Rothamsted Repository Download

A - Papers appearing in refereed journals

The regulation of three new members of the cytochrome P450 *CYP6* family and their promoters in the cotton aphid *Aphis gossypii* by plant allelochemicals

Fen Li,[a,b] Kangsheng Ma,[a] Xuewei Chen,[a] Jing-Jiang Zhou[b]* and Xiwu Gao[a]*

Abstract

BACKGROUND: The expression of P450 genes in insects can be induced by plant allelochemicals. To understand the induction mechanisms, we measured the expression profiles of three P450 genes and their promoter activities under the induction of plant allelochemicals.

RESULTS: The inducible expression of *CYP6CY19* was the highest among three genes, followed by those of *CYP6CY22* and *CYP6DA1*. The regions from −687 to +586 bp of *CYP6DA1*, from −666 to +140 bp of *CYP6CY19* and from −530 to +218 bp of *CYP6CY22* were essential for basal transcriptional activity. The *cis*-elements for plant allelochemicals induction were identified between −193 and +56 bp of *CYP6DA1*, between −157 and +140 bp of *CYP6CY19* and between −108 and +218 bp of *CYP6CY22*. These promoter regions were found to contain a potential aryl hydrocarbon receptor element binding site with a conservative sequence motif 5′-C/TAC/ANCA/CA-3′. All these four plant allelochemicals were able to induce the expression of these P450 genes. Tannic acid had a better inductive effect than other three plant allelochemicals.

CONCLUSIONS: Our study identified the plant allelochemical responsive *cis*-elements. This provides further research targets aimed at understanding the regulatory mechanisms of P450 genes expression and their interactions with plant allelochemicals in insect pests.
© 2018 Society of Chemical Industry

Supporting information may be found in the online version of this article.

Keywords: *Aphis gossypii*; *CYP6DA1*; *CYP6CY19*; *CYP6CY22*; plant allelochemicals; responsive *cis*-element

1 INTRODUCTION

Insect pests have the ability to utilize some cultivated plants as their host such as cotton, potato, eggplant, pepper/sweet-pepper or strawberry. They have developed sophisticated defense systems such as cytochrome P450 enzymes against plant toxic compounds.[1] In insects, cytochrome P450 enzymes can catalyze the metabolism of physiologically endogenous compounds but are best known for their roles in the detoxification of xenobiotics, such as pesticides, plant allelochemicals and anthropogenic pollutants. It was found that many xenobiotics can induce the transcriptional expression of P450 genes and thereby increase the levels of P450s activity.[2–4] Increased insecticide detoxification mediated by over-expressed P450s is a common mechanism of insecticide resistance.[5] There were a large number of studies which focused on the insects P450 inducibility. In the review of Liu et al., it was reported that many P450s were associated with insecticide resistance, especially the members of the *CYP6* subfamily.[6] Genes in the *CYP6* subfamily have been studied extensively in insect species. For example, Mao et al. first found *CYP6AE14* expression could be induced by gossypol in the cotton bollworm *Helicoverpa armigera*.[7] Li et al. found that *CYP6B6* of *H. armigera* could be up-regulated by 2-tridecanone.[8] However, little is known about the regulatory mechanisms of *CYP6* P450 genes expression such as the core promoter sequences and *cis*-regulatory elements responsive to plant allelochemicals. Further knowledge in P450 gene regulation may offer insights about how insect pests response to plant allelochemicals and useful information of their application in pest management.

The dual-luciferase assay has been widely used in cell lines to rapidly and accurately determine the promoter activity of a given gene.[9] For example, Zhao et al. found the core promoter

---

* Correspondence to: XW Gao, Department of Entomology, China Agricultural University, Beijing 100193, P. R. China. E-mail: gaoxiwu@263.net.cn; or J-J Zhou, Department of Biointeractions and Crop Protection, Rothamsted Research, Harpenden AL5 2JQ, UK. E-mail: jjzhouchina@163.com

a Department of Entomology, China Agricultural University, Beijing, P. R. China

b Department of Biointeractions and Crop Protection, Rothamsted Research, Harpenden, UK

region of two *CYP9A* genes (*CYP9a19* and *CYP9a22*) in the silkworm *Bombyx mori* using such a method.[3] Zhao et al. showed that the region between −827 and −722 bp was essential for the basal and 20-hydroxyecdysone induced transcriptional activities of *CYP6ab4* gene, and demonstrated its mediation by the BR-C Z binding site.[9] In a study of *Drosophila melanogaster*, five *CYP6DA1* alleles were found in four different strains based on the presence or absence of an intact TF Nrf2/Maf binding site in the 5′-promoter core region and the changes in this binding site were associated with the promoter activity of *CYP6DA1* as well as the resistance to DDT.[10]

The cotton aphids (*Aphis gossypii* Glover; Hemiptera: Aphididae), one of the most economically important pests in agriculture, cause economic damage through direct feeding and virus transmission.[11] Our group previously sequenced the transcriptome and constructed four digital gene expression libraries of the cotton aphids after they were fed on four plant allelochemicals with a very low concentration. We found that the expression level of three *CYP6* genes were regulated by the plant allelochemicals.[12] Thus in the current study, we characterized these *CYP6* genes by molecular cloning, and detailed measuring of the effects of four plant allelochemicals (quercetin, 2-tridecanone, gossypol and tannic acid) on their expression at physiological concentrations. We reported the plant allelochemical responsive *cis*-elements of these P450 genes as well as a potential aryl hydrocarbon receptor (AhR) binding sites with consensus sequence as 5′-CA/GNNCAA-3′ in the responsive elements. These results will help to understand the regulation mechanisms by plant allelochemicals of P450 gene expression in insects.

2 MATERIALS AND METHODS

2.1 Insects and sample collection
In this research, the cotton aphids were originally colonized from single aphid collected in the cotton field of Yuncheng, Shanxi Province, China in 2014. In the laboratory, they were maintained on cotton seedlings (*Gossypium hirsutum* L.) at 20–23 °C, 60% relative humidity and 16 h:8 h light:dark photoperiod.

