

# Comparative analyses of transcriptome and proteome in response to cotton bollworm between a resistant wild soybean and a susceptible soybean cultivar

Xiaoyi Wang<sup>1,2</sup> · Jianhua Lu<sup>1,2</sup> · Haifeng Chen<sup>1</sup> · Zhihui Shan<sup>1</sup> · Xinjie Shen<sup>1</sup> · Bingbing Duan<sup>1</sup> · Chanjuan Zhang<sup>1</sup> · Zhonglu Yang<sup>1</sup> · Xiaojuan Zhang<sup>1</sup> · Dezhen Qiu<sup>1</sup> · Shuilian Chen<sup>1</sup> · Xinan Zhou<sup>1</sup> · Yongqing Jiao<sup>1</sup>

Received: 25 October 2016 / Accepted: 22 February 2017 / Published online: 3 March 2017  
© The Author(s) 2017. This article is published with open access at Springerlink.com

**Abstract** Soybean is an important oil and protein crop in the world. Cotton bollworm (*Helicoverpa armigera*) is one of primary defoliating insects which seriously damage the soybean production. The wild soybean, ED059, was reported to be highly resistant to *H. armigera*. However, the molecular mechanism of resistance to *H. armigera* remains unknown. In this study, we conducted the transcript and protein profilings for ED059 and a susceptible soybean cultivar (cv.) Tianlong 2. In summary, 2213 genes and 116 proteins were specifically expressed in response to *H. armigera* in ED059, while 2179 genes and 9 proteins were specifically expressed in cv. Tianlong 2. Four hundred

and four genes were significantly up-regulated in response to *H. armigera* in ED059 while down-regulated or not changed in cv. Tianlong 2. GO analyses showed the specifically expressed genes in ED059 were mainly enriched in metabolism-related pathways. The function of one potential resistant gene involved in the phenylpropanoid pathway was confirmed by use of transgenic *Arabidopsis*. Our study provided useful resources for the soybean improvement, which could contribute to the elucidation of defense mechanisms against chewing insects in plants.

**Keywords** Soybean · Cotton bollworm · Transcriptome · Proteome · Defence

Xiaoyi Wang and Jianhua Lu have contributed equally to this work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11240-017-1196-5) contains supplementary material, which is available to authorized users.

✉ Xinan Zhou  
zhouxinan@caas.cn

✉ Yongqing Jiao  
jiaoyongqing@caas.cn

Xiaoyi Wang  
micky\_xiao@163.com

Jianhua Lu  
huajianlu666@163.com

Haifeng Chen  
smartchf@163.com

Zhihui Shan  
zhihuijimi@163.com

Xinjie Shen  
ylssxj@163.com

Bingbing Duan  
yuyufly4@163.com

Chanjuan Zhang  
zhangchanjuan@aliyun.com

Zhonglu Yang  
yangzhonglu@126.com

Xiaojuan Zhang  
xjzhang@oilcrops.cn

Dezhen Qiu  
qiudz6677@yahoo.cn

Shuilian Chen  
chenshuilianjy@163.com

- 1 Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Oil Crops Research Institute of the Chinese Academy of Agricultural Science, Wuhan, China
- 2 Graduate School of the Chinese Academy of Agricultural Sciences, Beijing, China

## Introduction

Plant–insect interactions have been widely studied (Ferry et al. 2004; Gatehouse 2002; Kessler and Baldwin 2002; Van Poecke 2007; Wu and Baldwin 2009). The resistances to herbivores in plants could be divided into two types, direct and indirect (Kessler and Baldwin 2002). The direct defense are conferred by some secondary metabolites, proteinase inhibitors (PIs) and physical barriers, which have toxic, repellent or anti-digestive effects on the performance of insect life (Wu and Baldwin 2009). The indirect defense means that the plants gain the protection by emitting some volatile compounds that attract the natural enemies of herbivores (Heil 2008; Kessler and Baldwin 2002). In addition to the above definition, the plant defense against herbivory insects can also be categorized as constitutive and induced defenses based on the timing of defense activation (Gatehouse 2002). No matter what kind of defenses, they are the burdens on plant growth or reproduction (Zavala and Baldwin 2004). Therefore, a sophisticated regulatory network is needed by plants to maintain a balance between the growth and defense responses.

Three kinds of phytohormones, jasmonic acid (JA), ethylene (ET) and salicylic acid (SA), were found to play important roles in response to herbivore attack (Wu and Baldwin 2009). JA positively regulated a series of genes conferring resistance to herbivory insects, whereas ET and SA mainly functioned as negative regulators, which have antagonistic effects on JA-dependent responses (Wu and Baldwin 2009). In addition to JA, ET and SA, other phytohormones such as abscisic acid (ABA), auxin, cytokinin and gibberellins, are also involved in herbivore-induced responses (Gatehouse 2002; Stam et al. 2014). In addition to phytohormones, anti-herbivore secondary metabolites such as phenylpropanoids, glucosinolates, alkaloids and phenolics, also had important functions in conferring resistance to herbivores in plants (Wu and Baldwin 2009). Phenylpropanoids were a diverse group of secondary metabolites including flavonoids, coumarins, monolignols and lignans, which were derived from the phenylalanine, an end product of the shikimate pathway (Fraser and Chapple 2011). The phenylpropanoid pathway was indispensable to plants and involved in plant defense, structural support and survival (Vogt 2010). Overexpression of a transcription factor gene *PAP1* that regulated the biosynthesis of phenylpropanoids in *Arabidopsis*, resulted in the reduced feeding rates of generalist *Spodoptera frugiperda* (Johnson and Dowd 2004).

