls (shown as *black squares*) all plot at position (0, 0) and the range for each plot is -2.4 to

2.4 for both the *X* and *Y* axes. Data points that plot in the *upper right quadrant* represent samples in which both genes in the plot were transcriptionally upregulated; data points in the *lower left quadrant* represent samples in which both genes in the plot were transcriptionally downregulated. The comparisons shown in this figure represent the four highest pairwise correlations in this data set. The complete set of pairwise correlations is shown in Fig. S2 of the “Electronic supplementary material”

Fig. 4 GmDcl2a shows a transcriptional increase relative to GmDcl2b in response to abiotic stress. Sequenom MassARRAY (X axis) and real-time PCR (Y axis) assays were used to estimate the relative transcriptional responses of GmDcl2a and GmDcl2b to four different abiotic stresses (see key) among three genotypes and three tissue types (each data point is the mean of three biological replications). The real-time PCR data were computed as the fold-change (FC) of GmDcl2a and GmDcl2b relative to the unstressed control. The GmDcl2a and GmDcl2b control FC values were set to 1.0, thus all of the control data plot along value 0.5 on the y axis (shown as *black squares*). Nearly all of the stress \times genotype \times tissue data points plot to the upper right of the control, indicating that the stresses elicited a transcriptional increase of GmDcl2a relative to GmDcl2b (cross-validated by the real-time PCR and Sequenom MassARRAY platforms)



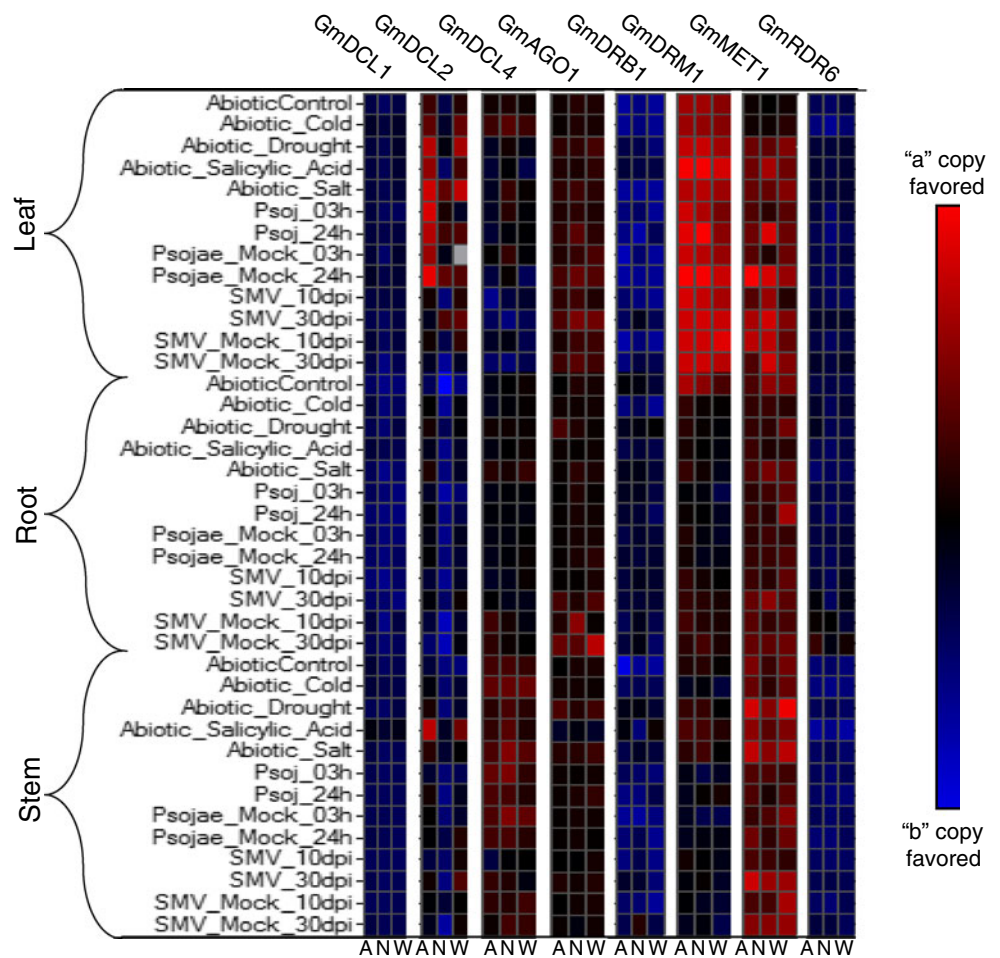
processes (GmAGO1, GmDRB1, GmDRM1, GmMET1, and GmRDR6). The MassARRAY system allowed us to measure the relative transcriptional changes between paralogous pairs across the four abiotic stresses described in the previous section and four biotic stress stages (*P. sojae* post-inoculation at 3 h, *P. sojae* post-inoculation at 24 h, SMV at 10 days post-inoculation (dpi), and SMV at 30 dpi).

