Identifying a *cis-element* in PtoCP1 promoter for efficiently controlling constitutive gene expression in *Populus tomentosa*

Yu Peng*, Xueqin Guo*, Yawei Fan, Han Liu, Leiqian Sun, Di Liu, Hui Li, Xin Wang, Hongli Guo and Hai Lu

State Key Laboratory of Tree Genetics and Breeding, National Engineering Research Center of Tree Breeding and Ecological Restoration, The Tree and Ornamental Plant Breeding and Biotechnology Laboratory of National Forestry and Grassland Administration, Co, Beijing Forestry University, Beijing, China

* These authors contributed equally to this work.

ABSTRACT

Gene expression is regulated by transcription factors binding to *cis-elements* in promoters. However, efficient *cis-elements* for genetic engineering are rarely reported. In this study, we identified an 11 bp *cis-element* in the PtoCP1 promoter that drives strong constitutive gene expression in *Populus tomentosa*. A 2,270 bp promoter region upstream of the *PtoCP1* gene's translation start site was cloned and named ProPtoCP1. This promoter controls GUS reporter gene expression in the roots, leaves, and stems of *Arabidopsis* seedlings. Based on the location and density of *cis-elements*, the PtoCP1 promoter was divided into four fragments by 5′-end deletions. GUS staining and RT-qPCR revealed a key *cis-element* at −466 to −441 bp essential for gene expression. Further analysis showed that the MYB-TGACG *cis-element* is a positive regulator, whereas neither MYB nor TGACG alone drove gene expression. This study enhances our understanding of gene expression regulation by *cis-elements* and provides a valuable tool for genetic engineering.

INTRODUCTION

The Chinese white poplar (*P. tomentosa Carr.*), a commercially valuable tree species used for timber production, belongs to the section *Populus* (Leuce). It is a diploid (2n = 2× = 38) and outcrossed perennial species (*Tian, Xie & Qu, 2013*; *Xun et al., 2021*). Proteases play a crucial role in the growth and development of Chinese white poplar. The main cysteine proteases in plants are categorized into the Papain and Legumain types, with other members of the family falling under the Caspase and Calpain types (*Yan, Yang & Han, 2005*). Among these, the Papain cysteine proteases (PLCPs) are the most prevalent and extensively studied (*Liu et al., 2018*). Cysteine proteases are expressed in various tissues and are involved in the degradation of related proteins during different biological processes in plants.

The cysteine protease CEP1 in *Arabidopsis* is responsible for mediating tapetal programmed cell death (PCD) and secondary wall thickening. The expression level of CEP1 is closely associated with the level of PCD in the tapetum layer and pollen fertility (*Zhang et al., 2014*). The *Arabidopsis* cysteine protease γVPE indirectly influences PCD in xylem fiber cells by activating the maturation of the CEP1 protease (*Cheng et al., 2019*). *HvPAP14*, identified in barley, is involved in the normal turnover of chloroplast proteins and potentially degrades macromolecular proteins during leaf senescence (*Frank et al., 2019*). The cysteine protease gene *PtCP5* impacts seed germination by mobilizing *Populus* storage proteins during this process (*Liu et al., 2021*).

Furthermore, the expression of cysteine proteases in many plant species is influenced by biotic and abiotic stressors. For instance, the transcription levels of four C1A cysteine protease genes increase in response to drought stress in barley plants. Knockout lines of proteases encoded by the genes *HvPap-1* and *HvPap-19* in these plants result in changes in cuticle thickness and stomatal area in the leaves, which aids in preventing fungal infections and mite predation (*Gomez-Sanchez et al., 2019*). In tomato plants, SA-induced PCD is associated with rapid upregulation of specific cysteine proteases in root tissues and increased proteolytic activity. This is attributed to the downregulation of protease inhibitors (*Kovács et al., 2016*). Ultimately, the tissue-specific expression of cysteine proteases and their involvement in the stress response are vital for plants to adapt better to their environment.

The function of cysteine proteases is directly influenced by their expression. In sweet potatoes, the *SPCP2* gene expression significantly increases in naturally aged leaves, but is nearly absent in mature green leaves, stems, and roots (*Chen et al., 2010*). Similarly, the expression level of the CP gene increases during leaf senescence in *Picrorhiza* kurrooa Royle ex Benth (*Jai, Sanjeeta & Shruti, 2015*). The *SAG12* protein, known as a reference gene for leaf senescence, also shows significant increase in expression during this process (*James et al., 2018*). However, there is limited research on the *cis-elements* and transcription factors responsible for regulating cysteine protease expression. Some studies suggest that WRKY75 acts as a positive regulator of leaf senescence by directly binding to the *SAG12* and *SAG29* promoters, thus activating their expression. Expression of age-related genes (SAG) is inhibited in wrky75 mutants and increased in WRKY75 overexpressing plants (*Du et al., 2012*).

To investigate the specificity of *PtoCP1* gene expression, we obtained the upstream promoter sequence of the gene and performed GUS staining in the leaves, roots, and stems of *Arabidopsis thaliana*. Our findings indicate that a key MYB-TGACG *cis-element* plays a regulatory role in gene expression, as demonstrated through 5′end deletion and site-specific mutation experiments. This study identifies important elements involved in cysteine protease gene expression and provides valuable information for genetic engineering.