Soybean (*Glycine max*) is one of the most important oil and protein crops in the world. Cotton bollworm (*Helicoverpa armigera*) as well as beet armyworm (*Spodoptera exigua* Hübner), common leafworm (*Spodoptera litura*), bean leaf beetle (*Cerotoma trifurcata*) and green cloverworm

(*Heliothis dipsacea*), are important defoliating insects that seriously damage soybean yield and quality. Despite Bt genes and chemical insecticides that were widely used in fighting against these insects (Stewart et al. 1996), utilization of soybean native resistant genes through either traditional breeding or genetic engineering, could broaden the options of the insect management and be economical and environment friendly (Wang et al. 2014, 2015a, b; Chen et al. 2015; Gunadi et al. 2016; Zhai et al. 2017). Thus, identifying native resistant genes and elucidating the molecular mechanism of resistance to defoliating insects are of great significance for the improvement of soybean. As the ancestors of cultivated soybean, wild soybean (*Glycine soja*) retained more genetic diversity and carried more useful stress-resistant genes, which might be lost during domestication in cultivated soybean (Guan et al. 2014; Li et al. 2014; Qi et al. 2014; Winter et al. 2007). Therefore, mining genes with resistance to stresses from the wild soybean sources becomes the concern of soybean researchers. In a previous evaluation of soybean germplasm for their resistance to herbivory insects, we found that a wild soybean, ED059, was the one with highest level of resistance to *H. armigera* (Wang et al. 2015a, b). In order to learn the resistance mechanism in ED059, we conducted comparative analyses of transcriptome and proteome in response to *H. armigera* between ED059 and a susceptible control, cultivar (cv.) Tianlong 2. The results suggested a potential important role for metabolic processes in conferring resistance to *H. armigera* in ED059. One candidate resistance gene was functionally confirmed by overexpressing it in *Arabidopsis*. Our studies provided insights into the mechanism of resistance to *H. armigera* in wild soybean ED059, which could contribute to the understanding of defence mechanism against herbivory insects in plants.

## Results

### Overview of RNA-seq data

Our previous studies showed that ED059 had a high level of resistance to *H. armigera* in either choice or no-choice tests (Wang et al. 2015a, b). To learn the resistant mechanism, we conducted transcriptome analyses by RNA-seq for the leaves at 24 h after treatment with or without *H. armigera* between ED059 and a susceptible cultivar (cv.), Tianlong 2. Four libraries were constructed, which included: 1L, Tianlong 2 at 24 h without treatment; 1-L, Tianlong 2 at 24 h with treatment of *H. armigera*; 2L, ED059 at 24 h without treatment; and 2-L, ED059 at 24 h with treatment of *H. armigera* (Table 1). The number of total raw tags for each library ranged from 5.82 to 6.23 million (Table 1). After filtered by removing 3'

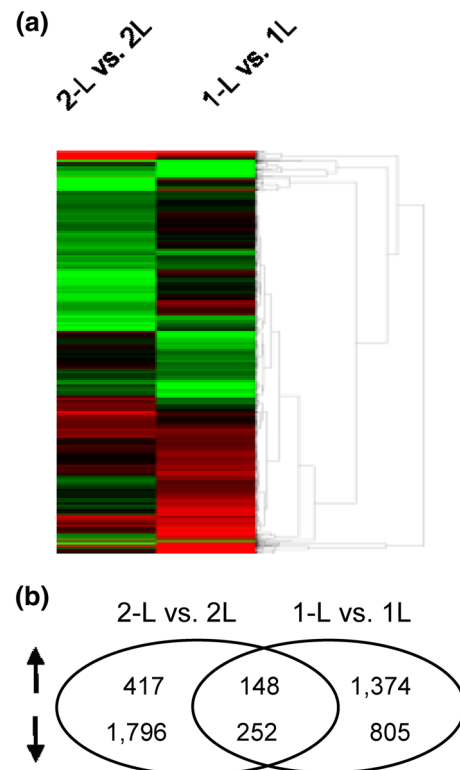
**Table 1** Summary of RNA-seq data for four samples, 1L, 1-L, 2L and 2-L. 1L, Tianlong 2 at 24 h without treatment of *H. armigera*; 1-L, Tianlong 2 at 24 h with treatment of *H. armigera*; 2L, ED059 at 24 h without treatment; and 2-L, ED059 at 24 h with treatment of *H. armigera*

Samples	Raw tags	Clean tags	Tags mapped to genes	Unique tags mapped to genes	Unknown tags
1-L	5871074	5679012	4925506 (86.73%)	2602211 (45.82%)	496552 (8.74%)
1L	5830235	5660877	5067657 (89.52%)	2619755 (46.28%)	314552 (5.56%)
2-L	5820915	5656114	4965780 (87.79%)	2599277 (45.96%)	433059 (7.66%)
2L	6229014	6031104	5118071 (84.86%)	2795739 (46.36%)	546216 (9.06%)

adaptor fragments, low-quality tags and several types of impurities, the clean tags were produced for each library, which accounted for 97% of raw tags (Table 1). More than 2.5 million tags for each library, accounting more than 45% of clean tags, were unambiguously mapped to the reference genome (<http://www.soybase.org/>) (Table 1). The analyses of sequencing saturation showed that the number of detected genes approached saturation when the number of tags reached two millions (Additional file 1), which indicated that our RNA-seq data for each library contained sufficient information of soybean genes for further studies of gene expression. The number of unambiguous clean tags for each gene was normalized to TPM (number of transcripts per million clean tags) to represent the expression level of each gene (Morrissey et al. 2009).

### Specifically expressed genes in response to *H. armigera* between ED059 and cv. Tianlong 2

To learn the difference of defense mechanisms between ED059 and cv. Tianlong 2, we compared the expression of genes in response to *H. armigera* between ED059 and cv. Tianlong 2. The genes with twofold change and FDR (false discovery rate)  $\leq 0.001$  were defined as significantly differentially expressed genes (DEGs). The hierarchical clustering analyses for the DEGs between 1-L versus (vs.) 1L and 2-L vs. 2L, suggested an apparently different defense responses against *H. armigera* between ED059 and Tianlong 2 (Fig. 1a). In summary, 2213 specific DEGs with 417 up-regulated and 1796 down-regulated were identified in 2-L vs. 2L (Fig. 1b; Additional file 2). Interestingly, some genes involved in the JA pathway were found to be significantly down-regulated in ED059, which included *OPR3* (*Glyma17g38160.1*), *JAR1* (*Glyma19g44310.1*) and *MYC2* (*Glyma08g36720.2*, *Glyma01g15930.1*, *Glyma01g02251.1*, *Glyma06g35330.1*, and *Glyma08g28010.1*). In comparison, 2179 specific DEGs with 1374 up-regulated and 805 down-regulated were identified in 1-L vs. 1L (Fig. 1b; Additional file 3). One hundred and forty-eight up- and



**Fig. 1** Analyses of differentially expressed genes between ED059 and cv. Tianlong 2. The hierarchical clustering analyses (a) and Venn diagram (b) of the DEGs between 1-L vs. 1L and 2-L vs. 2L. Red and green indicate high and low expression levels, respectively. The upward arrow represents the up-regulations of gene expression. The downward arrow represents the down-regulations of gene expression. 1L, Tianlong 2 at 24 h without treatment; 1-L, Tianlong 2 at 24 h with treatment of *H. armigera*; 2L, ED059 at 24 h without treatment; and 2-L, ED059 at 24 h with treatment of *H. armigera*. (Colour figure online)

two hundred and fifty-two down-regulated genes were commonly identified in both of two comparisons (Fig. 1b; Additional file 4).