For the experiments, healthy and lively adult aphids were harvested and randomly divided into treatment groups and control groups. Each group had three repeats and each repeat had 100 single adults. Sterilized glass tubes (3 cm × 2 cm) with openings at both ends were used for plant allelochemicals induction. In detail, one end of each tube was covered by two layers of parafilm, and the solution (artificial diet added with plant allelochemicals) was sandwiched between the two parafilm layers, while another end was covered with Chinese art paper in order to prevent the aphids from escaping but allow them access to enough air. In the plant allelochemicals induction experiments, aphids were fed on 90 μL of 0.5 mol L^{-1} sterile sucrose solution supplemented with either 10 μL of quercetin (500 mg/L, Sigma-Aldrich, St Louis, MO, USA) or 2-tridecanone (500 mg/L, Solarbio, Beijing, China) or gossypol (500 mg/L, Sigma-Aldrich) or tannic acid (200 mg/L, Sigma-Aldrich). In the control group, only 10 μL solvent used for making plant allelochemical solutions (sterilization water or acetone) was used. After 48 h, all alive aphids from each treatment were collected for RNA extractions. All experiments were performed in triplicates during different generations of aphids.

2.2 Cell culture
The Sf9 cells of *Spodoptera frugiperda* were kindly provided by the Institute of Plant Protection of Chinese Academy of Agricultural Sciences and cultured in Sf-900 II SFM liquid media (Invitrogen, Waltham, MA, USA) supplemented with 10% fetal bovine serum (FBS) (Evergreen, Hangzhou, China) and 100 U/mL penicillin, 100 mg/mL streptomycin. Cells were maintained at 27 °C without carbon dioxide (CO_2) condition.

2.3 Total RNA extraction, cDNA synthesis and P450 genes full-length obtained
Total RNA was extracted by using the RNA-Solv Reagent (Omega Bio-tek, Norcross, GA, USA) according to the manufacturer's instructions and stored at −80 °C until use. The HiScript II Q RT SuperMix for quantitative polymerase chain reaction (qPCR) (Vazyme Biotech, Nanjing, China) was used in order to eliminate potential genomic DNA contamination.

The 3′- and 5′-RACE PCR were conducted by the SMARTer RACE 5′/3′ Kit (TaKaRa, Dalian, China) following its protocol. The genomic DNA of the cotton aphids was isolated from adults using a Cell/Tissue DNA Extraction Kit (Bio TeKe, Beijing, China). The concentration and purity of the isolated DNA were analyzed using Nanodrop 1000 Spectrophotometer.

2.4 Quantitative RT-PCR
For quantitative reverse transcription polymerase chain reaction (qRT-PCR), SYBR Green Master Mix (Vazyme Biotech) was used following its instruction manual. The qRT-PCR primers for *CYP6DA1*, *CYP6CY19*, *CYP6CY22* and the internal reference gene elongation factor 1 alpha (EF1-*a*) (Genebank: EU019874.1) were synthesized by BGI (Beijing, China) and the sequences are listed in Table 1. EF1-*a* was used as a reference gene to normalize the target gene expression levels among samples because of its stable characteristics demonstrated by Ma et al.[11]

The gene expression at the mRNA level was calculated using the $2^{-\Delta\Delta Ct}$ method.[13] All experiments were performed in three independent biological replicates and the reactions of each sample were carried out in triplicate.

2.5 The promoter of *CYP6DA1*, *CYP6CY19* and *CYP6CY22* obtained and the deletion fragments constructed
For the molecular cloning of full-length promoter regions, the Genome Walking Kit (TaKaRa) was used. Briefly, according to the user manual guidelines, three gene-specific primers in the same direction of the known 5″-UTR of *CYP6* genes were designed, then annealing with four kinds of merger primers provided by the kit. The PCR cycling parameters were for 'touchdown' PCR, which involved using an annealing/extension temperature several degrees higher than the Tm of the primers during the initial PCR cycles. For primary PCR, 1 μL of genomic DNA (500–1000 ng) was used. For secondary and third PCR, 1 μL of 50 × dilution of the primary PCR product was used. The final product can then be sequenced, cloned, and further analyzed.

To construct the deletion promoter fragments of each *CYP6* gene with gradually decreasing lengths of promoter regions, the longest promoter fragment was used as template and amplified using PCR with several interval primers as listed in Table 2. The PCR products were digested by either *EcoR*V or *Hind*III and *Sac*I restriction enzymes and subcloned into the promoterless pGL4.11 [luc2P] vector (Promega, Madison, WI, USA) which was digested by the same enzymes, yielding pGL4-report-promoter deletion constructs. All sequences of the cloned promoter fragments were confirmed by sequencing (BGI). In this paper, the transcriptional start site (TSS) of each gene was predicted by Promoter 2.0 Prediction

Table 1. The primers used for qRT-PCR

Gene name	Forward (5′ to 3′)	Reverse (5′ to 3′)	Product size (bp)
CYP6DA1	CGTAGAAGTCATCCATTCGCTG	GCTTTCCACTGATCTCCTCG	225
CYP6CY19	GATTCTTCTGTTTGTCGCCGGTG	GGGTAAGTCTTAGTTGCTTCTC	252
CYP6CY22	GATGTACTACTTATATCCGAACGCC	GTGTCCTTTAGCTTTCCAGATGTG	255
EF1-*a*	GAAGCCTGGTATGGTTGTCGT	GGGTGGGTTGTTCTTTGTG	187