MATERIALS AND METHODS

Growth of plant materials

Wild-type tobacco (*Nicotiana benthamiana*) seeds were vernalized in the dark at 4 °C for 3 days and then planted in nutrient soil (substrate: vermiculite = 3:1) using toothpicks. The

pots were covered with plastic wrap and placed in a laboratory light incubator set at 25 °C, 60% humidity, and a 16-h light/8-h darkness cycle. The plastic wrap was removed when the plants reached two weeks of age. The plants were cultivated until they were 5 weeks old for transient transformation and GUS staining of tobacco leaves.

Col-0 wild-type *Arabidopsis* seeds were also vernalized in the dark at 4 °C for 3 days and then planted in nutrient soil (substrate: vermiculite = 3:1) using toothpicks. The pots were covered with plastic wrap and placed in a laboratory light incubator set at 25 °C, 60% humidity, with a consecutive alternating culture of 16 h light at 23 °C, 60% humidity, and 8 h dark. The plastic wrap was removed when the plants reached the stage of having six true leaves and were cultivated until they reached the flowering stage.

Wild-type *Populus alba* × *Populus glandulosa* were grown on ½ MS solid medium in a greenhouse at a constant temperature of 25 °C, with a 14 h light/10 h dark photoperiod. ½ MS solid medium included 1.5% (w/v) sucrose, 2.18 g l^{-1} ½ Murashige and Skoog salts, 100 mg l^{-1} inositol, 0.1 mg l^{-1} 1−naphthaleneacetic acid (NAA) and 6 g l^{-1} agar, pH 6.15.

Transformation of plant materials

Transient transformation of tobacco leaves: Prior to the immediate infection of tobacco leaves, it is necessary to ensure that they have absorbed an ample amount of water. During the infection process, remove the needle from a 1 mL syringe and carefully inject an appropriate quantity of the corresponding bacterial suspension into the lower epidermis of tobacco leaves that are 4–5 weeks old. Gently and evenly push the syringe until the desired infection area is reached, and immediately cease injection. Once the injection process is complete, use filter paper to wipe away any remaining bacterial liquid on the leaves. Proceed by marking the leaves and covering the plants with plastic film to prevent evaporation of the bacterial liquid and maintain moisture. Incubate the plants in darkness at room temperature for 36 h, allowing the bacterial solution to fully penetrate the leaves. Finally, transfer the plants to normal cultivation under 12 h of light for GUS histochemical staining.

Stable transformation of *Arabidopsis*: The *Agrobacterium*-mediated transformation method using the flower-dip technique was employed. To promote the growth of lateral branches, carefully cut the main stem of *Arabidopsis thaliana* with scissors, and then infect the inflorescence of wild *Arabidopsis thaliana* that has been cultivated for a specific duration. Before the initial infection, remove any mature fruit pods and immerse the entirety of the *Arabidopsis* inflorescence in water for approximately 30 s. Once infected, cover the entire plant with plastic wrap to maintain moisture. Allow the plant to undergo dark cultivation for 24 h, and then transfer it to a well-lit environment for growth. Sterilize and select transformant seeds on half-strength MS agar medium supplemented with 1.5% (w/v) sucrose and 50 mg/L kanamycin sulfate. To further verify positive seedlings, perform PCR identification using a rapid DNA extraction kit on plants that have successfully survived on the antibiotic-containing culture medium. In the PCR process, upstream primers should target sequences on the promoter, whereas downstream primers should target sequences on GUS (F: 5′-TAAAAGCCTCCACCTCTCGC-3′, R: 5′-AACTGTTCGCCCTTCACTGC-3′). Ultimately, utilize T3 generation homozygous positive plants for subsequent experiments.

Stable transformation of *Populus alba* × *Populus glandulosa*: Take small sections of *Populus alba* × *Populus glandulosa* leaves and place them on solid medium WPM supplemented with 0.1 mg/l NAA. Pre-culture the leaves in darkness at a temperature of 25 °C for 2–3 days. Subsequently, infect the leaves with *A. tumefaciens* GV3101 at an optical density of 0.6 for a duration of 10–15 min at a wavelength of 600 nm. Use carbenicillin sodium (250 mg/l) and kanamycin (20 mg/l) for selection purposes. Replace the selective medium every 2 weeks until adventitious buds begin to develop.

Cloning of the *PtoCP1* promoter

The genomic database for JGI *P. trichocarpa* (https://phytozome.jgi.doe.gov/pz/portal.html) was utilized to locate the corresponding *PtCP1* genomic sequence. Our method involved extracting the *PtoCP1* genome DNA from a 100 mg portion of fresh *P. tomentosa* leaves, following the manufacturer's instructions and using the extracted genomic DNA as the template. Specific primers (F: 5′-CTAAATGTGCAATTTGTAACCATCA-3′, R: 5′-CGAGAATATTGCAAACCTTAACTTC-3′) were designed with reference to the sequence of the *PtCP1* genome to clone the 2,270 bp *PtoCP1* gene promoter. The cloned promoter was then connected to the pMD18-T vector and sent to a biological company for sequencing. For promoter expression analysis, promoter fragments with correct sequencing were constructed into the pBI121 expression vector. We would like to highlight that the synthesis of primers and the sequencing services were provided by Beijing Tsingke Biotech Co., Ltd.