Figure 5 shows a heat map of the relative expression levels of the duplicate gene copies across the control and treated samples. The variation in relative response among the gene pairs is notable. The GmDCL1, GmAGO1, GmDRB1, and GmRDR6 paralogous pairs displayed only subtle relative responses to any of the treatments. The other four paralogous pairs each displayed unique patterns. GmMET1 showed a strong up-regulation of the “a” copy in some specific stresses. GmDCL4 and GmDRM1 showed tissue-specific changes: the GmDCL4a copy was up-regulated in stem tissues and the GmDRM1a copy was strongly up-regulated in leaf tissues.

GmDCL2 displayed the most chaotic patterns. First, the down-regulation of GmDCL2b in SA-treated stems is clearly evident as a red row in an otherwise blue set of tiles (Fig. 5). This result confirms the same trend observed in the quantitative PCR data (Fig. 3). Second, the range of relative expression changes was extreme, favoring the “b” transcript in root and stem tissues but favoring the “a” transcript in some leaf tissue treatments. Furthermore, the transcriptional response to abiotic stresses almost always showed a relative up-regulation of the “a” copy (also see Fig. 4), but this response was not as strong or universal across the biotic stress treatments. Interestingly, GmDCL2 was the only paralogous gene pair that displayed a consistent difference among the genotypes. In both control and treated samples, Noir 1 showed a favoring of the “b” transcript compared to Williams 82 and Archer.

A more detailed examination of the duplicate gene transcript ratios among the genotypes is shown in Fig. S3 of the “Electronic supplementary material”. The genotype \times genotype comparisons indicate that Williams 82 and Archer

Fig. 5 Heat map of the ratio of transcript abundance between duplicate gene pairs as measured by Sequenom MassARRAY. *Red* indicates that transcription of the ‘a’ copy is favored, *black* indicates that the duplicate pairs have approximately equal transcription levels, while *blue* indicates that transcription of the ‘b’ copy is favored. The scale on the *right* indicates coloration used to depict the relative transcript proportions between duplicate copies. The heat map is organized by gene pair and tissue type, with each column representing a genotype (*A* Archer, *W* Williams 82, *N* Noir 1)



display similar duplicate transcript ratios for the set of eight gene pairs relative to the Williams 82–Noir 1 and Archer–Noir 1 comparisons. A gene expression heat map grouped by genotype also illustrates this point (Fig. S4 of the “Electronic supplementary material”). Additionally, these data indicate that there were generally fewer transcriptional differences between genotypes in response to the abiotic stresses than were observed in response to the biotic stresses (Fig. S3 of the “Electronic supplementary material”).

The relative expression ratios from the MassARRAY data were compared among the eight gene duplicates to identify possible interactions among sets of duplicate pairs. Pairwise correlations among the eight gene pairs were calculated from the entire set of MassARRAY ratios (Table S2 of the “Electronic supplementary material”). Nearly half (13 of 28) of the pairwise comparisons were significant. We also analyzed the positive co-expression trends to stress response within each treatment × tissue combination (Table S3 of the “Electronic supplementary material”). The data for the three genotypes were combined for this analysis, leaving a total of 24 interaction tests for each pair of genes. There was clearly far more co-expression interactions in response to abiotic stresses (36 significant interactions) than biotic stresses (five

significant interactions). In fact, no significant interactions were observed in response to *P. sojae* inoculation. The total number of positive co-expression interactions for each gene pair is shown in Table S4 of the “Electronic supplementary material”.