Table 2. List of primers for promoter deletion plasmid construction

Gene		Sequence (5′ to 3′)	Lengths (bp)	Position
*CYP6DA1 Eco*R V/*Sac* I	E-1	GGATATCGTAAACGTAGCTCAGGAACAC		
	S-1	CGAGCTCGTTTCAGTTATGCCGGTG	243	+343/ + 586
	S-2	CGAGCTCGACAACTATTACTACTATAAAC	530	+56/ + 586
	S-3	CGAGCTCGTACTAACTCTCCTGTCC	779	−193/ + 586
	S-4	CGAGCTCCTTGTTTTGCTATCGCCC	1070	−484/ + 586
	S-5	CGAGCTCGTTGCAAATATTCTTCCTTAGG	1273	−687/ + 586
*CYP6CY19 Eco*R V/*Sac* I	E-1	GGATATCGTCACGGAAGTACTTATTAAGGTG		
	S-1	CGAGCTCGATATGGCGTCTTCGTG	297	−157/ + 140
	S-2	CGAGCTCGGTAAATCATTGTAC	568	−428/ + 140
	S-3	CGAGCTCCATTGTTGAGTCCTCCG	806	−666/ + 140
	S-4	CGAGCTCCCATTTTCCCACTATGATC	1046	−906/ + 140
*CYP6CY22 Hin*d III/ *Sac* I	H-1	CCCAAGCTTCCATTTATCGTACGTCG		
	S-1	CGAGCTCCTTGTATCTTTTCCACTCCATAAAC	326	−108/ + 218
	S-2	CGAGCTCGGTGTTTATTAAATGATGG	561	−343/ + 218
	S-3	CGAGCTCCAATAGTAACTACATGAACATGG	748	−530/ + 218
	S-4	CGAGCTCCTTATATTCCCAATCCTTTTACC	962	−744/ + 218

Server (http://www.cbs.dtu.dk/services/Promoter) and indicated by +1, with 5′ upstream sequences from it preceded by '−' and 3′ downstream sequences preceded by '+'. The name of each deletion construct includes a 'p' followed by initial letter of two restriction enzyme names and followed by the construct number (1–5) from the shortest to the longest fragment. The sizes between adjacent fragments were between 100 and 300 bp. The names and schematic diagrams of the corresponding pGL4-promoter deletion constructs are shown on Fig. 1.

General putative binding sites for transcription factors in the 5′-region of *CYP6DA1*, *CYP6CY19* and *CYP6CY22* genes were predicted using the New PLACE (https://sogo.dna.affrc.go.jp/cgi-bin/sogo.cgi?sid=&lang=en&pj=640&action=page&page=newplace) online software. The specific transcription factors binding sites were predicted by Alggen (http://alggen.lsi.upc.edu/), Jaspar (http://jaspar.genereg.net/) and Matinspector 8.0 Genomatix (https://www.genomatix.de/cgi-bin//eldorado/main.pl?s=5739a44394ae66817fe67993ac2cb359;SELECTION=reg).

2.6 Transfection and measurement of dual-luciferase activity

For transient transfection, Sf9 cells with 4×10^4 cells/mL density were added to each well of a 96-well microtiter plate. The plasmid pRL-TK (Promega) containing the Renilla firefly reporter gene was co-transfected into all cells to normalize the results. The pRL-TK vector (0.2 μg) and pGL4.11-reporter-promoter deletion construct (2.0 μg) were co-transfected in each well using 1 μL of Cellfectin II (Invitrogen). The molar ratio of pRL-TK vector to pGL4.11-reporter construct was 1:10. At 6 h post-transfection, the medium was replaced with the complete medium including 10% FBS. For plant

allelochemical induction experiments, plant allelochemical was added to each well (the final volume was still 100 μL) after 6 h of transfection, and incubated until the cells were harvest at 48 h. Acetone, the solvent for quercetin, gossypol, 2-tridecanone and sterile water, the solvent for tannic acid, were added to separate wells as negative controls. The activities of firefly and renilla luciferase in cells were measured 48 h later by a dual-luciferase reporter assay system (Promega) according to the manufacturer's instructions using a FLx800 System (BioTek, Winooski, VT, USA). Firefly luciferase activity was normalized to renilla luciferase activity. All experiments were performed in three independent biological replicates.

2.7 Transfection and measurement of dual-luciferase activity

All data were analyzed by *t*-test analysis. Significance was set at $P < 0.01$. All analysis was conducted with GraphPad Prism 5 software (San Diego, CA, USA). Data were expressed as mean ± standard error of the mean (SEM) from triplicate experiments.