Bioinformatics analysis of *PtoCP1* promoter sequence cis-element

The online sequence analysis of the *PtoCP1* promoter of the *P. tomentosa* cysteine protease gene was performed using the PlantCARE (http://bioinformatics.psb.Ugent.be/webtools/plantcare/html/) and PLACE (https://www.dna.affrc.go.jp/PLACE/?action=newplace) databases to predict the *cis-element* on the promoter.

Vector construction

In the vector construction experiment for the promoter fragments, the following upstream primers were employed to amplify the full-length promoter and 5′ deletion: P0 (F: 5′-CTATGACCATGATTACGCCAAGCTTTGTGCAATTTGTAACCATCA-3′), P1 (F: 5′-GACCATGATTACGCCAAGCTTCCTGGCTTGAAAGGTTGGACT-3′), P2 (F: 5′-GACCATGATTACGCCAAGCTTAACCGTTAGGCGTCACCAT-3′), P3 (F: 5′-GACCATGATTACGCCAAGCTTCACAGTAAGCGATAACGGAACTT-3′), and P4 (F: 5′-GACCATGATTACGCCAAGCTTCATGTACTTTGCTCCTTGACCAG-3′). The downstream universal primer used was PR (R: 5′-GACTGACCACCCGGGGATCCGATTTACTTGTTTTTGTGGGTGGTC-3′). The upstream primers contained a segment of the homologous arm sequence and a *Hin*d III restriction site upstream of the vector, whereas the downstream primer contained a segment of the homologous arm sequence and a *Bam*H I restriction site downstream of the vector. The PCR reaction involved a 50 μL mixture comprising 25 μL of 2× Phanta Max Buffer, 1 μL of the dNTP Mix, 2 μL of the 10 μM upstream primer, 2 μL of the 10 μM downstream primer, 1 μL of template, and

ddH$_2$O to complete the volume. The PCR reaction procedure included incubation at 95 °C for 3 min, followed by cycles at 95 °C for 15 s, 60 °C for 15 s, 72 °C for 15 s, and a final extension step at 72 °C for 5 min, with a subsequent holding step at 4 °C. We employed 35 cycles for steps 2 to 4. The amplified target fragments were homologously recombined into the pBI121 expression vectors, with the 35S promoters removed using *Hin*d III and *Bam*H I, following the instructions provided in the seamless cloning kit. The resulting conjugate product was then transformed into competent *E. coli* DH5α, and the positive single colony was screened in LB medium containing *Kan*. The correctness of the recombinant product was verified using PCR, double enzyme digestion, and sequencing. To construct the vector for *cis-elements* connecting 35Smini composite fragments, the upstream primers for the target fragment were as follows: ABC (F: 5′-GACCATGATTACGCCAAGCTTCCGTTA CGTCACATGTGCCTTCGCAAGACCCTTC-3′), AAA (F: 5′-GACCATGATTA CGCCAAGCTTCCGTTACCGTTACCGTTACCTTCGCAAGACCCTTC-3′), BBB (F: 5′-GACCATGATTACGCCAAGCTTCGTCACGTCACGTCACCTTCGCAAGA CCCTTC-3′), and CCC (F: 5′-GACCATGATTACGCCAAGCTTCATGTGCATGTGCA TGTGCCTTCGCAAGACCCTTC-3′). The downstream universal primer used was ABC (R: 5′-GACTGACCACCCGGGGATCCCGTGTTCTCTCCAAATGAAC-3′).

For the vector construction of site-mutated *cis-elements* connecting 35Smini composite fragments, the upstream primers for the target fragment were (A)BC (F: 5′-GACCATGA TTACGCCAAGCTTCTAGGACGTCACATGTGCCTTCGCAAGACCCTTC-′), A(B)C (F: 5′-GACCATGATTACGCCAAGCTTCCGTTACAGCACATGTGCCTTCGCAAGA CCCTTC-3′), and AB(C) (F: 5′-GACCATGATTACGCCAAGCTTCCGTTACGTCA CCGAGGCCTTCGCAAGACCCTTC-3′). The downstream universal primer used was ABC (R: 5′-GACTGACCACCCGGGGATCCCGTGTTCTCTCCAAATGAAC-3′).

GUS histochemical staining

Five 7-day-old stable transgenic *Arabidopsis* seedlings were randomly selected and placed in 2 mL centrifuge tubes. The leaves from transient tobacco infections were placed in a suitably sized petri dish. Pre-cooled 90% acetone (−20 °C) was added to the tubes and petri dish to completely cover the material. The tissue was treated at room temperature for 20–30 min to fix it and partially remove the chlorophyll. The plant material was then rinsed with distilled water, and a GUS staining buffer was added to cover it completely. The buffer contained 100 mM sodium phosphate (pH 7.2), 10 mM EDTA, 1% (v/v) Triton X-100, and 0.5 mg/L X-Gluc, as modified by *Berger et al. (2011)*. The tubes and petri dish were attached to a vacuum pump for 40 min, kept away from light at 37 °C, and incubated overnight. The material was then eluted through an ethanol gradient (50%, 70%, and 95%), gently shaking for 40 min each time. The 95% ethanol was used until the green color of the leaves was completely removed. GUS staining was observed by the naked eye or with an anatomic microscope.