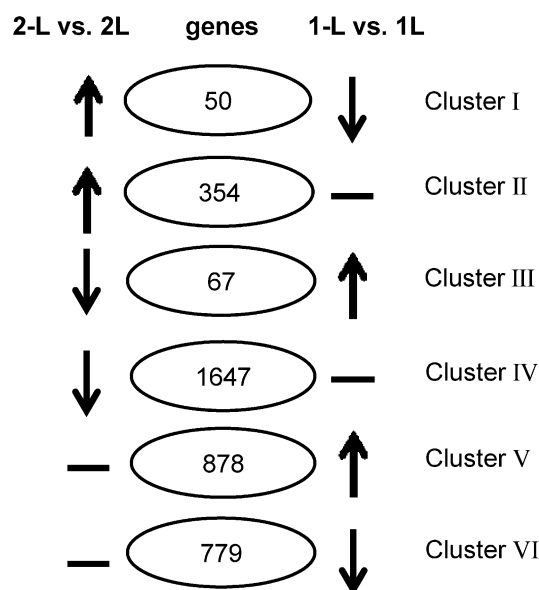
GO annotation by use of singular enrichment analysis (SEA) were conducted for three categories of genes (Du et al. 2010), which included: (I) specific differentially expressed genes in 2-L vs. 2L; (II) specific differentially expressed genes in 1-L vs. 1L; and, (III) common

differentially expressed genes between two comparisons. For the genes in the category I, 84 significant GOs in the biological processes were detected (Additional file 5). Of those, a large number of GOs were related to the regulation of metabolic processes (GO: 0009893, GO: 0009892, GO: 0031325, GO: 0031324, GO: 0010604, GO: 0010605, GO: 0051247 and GO: 0051248) (Additional file 5). For the genes in the category II, 52 significant GOs in the biological processes were detected, which were mainly related to response to stimulus (GO: 0050896), secondary metabolic process (GO: 0019748), cellular amino acid and derivative metabolic process (GO: 0006519), and cell wall macromolecule metabolic process (GO: 0044036) (Additional file 6). For the genes in the category III, 12 significant GOs were all related to response to stimulus (GO: 0050896) (Additional file 7).

Taking into account the expression pattern, we further used the method of parametric analysis of gene set enrichment (PAGE) to investigate the biological process for the genes between category I and II (Du et al. 2010). The results showed that the number of significant biological processes between category I and II were significantly reduced (Table 2). For category I, two significant biological processes were identified to be toxin metabolic process (GO: 0009404) and toxin catabolic process (GO: 0009407), which contained nine genes respectively (Table 2; Additional file 8). For category II, two significantly down-regulated biological processes were response to light intensity (GO: 0009642) and response to high light intensity (GO: 0009644), which contained 24 and 16 genes respectively (Table 2; Additional file 8). The biological processes between category I and II showed big differences, which indicated that there might be distinct defense mechanisms against *H. armigera* between ED059 and cv. Tianlong 2.

### The genes that were up-regulated in ED059 but down-regulated or not changed in cv. Tianlong 2 in response to *H. armigera*

To further learn the difference of defense mechanism, we analyzed the genes with different expression patterns in response to *H. armigera* between ED059 and cv.



**Fig. 2** Genes with different expression patterns between ED059 and cv. Tianlong 2. The upward arrow represents the up-regulations of gene expression. The downward arrow represents the down-regulations of gene expression. The horizontal line represents no changes of gene expression. 1L, Tianlong 2 at 24 h without treatment; 1-L, Tianlong 2 at 24 h with treatment of *H. armigera*; 2L, ED059 at 24 h without treatment; and 2-L, ED059 at 24 h with treatment of *H. armigera*

Tianlong 2. The genes with different expression patterns were grouped into six clusters (Fig. 2), which included: (I) 50 genes significantly up-regulated in 2-L vs. 2L while down-regulated in 1-L vs. 1L (Additional file 9); (II) 354 genes significantly up-regulated in 2-L vs. 2L while not changed in 1-L vs. 1L (Additional file 10); (III) 67 genes significantly down-regulated in 2-L vs. 2L while up-regulated in 1-L vs. 1L (Additional file 11); (IV) 1647 genes significantly down regulated in ED059 while not changed in 1-L vs. 1L (Additional file 12); (V) 878 genes not changed in 2-L vs. 2L while significantly up-regulated in 1-L vs. 1L (Additional file 13); and, (VI) 779 genes not changed in 2-L vs. 2L while significantly down-regulated in 1-L vs. 1L (Additional file 14).

**Table 2** The biological processes identified by use of the method of parametric analysis of gene set enrichment (PAGE) for the genes in category I and II

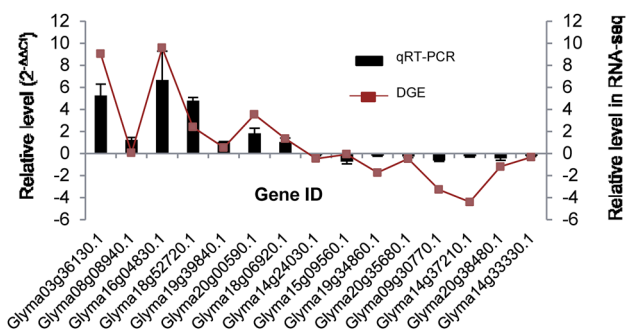
Category	Go term	Description	Gene number	Z-score	FDR
I	GO:0009404	Toxin metabolic process	9	3.8	0.045
	GO:0009407	Toxin catabolic process	9	3.8	0.045
II	GO: 0009642	Response to light intensity	24	-4.6	0.002
	GO: 0009644	Response to high light intensity	16	-4.5	0.002

(I) Specific differentially expressed genes in 2-L vs. 2L; (II) specific differentially expressed genes in 1-L vs. 1L. 1L, Tianlong 2 at 24 h without treatment; 1-L, Tianlong 2 at 24 h with treatment of *H. armigera*; 2L, ED059 at 24 h without treatment; and 2-L, ED059 at 24 h with treatment of *H. armigera*