Discussion

Profiling for duplicate gene co-expression

Stress conditions are known to trigger responses in gene expression which are regulated at both the transcriptional and post-transcriptional levels. In this study, PCR-based assays allowed us to screen the relative gene expression levels of eight gene pairs over three soybean genotypes, three tissue types, and eight stress conditions. These data revealed a wide range of patterns among the different gene pairs, including stress- and tissue-specific transcriptional responses.

The most profound co-expression patterns among the seven GmDCL genes were observed among non-paralogous copies, particularly GmDCL1b–GmDCL2b and

GmDCL2a–GmDCL3a. The GmDCL1b–GmDCL2b co-expression pattern was most strikingly observed as a down-regulation in stems exposed to SA. The GmDCL2a–GmDCL3a co-expression pattern was driven mainly by up-regulation of both genes in response to various stress × tissue treatments but was most profound in stressed roots. We did not find any mechanistic rationale, such as gene promoter similarities, that might explain the observed co-expression between the non-paralogous GmDCL genes. Furthermore, at this point, it is unclear if the respective GmDCL1b–GmDCL2b or GmDCL2a–GmDCL3a co-expression patterns are associated with shared or coordinated functions between the pathways assigned to each gene class (e.g., the miRNA, nat-siRNA, and/or rasiRNA pathways).

The co-expression analysis of homoeologous or paralogous GmDCL copies revealed a surprising and perhaps counter-intuitive relationship between the age of duplication and co-expression in response to stress. GmDCL2a and GmDCL2b are a tandem-arranged paralogous pair with age of duplication estimated to be 19.4 mya. Evidence suggests that the GmDCL1 and GmDCL4 gene pairs are more recent homoeologous duplications (Fig. 1), potentially resulting from the whole-genome duplication event 9–14 mya. One would expect that more recent duplicates will display stronger co-expression patterns than more ancient duplicates; however, our data revealed that the GmDCL2 paralogs exhibited much stronger patterns of co-expression than either the GmDCL1 or GmDCL4 homoeologous pairs. Based on our analysis, there is no clear explanation for this finding, as there are no obvious selective mechanisms (K_a/K_s) or promoter sequence conservation differences between the GmDCL2 paralogs as compared with the GmDCL1 and GmDCL4 duplicates. However, the GmDCL2 paralogs are the only tandem-arranged duplicates in this study and may thereby be exposed to similar chromatin and/or epigenetic states in response to stress and development. Furthermore, a recent study of soybean small RNAs identified a 22-nucleotide miRNA (miR1515) that specifically targets the GmDCL2b locus (Zhai et al. 2011). This class of miRNA has been shown to trigger the production of secondary small RNAs (e.g., tasiRNAs) (Chen et al. 2010) that, in turn, may target the GmDCL2a copy and/or other GmDCL copies. This regulatory cascade may explain the co-expression of the GmDCL2 duplicates across treatments, and similar mechanisms may also influence co-expression patterns among non-paralogous copies.

Regulatory pathways and stress response

Based on homology to characterized *Arabidopsis* genes, we can broadly divide the genes investigated in this study into three distinct processes involved in transcriptional and post-

transcriptional regulation: (1) DNA methylation (GmMET1 and GmDRM1), (2) RNAi processing (GmDCL2, GmDCL4, and GmRDR6), and (3) miRNA processing (GmDCL1, GmDRB1, and GmAGO1). All three of these processes have been implicated in plant stress responses. DNA methylation changes (Chinnusamy and Zhu 2009) and alterations of specific siRNAs and miRNAs (Kulcheski et al. 2011; Silva et al. 2011; Sunkar et al. 2007), as well as their targets (Borsani et al. 2005), have been reported across a wide range of plant species and stresses.

Our data allow us to compare the transcriptional response to stress for the genes in these three categories. At first glance, our results suggest that the DNA methylation and RNAi pathway genes are more responsive to stress than miRNA pathway genes in soybean. GmDCL3a and the GmDCL2 paralogs, particularly GmDCL2a, exhibited a wide range of transcriptional changes in response to stress (Fig. 3; Fig. S2 of the “Electronic supplementary material”). This finding suggests that these genes may play an important role in stress response. Though co-expressed, GmDCL2a consistently exhibited a relative up-regulation to stresses compared to the GmDCL2b copy (Fig. 4). GmDCL2a may function as a component of the nat-siRNA pathway and/or a surrogate component to DCL4 in anti-virus defense (Dunoyer et al. 2010). Furthermore, the relative transcript analysis of the DNA methyltransferase duplicate genes for GmDRM1 and GmMET1 showed evidence of transcriptional responses to the stress treatments (Fig. 5).