Expression analysis by real-time qPCR

To analyze differences in GUS expression levels driven by different promoter fragments, RNA was extracted from 7-day-old transgenic *Arabidopsis* leaves. RT-qPCR was

Figure 1 The *ProPtoCP1 cis*-acting element analysis.

Full-size DOI: 10.7717/peerj.18292/fig-1

performed using SYBR Green qPCR Mix (TIANGEN, Beijing, China) on an iQ5 Multicolor Real-Time PCR detection system (Bio-Rad Laboratories, Hercules, CA, USA). The primer sequence for detecting the reporter gene GUS was (F: 5′-TAACCACAAA CCGTTCTACTT-3′, R: 5′-ACTGATACTCTTCACTCCACA-3′), and the primer sequence for the internal reference gene Actin was (F: 5′-CGTATGAGCAAGGAGATCA C-3′, R: 5′-CACATCTGTTGGAAGGTGCT-3′). The PCR conditions were 95 °C for 15 min, 95 °C for 10 s, 55 °C for 20 s, 72 °C for 30 s, 39 cycles were employed from steps 2 to 4. Data were analyzed using iQ5 (Bio-Rad) software, and differences in gene expression were calculated using the $2^{-\Delta\Delta Ct}$ method. β-actin was used as an internal control to quantify the relative expression levels of genes in the samples.

Table 1 *Cis*-element information contained in *ProPtoCP1*.

Cis-element	Sequence (5′–3′)	Start position	Function
CARE	CAACTC	−1,799, −1,652, −1,405	GA response element
ATCT-motif	AATATAATCC	−1,142	Light response element
T-box	ACTTTG	(−1,794), (−1,740), −107	Light response element
BOX 4	ATTAAT	−2,079, −1,549, −1,351, −728, −652, −615	Light response element
GT1-motif	GRWAAW GGTTAA	−1,828, −1,748, (−1,661), (−1,636), (−1,458), (−1,435), (−1,166), (−1,016), (−1,008), (−969), −958, (−920), (−876), (−794), (−736), (−330), (−307), (−248), (−239), (−45),	Light response element
TGACG-motif	TGACG	(−454), (−444)	TGA binding site, MeJA, SA response element
W BOX	TTGACY	−987, −899, −97	WRKY binding site, SA, pathogen, damage response element
MYB	CNGTTR YAACKG WAACCA	−2,257, (−2,053), (−1,907), −1,744, (−464), −429	MYB binding site, ABA, drought *etc.*, response element
MYC	CANNTG	−1,972, −1,838, −1,754, −1,459, −1,471, −1,361, −450, −134	MYC binding site, ABA, drought *etc.*, response element

Note:
() indicates promoter sequence antisense chain; R: A/G; W: A/T; Y: C/T; N: A/T/C/G; K:G/T.