The genes in cluster I and II were speculated to be involved in the pathways that were activated in ED059 while suppressed or not changed in cv. Tianlong 2. For the 50 genes in cluster I, most of them were associated with metabolic process such as flavonoid biosynthesis (*Glyma06g03410.1*, *Glyma18g03020.1*, *Glyma18g06510.4* and *Glyma18g12210.1*), phytohormone biosynthesis (*Glyma04g03740.1* and *Glyma01g29930.1*), lipid metabolism (*Glyma18g13540.2*, *Glyma02g43440.1*, *Glyma09g37430.1*, *Glyma12g30870.1*, and *Glyma06g23560.1*) (Additional file 9). For the 354 genes in cluster II, some of them encode transcription factors (e.g. *Glyma18g26600.1*, *Glyma06g24730.1*, *Glyma11g01850.1* and *Glyma07g05351.1*), phytohormone-related genes (e.g. *Glyma15g08540.2*, *Glyma15g13320.1*, *Glyma19g28250.1*, *Glyma02g41280.1*, *Glyma17g35610.1*, *Glyma03g30460.1* and *Glyma11g21650.1*) and secondary metabolism associated genes (e.g. *Glyma15g06730.1*, *Glyma09g03270.1* and *Glyma05g25460.1*) (Additional file 10). The involved pathways for these genes might play important roles in conferring resistance to *H. armigera* in ED059.

### qRT-PCR confirmation of selected gene expressions

To validate the transcript profiles produced in this study, 15 genes were randomly selected for qRT-PCR analysis (Fig. 3; Additional file 15). Three housekeeping genes, *HDC*, *EF1b*, and *UKN2*, were used as reference genes to normalize the gene expression levels (Wang et al. 2015a, b). The expression patterns detected by qRT-PCR for these 15 genes were consistent with those in the profilings, which indicated our RNA-seq data were reliable (Fig. 3).

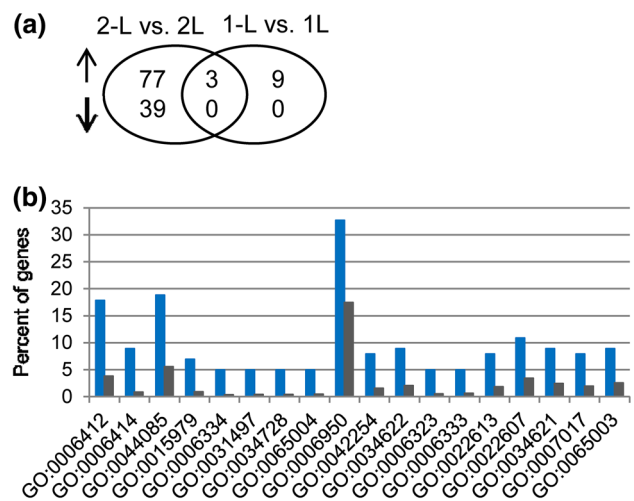


**Fig. 3** Real-time PCR validation of selected genes. The abscissa represents the name of selected genes; the *left ordinate* represents the relative expression levels of the selected genes, which were quantified by the  $2^{-\Delta\Delta Ct}$  method; the *right ordinate* represents the relative expression levels of the selected genes (DEGs), which were calculated by the absolute value of  $\log_2$  ratio

### Analysis of proteome in response to *H. armigera* between ED059 and cv. Tianlong 2

To learn the changes of protein expressions in response to *H. armigera* between ED059 and cv. Tianlong 2, we conducted the analysis of proteome for the samples, 2-L, 2L, 1-L and 1L, by use of isoaric tag for relative and absolute quantitation (iTRAQ) technology. Two biological replicates were performed for each sample and a high level of reproducibility in the iTRAQ analysis was observed between the two biological replicates of each sample (Additional file 16).

In summary, a total of 116 specifically expressed proteins with 77 up-regulated and 39 down-regulated, were significantly identified in a comparison of 2-L vs. 2L (Fig. 4a; Additional file 17). For the comparison of 1-L vs. 1L, 9 specific proteins were identified to be up-regulated, while no down-regulated proteins were detected (Fig. 4a; Additional file 18). Three proteins were commonly found in both of these two comparisons, all of which were up-regulated (Fig. 4a; Additional file 19). GO analysis showed that the specific expressed proteins in 2-L vs. 2L were mainly enriched in 18 significant GOs, of which the response to stress (GO: 0006950) contained the largest number of genes (Fig. 4b; Additional file 20). For the specific expressed proteins in 1-L vs. 1L, no significant GO was identified.



**Fig. 4** The differentially expressed proteins in response to *H. armigera* between ED059 and cv. Tianlong 2. **a** Venn diagram for the differentially expressed proteins between 2-L vs. 2L and 1-L vs. 1L; **b** Histogram of gene ontology classifications for the differentially expressed proteins. 1L, Tianlong 2 at 24 h without treatment; 1-L, Tianlong 2 at 24 h with treatment of *H. armigera*; 2L, ED059 at 24 h without treatment; and 2-L, ED059 at 24 h with treatment of *H. armigera*



## Functional confirmation of a potential resistant gene

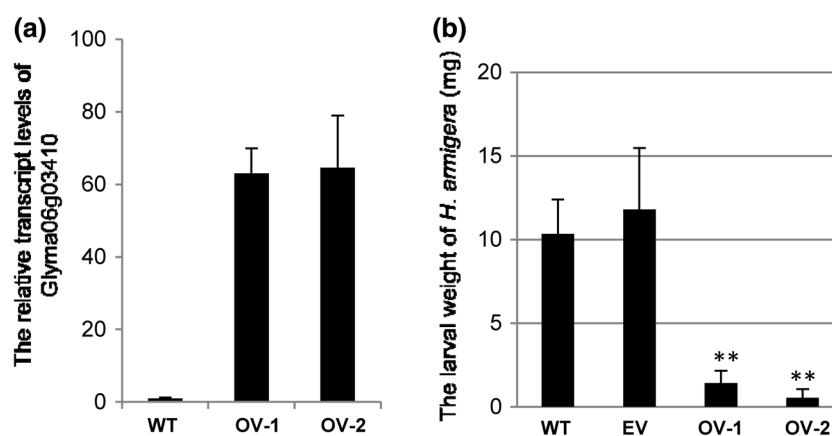
Because of our interests about the roles of phenylpropanoids in resistance to chewing insects in plants, we selected the gene *Glyma06g03410* that was predicted to be involved in the phenylpropanoid pathway for the functional analysis in the next step. *Glyma06g03410* encoded a putative phenylcoumaran benzylic ether reductase (PCBER), an enzyme responsible for the synthesis of lignans in the phenylpropanoid pathway (Karamloo et al. 2001; Nuoendagula et al. 2016; Vander et al. 2000). The expression of *Glyma06g03410* were up-regulated by 3.0 fold change in response to *H. armigera* in ED059 while down-regulated by 3.0 fold change in cv. Tianlong 2 (Additional file 9). Whole genome resequencing showed that, in comparison with Tianlong 2, there were one nucleotide deletion in the promoter region of *Glyma06g03410* and six point mutations in the gene region of *Glyma06g03410* in ED059, among which three point mutations were located in the introns and the other three were located in the exons (Additional file 21). The point mutations in the exons of *Glyma06g03410* in ED059 did not cause amino acids changes. No copy number variations were found for *Glyma06g03410* between ED059 and cv. Tianlong 2. We overexpressed *Glyma06g03410* in *Arabidopsis* and used two independent transgenic lines for the evaluation of gene function (Fig. 5a). The newly hatched *H. armigera* were placed on the transgenic plants and controls. And, the weights of *H. armigera* were determined at 6 days after treatment. The results showed that the *H. armigera* feeding on the transgenic plants gained less weight than the ones feeding on the controls (Fig. 5b), suggesting the role