The GmDCL1 duplicates showed some transcriptional changes in response to stress; however, the range was more subtle than the other GmDCL genes. GmDCL1b showed a strong co-expression with GmDCL2b, including a conspicuous down-regulation in response to salicylic acid treatment in stems. Taking into account the ability of many DCL and DRB proteins to compete and antagonize one another, it would be premature to dismiss the influence of the miRNA pathway in soybean stress response. In fact, recent reports in soybean have identified over 200 miRNAs, including several that exhibited differential expression under abiotic and biotic stress (Kulcheski et al. 2011; Li et al. 2011). However, the involvement of the different genes regulating and processing transcripts will remain unresolved until functional analysis can be carried out with appropriate soybean mutants.

The data set presented here may be particularly useful for designing targeted experiments that focus on functional divergence between duplicated genes. The development of new soybean mutant resources (Bolon et al. 2011; Hancock et al. 2011; Mathieu et al. 2009; Pham et al. 2010) and new methods capable of producing single and double mutants (Curtin et al. 2011) will be crucial for the advancement of soybean functional genomics and for studies of functional divergence between soybean duplicate genes. Mutant

phenotypes for duplicated genes are frequently difficult to identify, largely due to the genetic buffer provided by the duplicate copy(s) (Bouche and Bouchez 2001; Jander and Barth 2007). Phenotypes for loss of function mutants may be more attainable and informative when screened under conditions known to trigger transcriptional differentiation between the duplicate copies.

Acknowledgments We thank Ben Lockhart for providing the SMV pathogen and Dean Malvick for providing the *P. sojae* pathovar. We are grateful to Dinesha Walek and the University of Minnesota BMGC for running the MassARRAY experiments. We are very grateful to Mingbo Wang and Andrew Eamens for providing helpful comments on the manuscript. This work was supported by the United States Department of Agriculture—Agricultural Research Service (Agreement 58-3640-0-626) and a Pioneer Hi-Bred graduate fellowship to M.B.K.