RESULTS

ProPtoCP1 cloning and bioinformatics analysis of the *cis*-elements

To determine the expression of the *PtoCP1* gene, we designed primers (Forward: 5′-CTAAATGTGCAATTTGTAACCATCA-3′, Reverse: 5′-CGAGAATATTGCAAA CCTTAACTTC-3′) based on the *PtCP1* sequence (Genbank Accession Number: XM_006381559.1) available at the JGI (https://phytozome-next.jgi.doe.gov/blast-search). We successfully cloned a 2,270 bp *PtoCP1* promoter sequence from *P. tomentosa Carr.*, which we named *ProPtoCP1*. The fragment size was measured from the translation initiation site (ATG). The CDS sequence similarity between *PtoCP1* (*Populus tomentosa*) and *PtCP1* (*Populus trichocarpa*) was found to be 98.50%. We analyzed the *PtoCP1* promoter sequence using the Plant CARE (http://bioinformatics.psb.Ugent.be/webtools/ plantcare/html/) and PLACE (https://www.dna.affrc.go.jp/PLACE/?action=newplace) promoter analysis software to predict the presence of *cis*-elements (Fig. 1 and Table 1). The *PtoCP1* promoter region was found to contain numerous photoresponsive elements (T-box, BOX4, and GT1-motif), as well as plant hormone and abiotic stress response elements such as methyl jasmonic acid (MeJA) response element (TGACG-motif), SA response element (W BOX), gibberellic acid (GA) response element (CARE), abscisic acid (ABA), and drought response elements (CNGTTR, YAACKG, WAACCA, and CA NNTG). These findings suggest that the expression of the *PtoCP1* gene may be regulated by multiple factors, including abiotic stress and hormones.
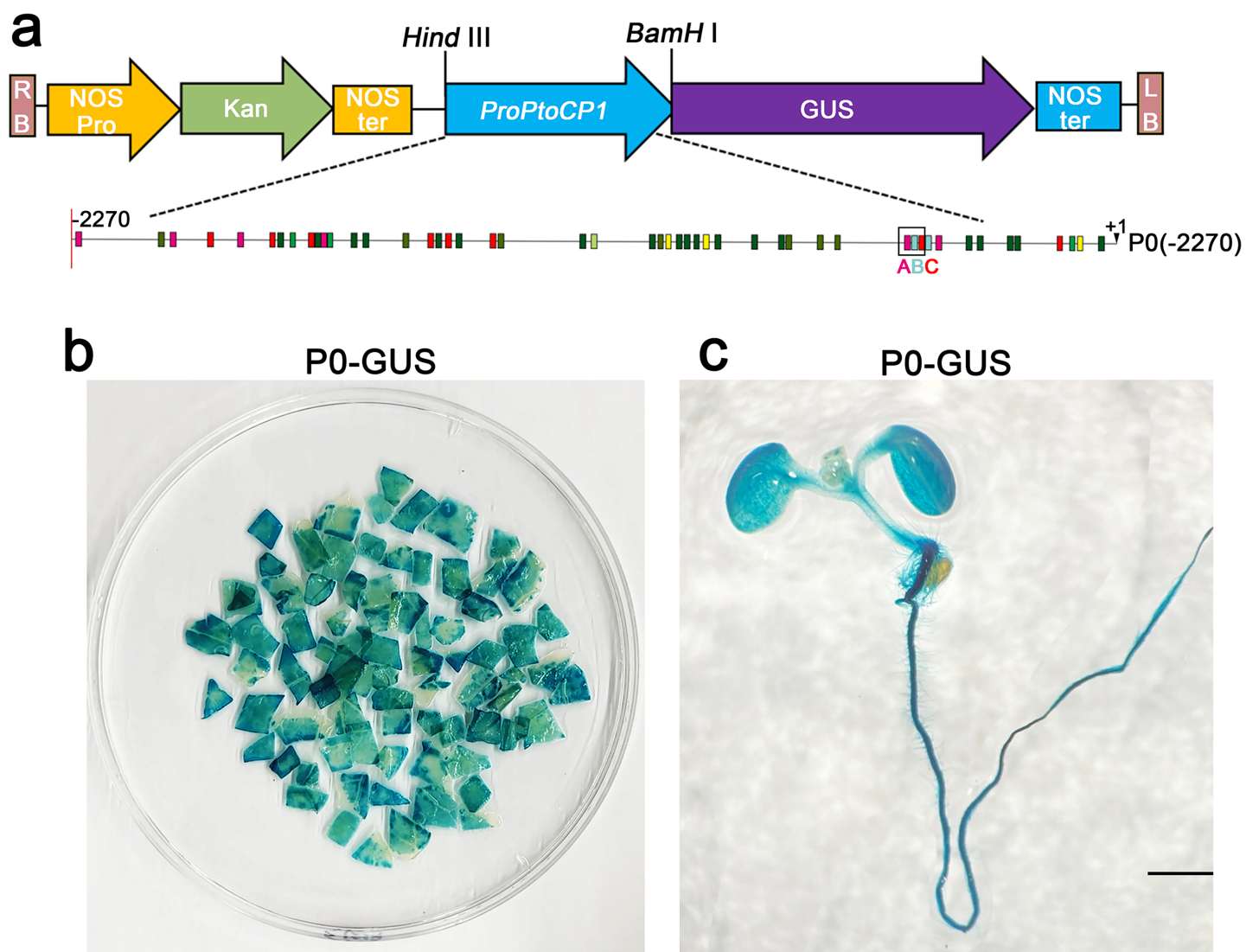
Figure 2 GUS staining of *PtoCP1* full-length promoters in tobacco leaf and 7-day-old transgenic *Arabidopsis thaliana* seedlings. (A) Schematic representation of the *PtoCP1* full-length promoter fragments fusion with GUS; (B) GUS staining of different tobacco leaves during the same period; (C) GUS staining of the 7-day-old transgenic *Arabidopsis* seedlings. Bar = 1 mm.   Full-size ⬛ DOI: 10.7717/peerj.18292/fig-2