of *Glyma06g03410* as a resistant gene. The functional confirmation of *Glyma06g03410* showed that our comparative analysis of transcriptome between resistant and susceptible responses was a feasible strategy of mining the herbivore-resistant genes in ED059.

## Discussion

### The possible mechanism of resistance to *H. armigera* in ED059

A wild soybean line ED059, was identified to have the highest level of resistance to chewing insects among more than one thousand soybean germplasm in the field. With the aim of mining useful genes and elucidating the resistant mechanism, we conducted the comparison of transcriptome and proteome in response to *H. armigera* between ED059 and the susceptible cv. Tianlong 2. The hierarchical clustering analyses of gene expressions showed that there were two obviously different molecular responses to *H. armigera* between ED059 and cv. Tianlong 2 (Fig. 1). In ED059, a large number of specifically expressed genes were significantly enriched in the metabolic process rather than the response to stimulus that were detected for the specifically expressed genes in the susceptible cv. Tianlong 2 (Additional file 5 and 3). Since the common regulated genes were significantly enriched in response to stresses in both ED059 and cv. Tianlong 2, we speculated that the divergence of downstream metabolic pathways might be responsible for the differences of phenotypes between these two soybean accessions. Four hundred and four genes that



**Fig. 5** Functional confirmation of *Glyma06g03410* in *Arabidopsis*. **a** Mean  $\pm$  SE of the relative transcript levels of *Glyma06g03410* in the transgenic plants (n=3; three independent biological replicates were used per genotype). The transcript levels of the selected genes in related transgenic plants were calculated by the  $2^{-\Delta\Delta Ct}$  method. Data are the means  $\pm$  SE of two or three strains. **b** *H. armigera* larval

performance on *A. thaliana* plants overexpressing *Glyma06g03410*. Mean  $\pm$  SE of *H. armigera* larval mass after 6 days of feeding on wild-type (WT) plants and transgenic plants with the empty vector (EV) and *Glyma06g03410* (OV-1 and OV-2) (mean  $\pm$  SEM; n=5; statistics by *t* test; \*\*p<0.01)

were up-regulated in response to *H. armigera* in ED059 while down-regulated or not changed in cv. Tianlong 2, were speculated to play important roles in conferring resistance to *H. armigera* in ED059. These genes might provide useful resources for the identification of resistant genes, which was suggested by the functional analysis of *Glyma06g03410* by use of transgenic *Arabidopsis* (Fig. 5). Future studies should testify the enzyme activity of *Glyma06g03410* as a PCBER and investigate the roles of its metabolic products in defense against herbivores. More changed genes related to metabolism were also worthy of being further studied in order to elucidate the roles of metabolic pathway in resistance to *H. armigera* in ED059.

### Genes related to phytohormones between resistant and susceptible responses

JA was known to be an important phytohormone that mediated herbivore defense in plants (Gilardoni et al. 2011; Van Poecke 2007). However, in this study, we found that some genes such as *OPR3* (*Glyma17g38160.1*), *JAR1* (*Glyma19g44310.1*) and *MYC2* (*Glyma08g36720.2*, *Glyma01g15930.1*, *Glyma01g02251.1*, *Glyma06g35330.1*, and *Glyma08g28010.1*) in the JA pathway, were significantly down-regulated in response to *H. armigera* in the resistant ED059 in comparison with the susceptible cv. Tianlong 2 (Additional file 2). These results indicated that the JA-regulated defense against *H. armigera* might be suppressed in ED059. Because only one time point, 24 h after treatment, was adopted, we had no idea about the earlier gene expression in response to *H. armigera* between these two soybean accessions. Thus, we could not exclude the possibility that the JA pathway might be activated earlier in ED059 than that in cv. Tianlong 2, which resulted in a quick defense response to *H. armigera* in ED059. The exact role of the JA pathway in defense response between ED059 and cv. Tianlong 2 need to be investigated further.

SA was an important signal factor that is involved in the local- and systemic-induced defense response against pathogens (Glazebrook 2005). However, the role of SA in chewing-insect defense is still unclear. Previous studies found that SA could inhibit the expression of some anti-herbivore related genes in potato (Sivasankar et al. 2000), and the decrease of SA content in response to *Manduca sexta* in *Nicotiana attenuata* was correlated to the induction of defense responses (Gilardoni et al. 2011). Interestingly, a minor increase of endogenous SA content in response to *M. sexta* in *N. attenuata* was also reported, suggesting a complicated role of SA in herbivory defense responses (Wu et al. 2007). In our studies, the SA responsive gene, *GmPRI* (*Glyma15g06780.1*), was induced in response to *H. armigera* in both cv. Tianlong 2 and ED059 (Additional file 4), which indicated that the SA pathway were

commonly activated in these two accessions and might not be responsible for the resistance in ED059.

### Genes related to metabolism between resistant and susceptible responses

Polyphenol oxidase (PPO) can alkylate the essential amino acids and reduce the nutritional values of plant proteins (Van Poecke 2007). Overproduction of PPO could enhance the ability of transgenic plants to defend themselves against herbivores (Wang and Constabel 2004). In plants, lipoxygenase (LOXs) not only participated in the JA-dependent defense pathway, but also directly fought against herbivores by producing peroxides and oxylipin products (Zhu-Salzman et al. 2008). In ED059, three PI-encoding genes including *Glyma18g46580.1*, *Glyma14g04250.1* and *Glyma03g37260.1*, two PPO-encoding genes including *Glyma04g14361.1* and *Glyma13g25181.1*, and two LOX-encoding genes including *Glyma07g03920.2* and *Glyma07g00900.1*, were specifically up-regulated in response to *H. armigera* in ED059 (Additional file 2), which might play important roles in conferring resistance to *H. armigera*.