3 RESULTS

3.1 Cloning and induction of three P450 genes

Three novel P450-like sequences from our previous transcriptome and expression profile analyses were predicted as unigenes.[12] Using the RACE technique, the full-length sequences of these genes were cloned in the current study. Comparing to the transcriptome unigene sequences, the 3′-UTR sequences of *CYP6DA1* and *CYP6CY19* were 14 and 98 bp longer, respectively.
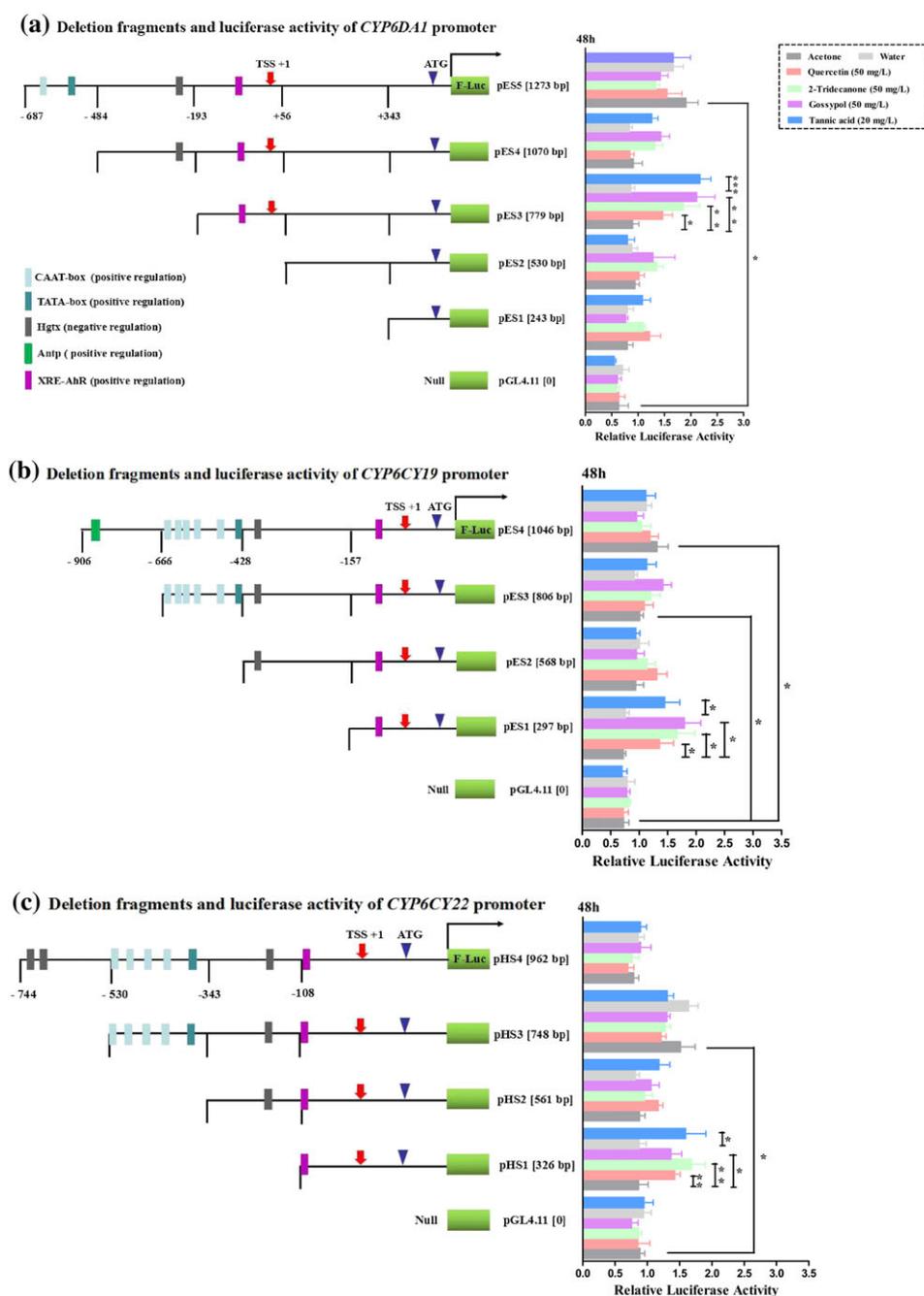
Figure 1. Diagrammatic drawing of the promoter deletion region and the luciferase activity driven by the promoter regions of different lengths. Left panel: the promoter fragments. Right panel: the luciferase activity. (a) The *CYP6DA1* deletion constructs; (b) the *CYP6CY19* deletion constructs; (c) the *CYP6CY22* deletion constructs. Green horizontal boxes represent the luciferase open reading frame (F-Luc). Different vertical boxes represent different *cis*-elements. The dark lines are the promoter region in the deletion constructs, the name and length of which are given in the right panel. The translation initiation site (ATG) and transcriptional start site (TSS) are indicated by red and blue arrows, respectively. The luciferase assays were performed in quadruplicate. The results are reported as mean ± standard error of the mean (SEM) of frefly/renilla ratios. For each treatment, the mean values for different cell lines were compared by two-tailed *t*-test to determine the *P*-values.

The full-length sequences of these three genes were named as *CYP6DA1* (2142 bp), *CYP6CY19* (1837 bp) and *CYP6CY22* (2171 bp) (personal communication with Professor Nelson, D.R., University of Tennessee, Memphis, TN, USA) and deposited in the GenBank database with the accession numbers KX945359 for *CYP6DA1*, KX945360 for *CYP6CY19*, and KX945361 for *CYP6CY22*.

As shown in Fig. 2, the expression of *CYP6CY19* and *CYP6CY22* were up-regulated by quercetin, 2-tridecanone, gossypol and