Analysis of key fragments of *PtoCP1* promoter expressed by driver gene

To assess the activity of the PtoCP1 promoter fragment, we fused the full-length PtoCP1 promoter fragment P0 (−2,270 bp) with the GUS reporter gene (Fig. 2A). This recombinant expression vector was then transiently transformed into tobacco leaves using the *Agrobacterium*-mediated method. Additionally, we transformed it into *A. thaliana* to generate stably transformed plants. The GUS staining results revealed that the PtoCP1 promoter fragment drives GUS expression in tobacco leaves (Fig. 2B). GUS staining was also detected in the leaves, roots, and stems of 7-day-old transgenic *Arabidopsis* seedlings (Fig. 2C).
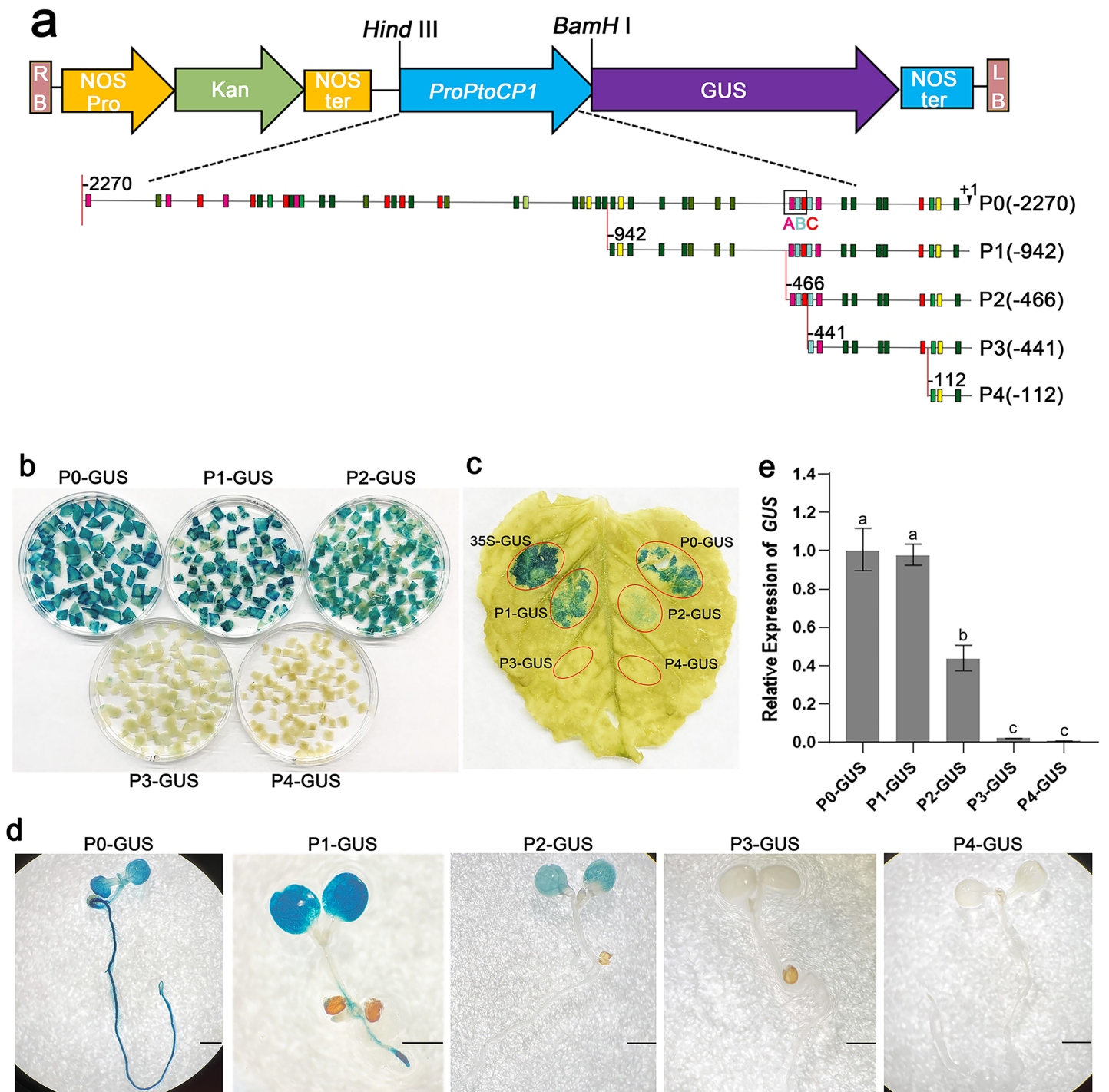
Figure 3 GUS activities analysis of *PtoCP1* promoters full-length and deletion series in transgenic *Arabidopsis* thaliana seedlings and tobacco leaf. (A) Schematic representation of diferent promoter fragments fusion with GUS; (B) GUS staining in leaves of tobacco; (C) GUS staining; (D) GUS staining of the transgenic *Arabidopsis* seedlings, Bar = 1 mm; (E) Expression levels of GUS under the control of a *PtoCP1* promoter full-length and deletion fragments in transgenic *Arabidopsis* thaliana seedlings. The relative difference of GUS gene expression was detected at the *P* value 0.05 level, and different letters represented significant difference.

Full-size 🖼 DOI: 10.7717/peerj.18292/fig-3

To determine the key fragments of the PtoCP1 promoter that regulate protease expression *via* MYB elements, we designed primers based on the MYB element locations in the promoter. We then truncated the full-length promoter by deleting the 5′-end, resulting in four promoter deletion fragments named P1 (−942 bp), P2 (−466 bp), P3 (−441 bp), and P4 (−112 bp). Each of these promoter deletion fragments was fused with GUS reporter genes (Fig. 3A). Subsequently, the recombinant vectors containing the full-length promoter and the deletion promoter fragments were introduced into tobacco leaves. The GUS staining results showed that compared to P0, the GUS staining of P1 exhibited no significant change, whereas the GUS staining intensity of P2 decreased. Furthermore, the GUS staining in P3 (−441 bp) and P4 (−112 bp) almost completely disappeared (Figs. 3B and 3C). To validate these findings, we stably transformed recombinant vectors containing the PtoCP1 promoter and the 5′-end deletion series into *Arabidopsis*. The GUS staining and RT-qPCR results indicated (Figs. 3D and 3E) that there was no significant difference in GUS expression in transgenic *Arabidopsis* leaves driven by P1 and P0. However, GUS expression driven by P2 decreased by approximately one fold, whereas P3 and P4 showed negligible GUS gene transcription. These results suggest that an essential gene expression element is located in the region between P2 and P3 (−466 to −441 bp).

## An important MYB-TGACG *cis-element* of *PtoCP1* driven gene expression

Upon comparing the fragments P2 (−466 bp) and P3 (−441 bp), we observed that three *cis-elements*, namely MYB, TGACG, and MYC (referred to as A, B, and C, respectively), were found in close proximity between the P2 and P3 regions. It is possible that these *cis-elements* play a crucial role in determining gene expression.