Glutathione S-transferases (GSTs) are ubiquitous in different types of plant organs, which were involved in the detoxification and processing of various xenobiotics but also endogenous toxic and non-toxic metabolites (Dixon et al. 2002). The expression of GSTs can be induced by a range of abiotic and biotic stresses such as drought, salt, cold, heavy metals, pathogens, and elicitors (Dixon et al. 2002). PAG analysis showed that two significant biological processes, toxin metabolic process (GO: 0009404) and toxin catabolic process (GO: 0009407), were specifically detected in 2-L vs. 2L, in which 5 GSTs with 3 up-regulated and 2 down-regulated were identified (Additional file 8). It is worthy of figuring out the targets of these GSTs in further studies.

### Comparison of transcriptome and proteome

In this study, we used RNA-seq and iTAQ technologies to conduct the analyses of transcriptome and proteome respectively. In comparison with the number of DEGs detected by RNA-seq, the number of DEGs detected by iTAQ was significantly reduced (Figs. 1b, 4a). In addition, some genes were found to have the opposite expression patterns between transcript and protein levels (Additional file 22). For example, photosystem I *psaG/psaK* (*Glyma5g22780.2*) and thioredoxin (*Glyma12g35190.1*) were down-regulated at the transcript level while up-regulated at protein level in 2-L vs. 2L (Additional file 2 and 17). This discrepancy between transcript and protein levels were also reported in other studies (Botella et al. 1996; Liang et al. 2016; Zhao

et al. 1996), which might be attributed to the limited capability of iTAQ technology for the detection of small differences in protein abundance, or the posttranscriptional or posttranslational modification of proteins.

In this paper, we investigated the mechanism of resistance to *H. armigera* in a wild soybean line ED059 through comparative analyses of transcriptome and proteome between ED059 and a susceptible cv. Tianlong 2. The results showed that the altered metabolism might be responsible for the resistance to *H. armigera* in ED059. Our studies provided us useful resources for identifying genes with resistance to *H. armigera* in soybean and also improve our understanding of plant defense mechanism against insects.

## Methods

### Samples preparation

The soybean cultivar (cv.), Tianlong 2, and the wild soybean, ED059, were obtained from the Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, China. Seeds were pre-germinated on moistened filter paper in a plant growth chamber at 27 °C, 85% ambient humidity and a 16:8 (light:dark) photoperiod for 3–4 days. The seedlings were then transferred into 18-cm×18-cm individual plastic pots with nutrient-rich soil (Pindstrup Substrate, Denmark) and vermiculite at a ratio of 2:3 at 27 °C under 16 h of light. All plants used in the experiments had three fully expanded trifoliolate leaves. For *H. armigera* treatment, the second fully expanded leaves were infected with three third-instar larvae of *H. armigera*. After 24 h, the wounded local leaves were harvested, flash frozen in liquid nitrogen, and stored at –80 °C until RNA isolation. Samples from untreated plants were used as controls. For each treatment, we used five leaves which were from five independent plants respectively. For biological replicates, all treatments and samples were set up and collected independently following the same method described above.

### Solexa/Illumina sequencing and iTRAQ-based proteome analysis

At least 10 µg of total RNA ( $\geq 300$  ng/µL) from each sample was extracted for RNA-seq. Sample preparation, library construction, sequencing and bioinformatics analysis were performed as described previously (Hao et al. 2011). The samples were sequenced by an Illumina HiSeq2000 platform. Sequencing-received raw image data were transformed by base calling into raw sequences. The raw sequences were transformed into clean tags by

removing the 3' adaptor sequences and low quality tags. A virtual library containing all the possible CATG+17 bases length sequences was generated using the SoyBase soybean genomics database. All clean tags were mapped to these reference sequences, and those tags with no more than one mismatch were considered.

The plant materials used in DGEseq were also used for proteome quantitation analysis. In addition to that, one more biological replicate was also set up. Thus, the samples from two biological replicates were used for proteome analyses. Total protein (100 µg) was extracted from each sample solution, and the proteins were digested with Trypsin Gold (Promega, Madison, WI, USA) at a protein:trypsin ratio of 30:1 at 37 °C for 16 h, according to a previously reported method (Lan et al. 2012). After trypsin digestion, peptides were labeled with the iTRAQ tags according to the manufacturer's protocol for an 8-plex iTRAQ reagent (Applied Biosystems, Foster City, CA, USA). Subsequently, the labeled peptide mixtures were pooled and pre-separated by strong cation exchange chromatography with the LC-20AB HPLC pump system (Shimadzu, Kyoto, Japan). iTRAQ analysis was performed on a TripleTOF 5600 system (AB SCIEX, Concord, ON, Canada) combined with a Famos autosampler (LC Packings) and LC20-AD Nano HPLC (Shimadzu, Kyoto, Japan), according to a previously reported study (Du et al. 2014). Proteome Discoverer software (ver.1.2.0.339; Thermo Fisher Scientific, San Jose, CA, USA) was used to transform raw data files into MGF files. Protein identification and quantitation for iTRAQ was performed using the Mascot search engine (ver. 2.3.0; Matrix Science, London, U.K.) against data in the Soybase database.

### qRT-PCR

Total RNA was extracted using Trizol reagent (Sigma). Total RNA (2 µg) was reverse transcribed to cDNA using Superscript II reverse transcriptase (Invitrogen). qRT-PCR analyses were performed with the SuperReal PreMix (SYBR Green, Tiangen) using a Rotor-Gene Q (Qiagen) real-time PCR system. The amplification program was initiated with a denaturation step at 95 °C for 15 min; followed by 40 cycles of 95 °C for 10 s, 60 °C for 20 s, and 72 °C for 30 s. The samples for qRT-PCR were from additional three biological replicates, which were different from the ones for RNA-seq. Relative quantification of the expression levels of genes was performed by the  $2^{-\Delta\Delta CT}$  method (Livak and Schmittgen 2001). Three soybean genes, namely, *HDC*, *EF1b*, and *UKN2*, were used as the internal controls to normalize the expression levels of the selected genes (Wang et al. 2015a, b).