tannic acid. The treatment with quercetin did not significantly affect the expression of *CYP6DA1* ($P > 0.01$) while the treatments with the other three plant allelochemicals induced its expression by over 1.4-folds (Fig. 2(a)). Comparing to that of the untreated groups, the mRNA expression of *CYP6CY19* was increased by about 11.2-folds by the tannic acid treatment, 4.9-folds by the 2-tridecanone treatment, 4.6-folds by the quercetin treatment and 4.2-folds by the gossypol treatment (Fig. 2(b)). The inductions of
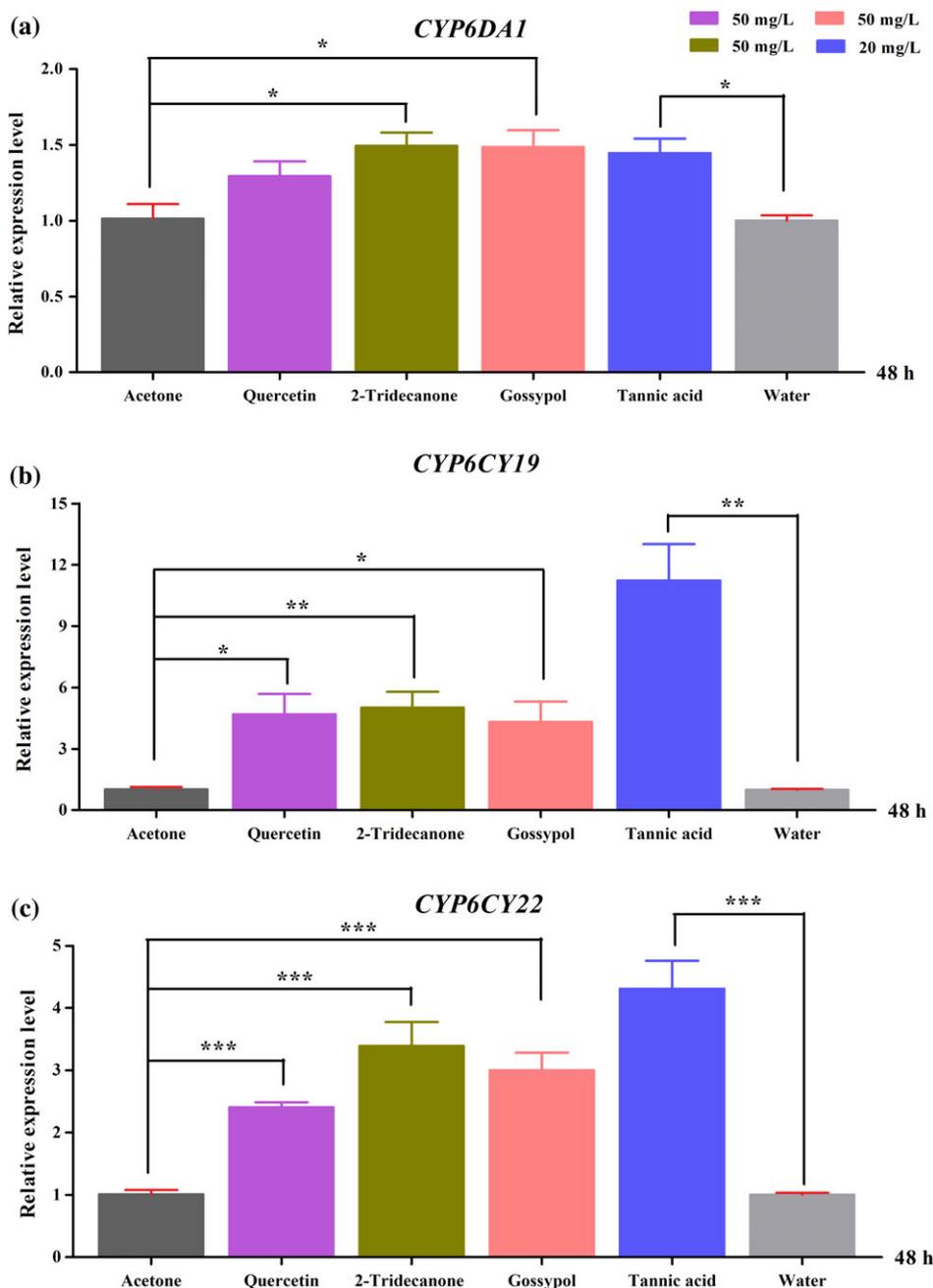
Figure 2. Transcriptional level of *CYP6* genes in the cotton aphids induced by plant allelochemicals. (a) The expression level of *CYP6DA1*; (b) the expression level of *CYP6CY19*; (c) the expression level of *CYP6CY22*. All error bars represent standard error of the mean, as determined from three independent experiments.

CYP6CY22 expression were also significant ($P < 0.0001$) with 4.3-, 3.4-, 3.0- and 2.4-folds increase by the tannic acid, 2-tridecanone, gossypol and quercetin treatments, respectively (Fig. 2(c)).

3.2 Identification of promoter regions and their background transcriptional activity

Using the genome walking approach, the promoter regions of *CYP6DA1* (762 bp), *CYP6CY19* (1100 bp) and *CYP6CY22* (958 bp) were obtained. The sequences were deposited in the GenBank with the accession numbers of KX950715 for *CYP6DA1*, KX950716 for *CYP6CY19* and KX950717 for *CYP6CY22*. To identify the critical region and the core elements required for gene transcription,

a series of 5′-deletions in the promoter regions of these three *CYP6* family genes were generated and ligated into pGL4.11 vector. The firefly luciferase activity of the pGL4.11 vector driven by the deletion fragments were measured and normalized to the renilla luciferase activity co-expressed by the pRL-TK vector. Unless the deletion fragment has a promoter activity, it cannot stimulate the transcription of firefly luciferase of the pGL4.11 vector. In other words, there would be a basal transcription activity if the deletion fragment has promoter activity. As shown in Fig. 1, the constructs of pES5 of *CYP6DA1* (Fig. 1(a)), pES3 and pES4 of *CYP6CY19* (Fig. 1(b)) and pHS3 of *CYP6CY22* (Fig. 1(c)) had a significantly higher luciferase activity comparing to the

pGL4.11 control construct ($P < 0.01$). In the promoter fragments of *CYP6CY19*, the basal transcriptional activity of pES3 was higher than pES2 (1.07 times), but there was no significant difference between them ($P > 0.01$). However, pES4 showed higher basal activity than pES3 (1.30 times, $P < 0.001$). These data suggest that the basal transcriptional activity of *CYP6DA1*, *CYP6CY19* and *CYP6CY22* promoter regions resides within 687 bp (-687 to $+1$), 666 bp (-666 to $+1$) and 530 bp (-530 to $+1$) from the TSS of each promoter, respectively. These results showed that, relative to the luciferase activity of the empty pGL4.1 vector, the critical responsive elements for the basal transcription could be between pES5 and pES4 of *CYP6DA1*, between pES3 and pES2 of *CYP6CY19*, between pHS3 and pHS2 of *CYP6CY22*, respectively. However, the construct pHS4 which containing the full-length promoter region of *CYP6CY22* did not show any luciferase activity, suggested the region between -744 and -530 bp can abolish the promoter activity.