To further explore the role of these three elements, we studied the three elements together (ABC) or each element individually in three groups (AAA, BBB, and CCC). These elements were then combined with the −56 bp CaMV 35S minimum promoter (35Smini) to create four fused fragments (ABC-35Smini, AAA-35Smini, BBB-35Smini, and CCC-35Smini). Each fused fragment was then coupled with GUS reporter genes (Fig. 4A). The recombinant expression vector was transiently transformed into tobacco leaves, and GUS staining revealed that ABC-35Smini effectively drives GUS transcription, whereas AAA-35Smini, BBB-35Smini, and CCC-35Smini do not (Figs. 4B and 4C). These results indicate that A, B, or C elements alone cannot determine gene expression.

Furthermore, we introduced the ABC-35Smini recombinant vector into *Arabidopsis* and *Populus alba* × *Populus glandulosa* to investigate whether A, B, and C elements are sufficient to drive gene expression. GUS staining was observed in the leaves, roots, and stems of transgenic *Populus alba* × *Populus glandulosa* seedlings, as well as in transgenic 7-day-old *Arabidopsis* seedlings (Figs. 4D–4F). This result demonstrates that the ABC *cis-element* is a positive active element for gene expression.

To further explore the core structure of the ABC *cis-elements*, we subjected the three candidate elements A, B, and C to site-specific mutation and connected them to the 35Smini promoters, resulting in (A)BC-35Smini, A(B)C-35Smini, and AB(C)-35Smini, respectively (Fig. 5A). GUS staining results indicated that a mutation in element A or B
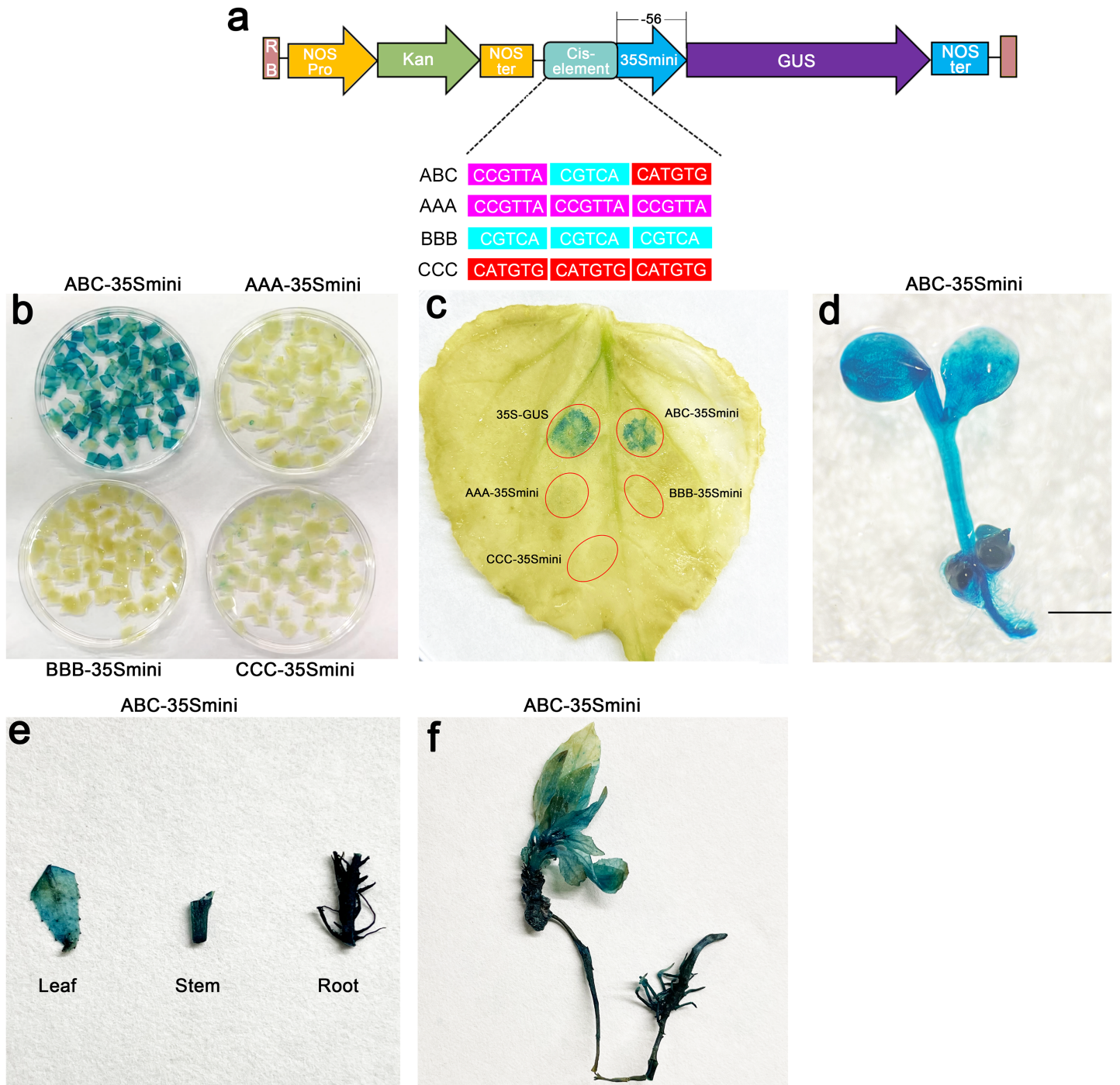
Figure 4 GUS activities analysis under the control in fusion of the *PtoCP1* promoter candidate cis-element with the 35Smini. (A) Schematic representation that the combination of *cis*-element and 35Smini fragment is fused with GUS. Purple background: MYB, blue background: TGACG-motif, red background: MYC, white capital base: element motif. (B) GUS staining in different leaves of the same tobacco plant; (C) GUS staining on the one tobacco leaf; (D) GUS staining in transgenic *Arabidopsis* thaliana seedlings. Bar = 1 mm; (E and F) GUS staining in transgenic *Populus alba* × *Populus glandulosa* seedlings.