## Whole-genome sequencing and copy number variation analysis

Whole-genome sequencing of ED059 and cv. Tianlong 2 was conducted using Illumina technology at the BIOMARKER company, Beijing, China, following a described protocol (Jiao et al. 2015). The sequencing depth for each sample was about 50× coverage. The CNV analysis was conducted using CNV-seq software (Xie and Tammi 2009).

## Plasmid construction and transgenic confirmation

The PCR products of *Glyma06g03410* were digested with *XcmI* and inserted into plasmid PCX-DG, as previously described (Chen et al. 2009). The transgenic *Arabidopsis* were obtained by use of the *Agrobacterium*-mediated floral dip method (Bent 2006). The empty vector (EV) was transformed into the plants as a control. The T<sub>2</sub> plants with a segregation ratio of 3:1 (the number of positive plants: the number of negative plants), were propagated, and the homozygous lines were used for the further studies. The evaluation of transgenic plants for the resistance to *H. armigera* was performed in a growth chamber at 27 °C, 85% ambient humidity, and a 16:8 (L:D) photoperiod. The eggs of *H. armigera* were hatched at 27 °C. One freshly hatched larva of *H. armigera* was placed in a Petri dish (100×25 mm) and starved for 12 h before being used in the test. Five mature rosette leaves from 4-week-old plants were collected and placed in each dish. The leaves were changed every 2 days (days), and the test was stopped at 6 days after treatment. The larval weight from each Petri dish were measured and recorded at 2, 4, and 6 days after treatment. The dish experiment was conducted in a randomized complete block design with 10 replications.

**Acknowledgements** We thank Wei Yan and Zechuan Lin, Peking University, for the advices of data analyses. This work was supported by National Key Research Plan of China (2016YFD0100603), National Transgenic Project (2014ZX08004005, 2014ZX08004003, 2015ZX08004003 and 2016ZX08004003), National Natural Science Foundation of China (31371654, 31522042, 31501655 and 31501334), the Modern Agro-industry Technology Research System (CAAS-04-PS08), Agricultural Science and Technology Innovation Program, and Wuhan Chenguang Plan (2015070404010193).

**Author contributions** XZ and YJ conceived the experiments. XW performed the experiments. XW and YJ analyzed the data and wrote the manuscript. HC, ZS, XS, JL, BD, CZ, ZY, XZ, DQ, SC helped with some experiments. All authors reviewed the manuscript.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interests.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Bent A (2006) *Arabidopsis thaliana* floral dip transformation method. *Methods Mol Biol* 343:87–103
- Botella MA, Xu Y, Prabha TN, Zhao Y, Narasimhan ML, Wilson KA, Nielsen SS, Bressan RA, Hasegawa PM (1996) Differential expression of soybean cysteine proteinase inhibitor genes during development and in response to wounding and methyl jasmonate. *Plant Physiol* 112:1201–1210
- Chen S, Songkumarn P, Liu J, Wang GL (2009) A versatile zero background T-vector system for gene cloning and functional genomics. *Plant Physiol* 150:1111–1121
- Chen L, Jiang B, Wu C, Sun S, Hou W, Han T (2015) The characterization of *GmTIP*, a root-specific gene from soybean, and the expression analysis of its promoter. *Plant Cell Tissue Organ Cult* 121:259–274. doi:10.1007/s11240-014-0682-2
- Dixon DP, Davis BG, Edwards R (2002) Functional divergence in the glutathione transferase superfamily in plants. Identification of two classes with putative functions in redox homeostasis in *Arabidopsis thaliana*. *J Biol Chem* 277:30859–30869
- Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res* 38:W64–W70
- Du C, Liang JR, Chen DD, Xu B, Zhuo WH, Gao YH, Chen CP, Bowler C, Zhang W (2014) iTRAQ-based proteomic analysis of the metabolism mechanism associated with silicon response in the marine diatom *Thalassiosira pseudonana*. *J Proteome Res* 13:720–734
- Ferry N, Edwards MG, Gatehouse JA, Gatehouse AM (2004) Plant-insect interactions: molecular approaches to insect resistance. *Curr Opin Biotechnol* 15:155–161
- Fraser CM, Chapple C (2011) The phenylpropanoid pathway in *Arabidopsis*. *Arabidopsis Book* 9:e0152
- Gatehouse JA (2002) Plant resistance towards insect herbivores: a dynamic interaction. *New Phytol* 156:25
- Gilardoni PA, Hettenhausen C, Baldwin IT, Bonaventure G (2011) *Nicotiana attenuata* LECTIN RECEPTOR KINASE1 suppresses the insect-mediated inhibition of induced defense responses during *Manduca sexta* herbivory. *Plant Cell* 23:3512–3532
- Glazebrook J (2005) Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu Rev Phytopathol* 43:205–227
- Guan R, Qu Y, Guo Y, Yu L, Liu Y, Jiang J, Chen J, Ren Y, Liu G, Tian L, Jin L, Liu Z, Hong H, Chang R, Gilliam M, Qiu L (2014) Salinity tolerance in soybean is modulated by natural variation in *GmSALT3*. *Plant J* 80:937–950
- Gunadi A, Rushton PJ, McHale LK, Gutek AH, Finer JJ (2016) Characterization of 40 soybean (*Glycine max*) promoters, isolated from across 5 thematic gene groups. *Plant Cell Tissue Organ Cult* 127:145–160. doi:10.1007/s11240-016-1038-x
- Hao QN, Zhou XA, Sha AH, Wang C, Zhou R, Chen SL (2011) Identification of genes associated with nitrogen-use efficiency by genome-wide transcriptional analysis of two soybean genotypes. *BMC Genomics* 12:525
- Heil M (2008) Indirect defence via tritrophic interactions. *New Phytol* 178:41–61