3.3 Identification of the plant allelochemical responsive elements in the promoter region

To determine the induction effects of plant allelochemicals on the promoter activities, the constructs with the deletion fragments of different lengths from three P450 gene promoter regions were evaluated for the induction ability of the luciferase activity by quercetin, 2-tridecanone, gossypol and tannic acid. We first determined the working concentrations of these plant allelochemicals that would give significant induction of the *CYP6* promoter activity without causing any cell death. As shown in Fig. 1, the treatments at 50 mg/L of quercetin, 2-tridecanone, gossypol and at 20 mg/L of tannic acid significantly increased the luciferase activity of the *CYP6* promoter deletion constructs. The luciferase activities of the deletion constructs in the induced cells were higher than those of uninduced cells. The pES3 construct of *CYP6DA1* drove 2.5-, 2.4-, 2.1- and 1.6-fold higher luciferase activities compared to that of uninduced cells under the treatments of tannic acid, gossypol, 2-tridecanone and quercetin, respectively, suggesting the fragment between -193 and $+56$ bp from the TSS site in pES3 is critical for the plant allelochemical induction (Fig. 1(a)). The luciferase activity driven by the deletion construct pES1 of the *CYP6CY19* promoter region was 1.9-fold higher relative to the untreated controls under the inductions of quercetin and tannic acid ($P < 0.01$, two tailed *t*-test), and at least 2.5- and 2.3-fold higher under the inductions of gossypol and 2-tridecanone, respectively, suggesting the fragment between -157 and $+140$ bp from the TSS site in the deletion construct pES1 is critical for the plant allelochemical induction for *CYP6CY19* (Fig. 1(b)). For *CYP6CY22* promoter, with the deletion construct pHS1, the 2-tridecanone treatment induced the highest level of the luciferase activity (1.9-fold), the gossypol and quercetin treatments gave 1.6-fold higher induction and the tannic acid treatment caused 1.8-fold increase of the luciferase activity relative to the untreated samples, suggesting the fragment between -108 and $+218$ bp from the TSS site in the deletion construct pHS1 is critical for the plant allelochemical induction for *CYP6CY22* (Fig. 1(c)).

3.4 Prediction of universal and specific transcription *cis*-elements within the *CYP6DA1*, *CYP6CY19* and *CYP6CY22* promoters

To further confirm the core elements for the basal transcription in the promoter regions, the binding sites of regulatory factors within the full-length promoter region of each P450 gene was predicted by the plant *cis*-acting regulatory DNA elements website

NEW PLACE (https://sogo.dna.affrc.go.jp/cgi-bin/sogo.cgi?sid=&lang=en&pj=640&action=page&page=newplace). The sequences known to bind regulatory factors were framed with the factor's name as shown in Supporting Information File S1. The *cis*-acting regulatory DNA elements in the promoter region between pES5 and pES4 (-687 to -484 bp) of *CYP6DA1*, between pES3 and pES2 (-666 to -428 bp) of *CYP6CY19*, between pHS3 and pHS2 (-530 to -343 bp) of *CYP6CY22* were compared with each other as well as with the rest of their own promoter region, respectively. Firstly, two such binding sites were found in these critical promoter regions: CAAT-box1 (one in *CYP6DA1* promoter at -636 bp, five in *CYP6CY19* promoter at -631, -590, -574, -566 and -489 bp, four in *CYP6CY22* promoter at -530, -502, -469 and -431 bp) and TATA-box (one TATA-box2 in *CYP6DA1* promoter at -566 bp, one TATA-box5 in *CYP6CY19* promoter at -434 bp, one TATA-box2 in *CYP6CY22* promoter at -364 bp). Secondly, the *cis*-acting regulatory DNA elements in the region between pHS4 and pHS3 (-744 to -530) of *CYP6CY22* promoter were compared with others region of its own promoter. However, no special *cis*-elements were found in this region.

This analysis of the presence of regulatory factor binding sites further demonstrates that the promoter region in the deletion construct pES3 of *CYP6DA1*, pES1 of *CYP6CY19* and pHS1 of *CYP6CY22* could be induced by all four kinds of plant allelochemicals. In order to get an accurate site, we chose the repetitive prediction sites both from Alggen (http://alggen.lsi.upc.edu/) and Jaspar (http://jaspar.genereg.net/) at the same time to compare these elements to the regions in the deletion construct pES4 of *CYP6DA1*, pES2 of *CYP6CY19* and pHS2 of *CYP6CY22*, respectively. We found some sites such as Ftz (fushi tarazu), Eve (Even-skipped), En (engrailed), Dfd (deformed), Zen (zerknult) in the promoter regions. Using the Genomatix software, two elements Antp (antennapedia) and Hgtx (Hgtx homeobox transcription factor) were found in the promoter region between the deletion constructs pES4 and pES3 of *CYP6DA1*, between pES2 and pES1 of *CYP6CY19*, between pHS4 and pHS3, between pHS2 and pHS1 of *CYP6CY22*. Some *cis*-elements were also identified, such as core promoter initiator elements, DNA replication-related element factor, heat shock factors, Nrf2 and so on.

Furthermore, three potential AhR binding sites were found in the pES3 promoter region of *CYP6DA1*, in the pES1 promoter region of *CYP6CY19* and in the pHS1 promoter region of *CYP6CY22*, respectively. They were the 5'-TACCCAA-3' motif located in -61 to -54 bp of *CYP6DA1* promoter region, the 5'-CAAGCAA-3' motif located between -81 and -74 bp in *CYP6CY19* promoter region and the 5'-CACTCCA-3' motif located between -95 and -88 bp in the *CYP6CY22* promoter region, resulting in the consensus sequence of 5'-C/TAC/ANCA/CA-3'.

4 DISCUSSION

Plants produce a wide variety of secondary metabolites, or allelochemicals, that serve as defensive agents against herbivores and pathogens.[14] In our research, four different categories of plant allelochemicals were chosen to examine their effect on the P450 gene expression and regulation in the cotton aphids. Quercetin is a natural flavonoid found in some fruits and brassica vegetables, including apples, berries, grapes, onions, and tomatoes.[15,16] Tannins are the second most abundant group of natural polyphenolic compounds in the vascular plants.[17] 2-Tridecanone belongs to methyl ketones which is a group of volatile compounds. It has been detected in the leaves and stems of wild tomatoes and

has important natural and commercial roles including acting as pheromones in plants and natural insecticides.[18] Gossypol and derivatives play important roles in defense against the pests and pathogens due to their cytotoxicity.[19]

Studies in mammals, including humans, have shown that the members of the *CYP1*, *CYP2*, *CYP3* and *CYP4* subfamilies are involved in detoxification of xenobiotic compounds such as drugs, environmental pollutants, many natural products and ethanol.[2] This current research demonstrated that these four plant allelochemicals can regulate the expression of three newly identified members of the *CYP6* subfamily. However, the mRNA expression level of *CYP6DA1* showed only slight increase when it was induced by quercetin (Fig. 2(a)). But its promoter activity could be induced by quercetin with a much smaller level (Fig. 1(a)). It is possible that quercetin may not be a strong inducer or its induction concentration used in this experiment needs to be further optimized.