References

- Adams KL (2007) Evolution of duplicate gene expression in polyploid and hybrid plants. *J Hered* 98:136–141
- Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121:207–221
- Arabidopsis Interactome Mapping Consortium (2011) Evidence for network evolution in an *Arabidopsis* interactome map. *Science* 333:601–607
- Ben Amor B, Wirth S, Merchan F, Laporte P, d'Aubenton-Carafa Y, Hirsch J, Maizel A, Mallory A, Lucas A, Deragon JM, Vaucheret H, Thermes C, Crespi M (2009) Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. *Genome Res* 19:57–69
- Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16:1667–1678
- Bolon YT, Haun WJ, Xu WW, Grant D, Stacey MG, Nelson RT, Gerhardt DJ, Jeddeloh JA, Stacey G, Muehlbauer GJ, Orf JH, Naeve SL, Stupar RM, Vance CP (2011) Phenotypic and genomic analyses of a fast neutron mutant population resource in soybean. *Plant Physiol* 156:240–253
- Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* 123:1279–1291
- Bouche N, Bouchez D (2001) *Arabidopsis* gene knockout: phenotypes wanted. *Curr Opin Plant Biol* 4:111–117
- Bouche N, Laussergues D, Gascioli V, Vaucheret H (2006) An antagonistic function for *Arabidopsis* DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J* 25:3347–3356
- Boyko A, Blevins T, Yao Y, Golubov A, Bilichak A, Ilnytskyy Y, Hollunder J, Meins F Jr, Kovalchuk I (2010) Transgenerational adaptation of *Arabidopsis* to stress requires DNA methylation and the function of Dicer-like proteins. *PLoS One* 5:e9514
- Brosnan CA, Mitter N, Christie M, Smith NA, Waterhouse PM, Carroll BJ (2007) Nuclear gene silencing directs reception of long-distance mRNA silencing in *Arabidopsis*. *Proc Natl Acad Sci USA* 104:14741–14746
- Buggs RJ, Zhang L, Miles N, Tate JA, Gao L, Wei W, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE (2011) Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Curr Biol* 21:551–556
- Cao X, Jacobsen SE (2002) Role of the *Arabidopsis* DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* 12:1138–1144
- Capitao C, Paiva JA, Santos DM, Fevereiro P (2011) In *Medicago truncatula*, water deficit modulates the transcript accumulation of components of small RNA pathways. *BMC Plant Biol* 11:79
- Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat Rev Genet* 6:351–360
- Chapman S, Schenk P, Kazan K, Manners J (2002) Using biplots to interpret gene expression patterns in plants. *Bioinformatics* 18:202–204
- Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF (2009) Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (gossypium). *Genetics* 182:503–517
- Chen HM, Chen LT, Patel K, Li YH, Baulcombe DC, Wu SH (2010) 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc Natl Acad Sci USA* 107:15269–15274
- Chinnusamy V, Zhu JK (2009) Epigenetic regulation of stress responses in plants. *Curr Opin Plant Biol* 12:133–139
- Curtin SJ, Zhang F, Sander JD, Haun WJ, Starker C, Baltes NJ, Reyon D, Dahlborg EJ, Goodwin MJ, Coffman AP, Dobbs D, Joung JK, Voytas DF, Stupar RM (2011) Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. *Plant Physiol* 156:466–473
- Dong S, Adams KL (2011) Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. *New Phytol* 190:1045–1057
- Dunoyer P, Brosnan CA, Schott G, Wang Y, Jay F, Alioua A, Himber C, Voinnet O (2010) An endogenous, systemic RNAi pathway in plants. *EMBO J* 29:1699–1712
- Eamens A, Vaistij FE, Jones L (2008a) NRPD1a and NRPD1b are required to maintain post-transcriptional RNA silencing and RNA-directed DNA methylation in *Arabidopsis*. *Plant J* 55:596–606
- Eamens A, Wang MB, Smith NA, Waterhouse PM (2008b) RNA silencing in plants: yesterday, today, and tomorrow. *Plant Physiol* 147:456–468
- Eamens AL, Smith NA, Curtin SJ, Wang MB, Waterhouse PM (2009) The *Arabidopsis thaliana* double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA* 15:2219–2235
- Finnegan EJ, Dennis ES (1993) Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Res* 21:2383–2388
- Flagel LE, Chen L, Chaudhary B, Wendel JF (2009) Coordinated and fine-scale control of homoeologous gene expression in allotetraploid cotton. *J Hered* 100:487–490
- 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
- Hancock CN, Zhang F, Floyd K, Richardson AO, Lafayette P, Tucker D, Wessler SR, Parrott WA (2011) The rice miniature inverted repeat transposable element mPing is an effective insertional mutagen in soybean. *Plant Physiol* 157:552–562
- Hovav R, Udall JA, Chaudhary B, Rapp R, Flagel L, Wendel JF (2008) Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proc Natl Acad Sci USA* 105:6191–6195
- Jackson S, Chen ZJ (2010) Genomic and expression plasticity of polyploidy. *Curr Opin Plant Biol* 13:153–159
- Jander G, Barth C (2007) Tandem gene arrays: a challenge for functional genomics. *Trends Plant Sci* 12:203–210
- Jian B, Liu B, Bi Y, Hou W, Wu C, Han T (2008) Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Mol Biol* 9:59