- Jiao Y, Vuong TD, Liu Y, Meinhardt C, Liu Y, Joshi T, Cregan PB, Xu D, Shannon JG, Nguyen HT (2015) Identification and evaluation of quantitative trait loci underlying resistance to multiple HG types of soybean cyst nematode in soybean PI 437655. *Theor Appl Genet* 128:15–23
- Johnson ET, Dowd PF (2004) Differentially enhanced insect resistance, at a cost, in *Arabidopsis thaliana* constitutively expressing a transcription factor of defensive metabolites. *J Agric Food Chem* 52:5135–5138
- Karamloo F, Wangorsch A, Kasahara H, Davin LB, Hausteine D, Lewis NG, Vieths S (2001) Phenylcoumaran benzylic ether and isoflavonoid reductases are a new class of cross-reactive allergens in birch pollen, fruits and vegetables. *Eur J Biochem* 268:5310–5320
- Kessler A, Baldwin IT (2002) Plant responses to insect herbivory: the emerging molecular analysis. *Annu Rev Plant Biol* 53:299–328
- Lan P, Li W, Schmidt W (2012) Complementary proteome and transcriptome profiling in phosphate-deficient *Arabidopsis* roots reveals multiple levels of gene regulation. *Mol Cell Proteomics* 11:1156–1166
- Li YH, Zhou G, Ma J, Jiang W, Jin LG, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang SS, Zuo Q, Shi XH, Li YF, Zhang WK, Hu Y, Kong G, Hong HL, Tan B, Song J, Liu ZX, Wang Y, Ruan H, Yeung CK, Liu J, Wang H, Zhang LJ, Guan RX, Wang KJ, Li W, Chen SY, Chang RZ, Jiang Z, Jackson SA, Li R, Qiu LJ (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat Biotechnol* 32:1045–1052
- Liang C, Cheng S, Zhang Y, Sun Y, Fernie AR, Kang K, Panagiotou G, Lo C, Lim BL (2016) Transcriptomic, proteomic and metabolic changes in *Arabidopsis thaliana* leaves after the onset of illumination. *BMC Plant Biol* 16:43
- Livak K, Schmittgen T (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25:402–408
- Morrissy AS, Morin RD, Delaney A, Zeng T, McDonald H, Jones S, Zhao Y, Hirst M, Marra MA (2009) Next-generation tag sequencing for cancer gene expression profiling. *Genome Res* 19:1825–1835
- Nuoendagula KN, Mori T, Nakabayashi R, Tsuji Y, Hishiyama S, Saito K, Masai E, Kajita S (2016) Expression and functional analyses of a putative phenylcoumaran benzylic ether reductase in *Arabidopsis thaliana*. *Plant Cell Rep* 35:513–526
- Qi X, Li MW, Xie M, Liu X, Ni M, Shao G, Song C, Kay-Yuen YA, Tao Y, Wong FL, Isobe S, Wong CF, Wong KS, Xu C, Li C, Wang Y, Guan R, Sun F, Fan G, Xiao Z, Zhou F, Phang TH, Liu X, Tong SW, Chan TF, Yiu SM, Tabata S, Wang J, Xu X, Lam HM (2014) Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. *Nat Commun* 5:4340
- Sivasankar S, Sheldrick B, Rothstein SJ (2000) Expression of allene oxide synthase determines defense gene activation in tomato. *Plant Physiol* 122:1335–1342
- Stam JM, Kroes A, Li Y, Gols R, van Loon JJ, Poelman EH, Dicke M (2014) Plant interactions with multiple insect herbivores: from community to genes. *Annu Rev Plant Biol* 65:689–713
- Stewart CN, Adang MJ, All JN, Boerma HR, Cardineau G, Tucker D, Parrott WA (1996) Genetic transformation, recovery, and characterization of fertile soybean transgenic for a synthetic *Bacillus thuringiensis* cryIAc gene. *Plant Physiol* 112:121–129
- Van Poecke RM (2007) *Arabidopsis*-insect interactions. *Arabidopsis Book* 5:e0107
- Vander Mijnsbrugge K, Beeckman H, De Rycke R, Van Montagu M, Engler G, Boerjan W (2000) Phenylcoumaran benzylic ether reductase, a prominent poplar xylem protein, is strongly associated with phenylpropanoid biosynthesis in lignifying cells. *Planta* 211:502–509
- Vogt T (2010) Phenylpropanoid biosynthesis. *Mol Plant* 3:2–20
- Wang J, Constabel CP (2004) Polyphenol oxidase overexpression in transgenic *Populus* enhances resistance to herbivory by forest tent caterpillar (*Malacosoma disstria*). *Planta* 220:87–96
- Wang Z, Huang W, Chang J, Sebastian A, Li Y, Li H, Wu X, Zhang B, Meng F, Li W (2014) Overexpression of *SiDGATI*, a gene encoding acyl-CoA:diacylglycerol acyltransferase from *Sesamum indicum* L. increases oil content in transgenic *Arabidopsis* and soybean. *Plant Cell Tissue Organ Cult* 119:399–410. doi:10.1007/s11240-014-0543-z
- Wang X, Chen H, Sha A, Zhou R, Shan Z, Zhang X, Zhang C, Chen L, Hao Q, Yang Z, Qiu D, Chen S, Zhou X (2015a) Laboratory testing and molecular analysis of the resistance of wild and cultivated soybeans to cotton bollworm, *Helicoverpa armigera* (Hübner). *Crop J* 3:10
- Wang Y, Wang H, Ma Y, Yang W, Yang Q, Yu D (2015b) Identification of soybean herbivory-regulated genes and a transgenic investigation of their potential in insect resistance. *Plant Cell Tissue Organ Cult* 123:321–340. doi:10.1007/s11240-015-0837-9
- Winter SM, Shelp BJ, Anderson TR, Welacky TW, Rajcan I (2007) QTL associated with horizontal resistance to soybean cyst nematode in *Glycine soja* PI464925B. *Theor Appl Genet* 114:461–472
- Wu J, Baldwin IT (2009) Herbivory-induced signalling in plants: perception and action. *Plant Cell Environ* 32:1161–1174
- Wu J, Hettnerhausen C, Meldau S, Baldwin IT (2007) Herbivory rapidly activates MAPK signaling in attacked and unattacked leaf regions but not between leaves of *Nicotiana attenuata*. *Plant Cell* 19:1096–1122
- Xie C, Tammi MT (2009) CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinform* 10:80
- Zavala JA, Baldwin IT (2004) Fitness benefits of trypsin proteinase inhibitor expression in *Nicotiana attenuata* are greater than their costs when plants are attacked. *BMC Ecol* 4:11
- Zhai Y, Shao S, Sha W, Zhao Y, Zhang J, Ren W, Zhang C (2017) Overexpression of soybean *GmERF9* enhances the tolerance to drought and cold in the transgenic tobacco. *Plant Cell Tissue Organ Cult* 128:607–618. doi:10.1007/s11240-016-1137-8
- Zhao Y, Botella MA, Subramanian L, Niu X, Nielsen SS, Bressan RA, Hasegawa PM (1996) Two wound-inducible soybean cysteine proteinase inhibitors have greater insect digestive proteinase inhibitory activities than a constitutive homolog. *Plant Physiol* 111:1299–1306
- Zhu-Salzman K, Luthe DS, Felton GW (2008) Arthropod-inducible proteins: broad spectrum defenses against multiple herbivores. *Plant Physiol* 146:852–858