The promoter regions −687 to −484 bp between pES5 and pES4 of *CYP6DA1*, −666 to −428 bp between pES3 and pES2 of *CYP6CY19* and −530 to −343 bp between pHS3 and pHS2 of *CYP6CY22* may hold a decisive role for basal transcription of each promoter because their strong ability to drive the luciferase expression in both control treatments (acetone and water) and allelochemical treatment. Comparing these regions with each other, two regulatory elements, CAAT-box1 and TATA-box, were found. These elements are very important for basal transcription activity, and essential for the glutamine synthetase gene expression in pea (*Pisum sativum*).[20] When these elements were deleted, the basal activity decreased or disappeared. Furthermore, Hgtx and Antp transcription factor binding sites were found in the promoter regions of three P450 genes. Associated to Gene Ontology (GO) terms, Antp acts in the positive regulation of transcription from RNA polymerase II promoter (GO:0045944) while Hgtx is involved in negative regulations of gene expression (GO:0010629) and transcription from RNA polymerase II promoter (GO:0000122). In this study, an Antp transcription factors binding site was found in the *CYP6CY19* pES4 construct of which the promoter activity was higher than the pES3 without the Antp binding site. On the contrary, with the deletion constructs that only carried Hgtx transcription factor binding site, such as pES4 of *CYP6DA1*, pES2 of *CYP6CY19* and pHS2 of *CYP6CY22*, the promoter activity decreased or disappeared. Addition, there were two Hgtx transcription factor binding sites in the longest promoter of *CYP6CY22*, they may also have ability to abolish its promoter activity.

The inducible increase in the luciferase activity was obtained with the promoter region in the pES3 of *CYP6DA1*, in the pES1 of *CYP6CY19* and in the pHS1 of *CYP6CY22* (Fig. 1), suggesting that these promoter regions could be plant allelochemical responsive *cis*-elements. However, we could not exclude the possibility that these elements respond not only to plant allelochemicals but also to other chemicals. It is critical for future confirmation to extend the current study for the effects of more plant chemicals and, more importantly, non-plant chemicals as well as using plants containing these chemicals.

The aryl-hydrocarbon receptor is a major regulator of drug-metabolizing enzymes, its recognition site in the promoter region of the targeted genes is XRE-AhR (xenobiotic responsive element to the aryl hydrocarbon receptor), which also exists in insect genes.[21] According to previous studies, the promoter regions of both *CYP6B1* and *CYP6B4* contain XRE-AhR elements that are activated by binding to aryl hydrocarbon receptor (AhR)-ARNT complexes.[22] The consensus sequence 5′-CACGCNA-3′ very similar to the AhR binding site sequences was identified in the promoter fragment pES3 of *CYP6DA1*, pES1 of *CYP6CY19* and pHS1 of *CYP6CY22*. These AhR elements from each *CYP6* gene had some commonality. Firstly, they were all located inside the short promoter fragment of the CYP gene. Secondly, they were very near the TSS site, the position was −61 to −55 bp of *CYP6DA1*, −81 to −75 bp of *CYP6CY19* and −95 to −89 bp of *CYP6CY22*. Thirdly, they had a very similar sequence as 5′-TACCCAA-3′ of *CYP6DA1*, 5′-CAAGCAA-3′ of *CYP6CY19* and 5′-CACTCCA-3′ of *CYP6CY22*, the consensus sequence motif was 5′-C/TAC/ANCA/CA-3′. These results suggested that the consensus sequence 5′-C/TAC/ANCA/CA-3′ of these three *CYP6* genes may be the potential plant allelochemical responsive *cis*-elements. However, whether or not these *cis*-elements were specificity or communal binding to the xenobiotic needs further studies in the future.

Other transcription factors Ftz, Eve, En, Dfd, Zen and Nrf2 were also identified in the current study and all belong to homeodomain transcription factor, which coregulate the expression of hundreds of downstream target genes involved in the development of insects.[23] Nrf2 is the primary transcriptional factor involved in the regulation of the expression of antioxidant and metabolizing enzymes important for the protection of cells against oxidative damage. Moreover, Nrf2 promotes the expression of various cytoprotective genes such as genes for superoxide dismutase, catalase and glutathione-*S*-transferase by binding to antioxidant responsive elements (ARE) in the 5′-flanking regions of many detoxifying genes in response to xenobiotic and oxidative stress.[5] This mechanism may be related to the plant–insect interactions and requires further research. Additional experiments are required to determine how transcription factors are involved in the interaction in order to understand the resistance mechanisms of the *A. gossypii* in the field. Further biochemical and molecular characterization using electrophoretic mobility shift assays and DNAse I protection assays could elucidate this key element and other transcription factor binding sites of these three newly identified P450 genes.

In conclusion, the core promoter regions of three P450 genes *CYP6DA1*, *CYP6CY19* and *CYP6CY22* in *A. gossypii* were reported for the first time in this paper. Further confirmation of their roles in the regulation of *CYP6* genes may be essential for elucidating the contribution of P450 proteins in the resistance mechanisms of insect pests to plant allelochemicals and other xenobiotics.

4.1 CONTRIBUTIONS OF AUTHORS

FL, XWG and JJZ designed the research and wrote the paper. FL and KSM performed the experiments. XW Chen participated in the data analysis. All authors read and approved the final manuscript. The authors declare that there are no conflicts of interest.