- Kachroo A, Fu DQ, Havens W, Navarre D, Kachroo P, Ghabrial SA (2008) An oleic acid-mediated pathway induces constitutive defense signaling and enhanced resistance to multiple pathogens in soybean. *Mol Plant Microbe Interact* 21:564–575
- Katiyar-Agarwal S, Morgan R, Dahlbeck D, Borsani O, Villegas A Jr, Zhu JK, Staskawicz BJ, Jin H (2006) A pathogen-inducible endogenous siRNA in plant immunity. *Proc Natl Acad Sci USA* 103:18002–18007
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–1645
- Kulcheski FR, de Oliveira LF, Molina LG, Almerao MP, Rodrigues FA, Marcolino J, Barbosa JF, Stolf-Moreira R, Nepomuceno AL, Marcelino-Guimaraes FC, Abdelnoor RV, Nascimento LC, Carazzolle MF, Pereira GA, Margis R (2011) Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics* 12:307
- la Grange A, le Roux N, Gardner-Lubbe S (2009) BiplotGUI: interactive biplots in R. *J Stat Softw* 30:1–30
- Li H, Dong Y, Yin H, Wang N, Yang J, Liu X, Wang Y, Wu J, Li X (2011) Characterization of the stress associated microRNAs in *Glycine max* by deep sequencing. *BMC Plant Biol* 11:170
- Lin JY, Stupar RM, Hans C, Hyten DL, Jackson SA (2010) Structural and functional divergence of a 1-Mb duplicated region in the soybean (*Glycine max*) genome and comparison to an orthologous region from *Phaseolus vulgaris*. *Plant Cell* 22:2545–2561
- Liu Z, Adams KL (2007) Expression partitioning between genes duplicated by polyploidy under abiotic stress and during organ development. *Curr Biol* 17:1669–1674
- Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ, Waterhouse PM (2006) The evolution and diversification of Dicers in plants. *FEBS Lett* 580:2442–2450
- Mathieu M, Winters EK, Kong F, Wan J, Wang S, Eckert H, Luth D, Paz M, Donovan C, Zhang Z, Somers D, Wang K, Nguyen H, Shoemaker RC, Stacey G, Clemente T (2009) Establishment of a soybean (*Glycine max* Merr. L) transposon-based mutagenesis repository. *Planta* 229:279–289
- Mlotshwa S, Pruss GJ, Peragine A, Endres MW, Li J, Chen X, Poethig RS, Bowman LH, Vance V (2008) DICER-LIKE2 plays a primary role in transitive silencing of transgenes in *Arabidopsis*. *PLoS One* 3:e1755
- Navarro L, Jay F, Nomura K, He SY, Voinnet O (2008) Suppression of the microRNA pathway by bacterial effector proteins. *Science* 321:964–967
- Nomura T, Ishihara A, Yanagita RC, Endo TR, Iwamura H (2005) Three genomes differentially contribute to the biosynthesis of benzoxazinones in hexaploid wheat. *Proc Natl Acad Sci USA* 102:16490–16495
- Park M, Lee JW, Leec JB, Song SH (2008) Several biplot methods applied to gene expression data. *J Stat Plan Infer* 138:500–515
- Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29:e45
- Pham AT, Lee JD, Shannon JG, Bilyeu KD (2010) Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. *BMC Plant Biol* 10:195
- Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet* 3:827–837
- Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC (2004) Mining EST databases to resolve evolutionary events in major crop species. *Genome* 47:868–876
- Schlueter JA, Lin JY, Schlueter SD, Vasylenko-Sanders IF, Deshpande S, Yi J, O'Bleness M, Roe BA, Nelson RT, Scheffler BE, Jackson SA, Shoemaker RC (2007) Gene duplication and paleopolyploidy in soybean and the implications for whole genome sequencing. *BMC Genomics* 8:330
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL et al (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183
- Silva TF, Romanel EA, Andrade RR, Farinelli L, Osteras M, Deluen C, Correa RL, Schrago CE, Vaslin MF (2011) Profile of small interfering RNAs from cotton plants infected with the polerovirus Cotton leaf roll dwarf virus. *BMC Mol Biol* 12:40
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. *Am J Bot* 96:336–348
- Stamati K, Mackay I, Powell W (2009) A quantitative genomic imbalance gene expression assay in a hexaploid species: wheat (*Triticum aestivum*). *Genome* 52:89–94
- Sunkar R, Chinnusamy V, Zhu J, Zhu JK (2007) Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. *Trends Plant Sci* 12:301–309
- Vaucheret H (2008) Plant ARGONAUTS. *Trends Plant Sci* 13:350–358
- Wassenegger M, Krczal G (2006) Nomenclature and functions of RNA-directed RNA polymerases. *Trends Plant Sci* 11:142–151
- Yan Y, Zhang Y, Yang K, Sun Z, Fu Y, Chen X, Fang R (2011) Small RNAs from MITE-derived stem-loop precursors regulate abscisic acid signaling and abiotic stress responses in rice. *Plant J* 65:820–828
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591
- Zhai J, Jeong DH, De Paoli E, Park S, Rosen BD, Li Y, Gonzalez AJ, Yan Z, Kitto SL, Grusak MA, Jackson SA, Stacey G, Cook DR, Green PJ, Sherrier DJ, Meyers BC (2011) MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes Dev* 25:2540–2553