ACKNOWLEDGEMENTS

SUPPORTING INFORMATION

Supporting information may be found in the online version of this article.

REFERENCES

1 Carletto J, Martin T, Vanlerberghe-Masutti F and Brevaultc T, Insecticide resistance traits differ among and within host races in *Aphis gossypii*. *Pest Manag Sci* **66**:301–307 (2010).

2 Morra R, Kuruganti S, Lam V, Lucchesi JC and Ganguly R, Functional analysis of the cis-acting elements responsible for the induction of the *Cyp6a8* and *Cyp6g1* genes of *Drosophila melanogaster* by DDT, phenobarbital and caffeine. *Insect Mol Biol* **19**:121–130 (2010).

3 Zhao SS, Zhao GD, Di TY, Ding H, Wan XL, Li B *et al.*, Analyzing the promoters of two *CYP9A* genes in the silkworm *Bombyx mori* by dual-luciferase reporter assay. *Mol Biol Rep* **40**:1701–1710 (2013).

4 Wu K, Zhang J, Zhang QL, Zhu SL, Shao QM, Clark KD *et al.*, Plant phenolics are detoxifed by prophenoloxidase in the insect gut. *Sci Rep* **5**:16823 (2015).

5 Mohammed BR, Wilding CS, Collier PJ and Deeni YY, Cloning of *Anopheles Gambiae CYP6M2* gene promoter and construction of its luciferase reporter system. *Int J Sci Technol Res* **3**:259–264 (2014).

6 Liu NN, Li M, Gong YH, Liu F and Li T, Cytochrome P450s – Their expression, regulation, and role in insecticide resistance. *Pestic Biochem Physiol* **120**:77–81 (2015).

7 Mao YB, Cai WJ, Wang JW, Hong GJ, Tao XY, Wang LJ *et al.*, Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nat Biotechnol* **25**:1307–1313 (2007).

8 Li F, Liu XN, Zhu Y, Ma J, Liu N and Yang JH, Identification of the 2-tridecanone responsive region in the promoter of cytochrome P450 *CYP6B6* of the cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae). *Bull Entomol Res* **104**:801–808 (2014).

9 Zhao GD, Zhang YL, Liu YL, Li B, Chen YH, Xu YX *et al.*, Promoter analysis and RNA interference of *CYP6ab4* in the silkworm *Bombyx mori*. *Mol Genet Genomics* **290**:1943–1953 (2015).

10 Wan H, Liu Y, Li M, Zhu SY, Li XC, Pittendrigh BR *et al.*, Nrf2/Maf-binding-site-containing functional *Cyp6a2* allele is associated with DDT resistance in *Drosophila melanogaster*. *Pest Manag Sci* **70**:1048–1058 (2014).

11 Ma KS, Li F, Liang PZ, Chen XW, Liu Y and Gao XW, Identification and validation of reference genes for the normalization of gene expression data in qRT-PCR analysis in *Aphis gossypii* (Hemiptera: Aphididae). *J Insect Sci* **16**:1–9 (2016).

12 Li F, Ma KS, Liang PZ, Chen XW, Liu Y and Gao XW, Transcriptional responses of detoxification genes to four plant allelochemicals in *Aphis gossypii*. *J Econ Entomol* **110**:624–631 (2017).

13 Livak KJ and Schmittgen TD, Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* **25**:402–408 (2001).

14 Li XC, Berenbaum MR and Schuler MA, Plant allelochemicals differentially regulate *Helicoverpa zea* cytochrome P450 genes. *Insect Mol Biol* **11**:343–351 (2002).

15 Rojas Á, Del Campo JA, Clement S, Lemasson M, García-Valdecasas M, Gil-Gómez A *et al.*, Effect of quercetin on Hepatitis C Virus life cycle: from viral to host targets. *Sci Rep* **6**:31777 (2016).

16 Abdelkawy KS, Balyshev ME and Elbarbry F, A new validated HPLC method for the determination of Quercetin: application to study pharmacokinetics in rats. *Biomed Chromatogr* **31**:e3819 (2017).

17 Muccilli V, Cardullo N, Spatafora C, Cunsolo V and Tringali C, α-Glucosidase inhibition and antioxidant activity of an oenological commercial tannin. Extraction, fractionation and analysis by HPLC/ESI-MS/MS and ^{1}H NMR. *Food Chem* **215**:50–60 (2016).

18 Antonious GF, 2-Undecanone and 2-tridecanone in field-grown onion. *J Environ Sci Health B* **48**:302–307 (2013).

19 Tian X, Ruan JX, Huang JQ, Fang X, Mao YB, Wang LJ *et al.*, Gossypol: phytoalexin of cotton. *Sci China Life Sci* **59**:122–129 (2016).

20 Tjaden G, Edwards JW and Coruni GM, *cis* elements and frans-acting factors affecting regulation of a nonphotosynthetic light-regulated gene for chloroplast glutamine synthetase. *Plant Physiol* **108**:1109–1117 (1995).

21 Giraudo M, Hilliou F, Fricaux T, Audant P, Feyereisen R and Le Goff G, Cytochrome P450s from the fall armyworm (*Spodoptera frugiperda*): responses to plant allelochemicals and pesticides. *Insect Mol Biol* **24**:115–128 (2015).

22 McDonnell CM, Brown RP, Berenbaum MR and Schuler MA, Conserved regulatory elements in the promoters of two allelochemical-inducible cytochrome P450 genes differentially regulate transcription. *Insect Biochem Mol Biol* **34**:1129–1139 (2004).

23 Kim DH, Lee BY, Kim HS, Jeong CB, Hwang DS, Kim IC *et al.*, Identification and characterization of homeobox (*Hox*) genes and conservation of the single *Hox* cluster (324.6 kb) in the water flea *Daphnia magna*. *J Exp Zool (Mol Dev Evol)* **330**:76–82 (2018).