Full-size 🖼 DOI: 10.7717/peerj.18292/fig-4
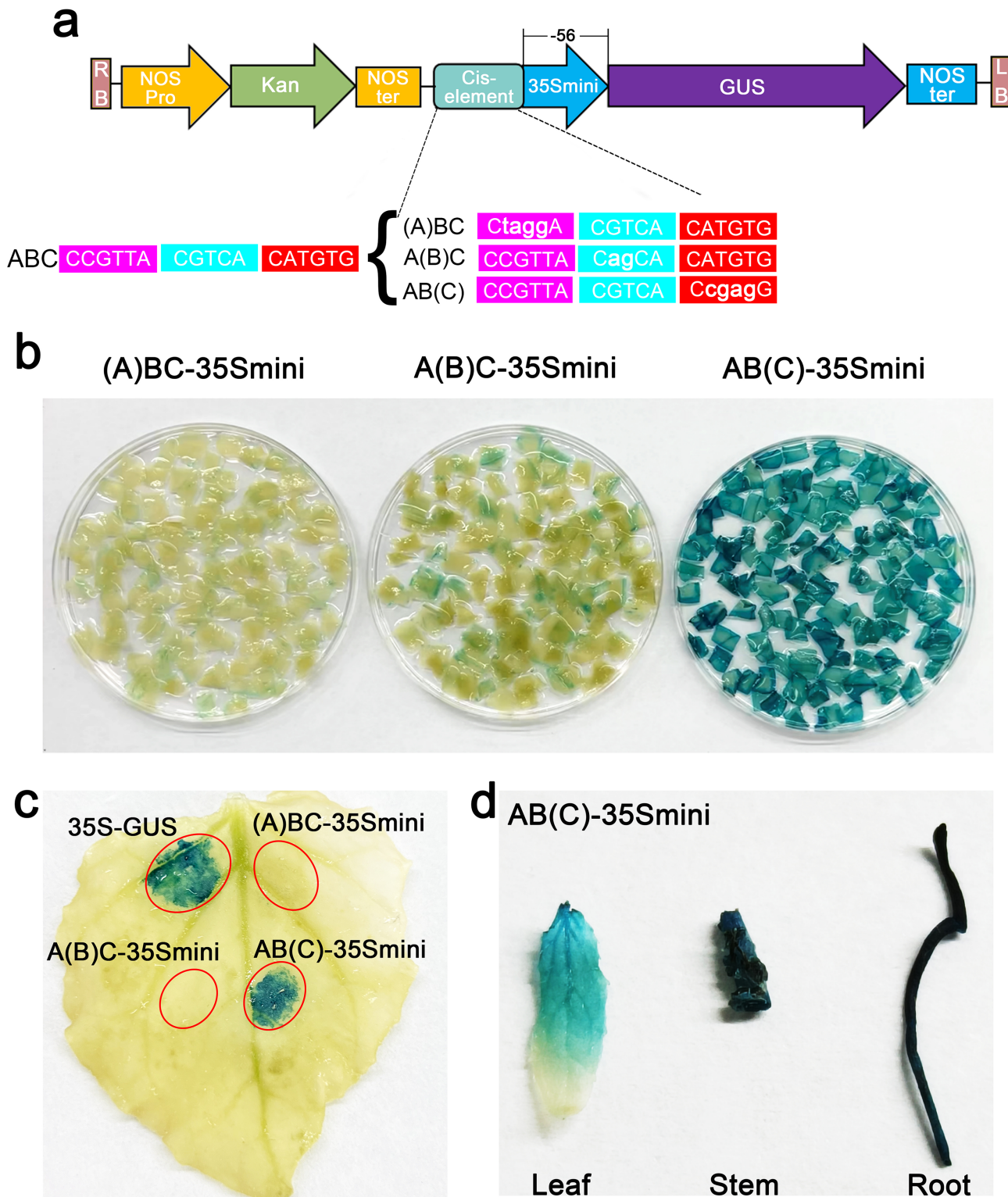
Figure 5 GUS activities analysis under the control in fusion of the site-mutation *cis*-element with the 35Smini. (A) Schematic representation that the combination of site- mutation *cis*-element and 35Smini fragment is fused with GUS. Purple background: MYB, blue background: TGACG-motif, red background: MYC, white capital base: element motif, ( ): mutant elements, white bold lowercase base: mutation site; (B) GUS staining in different leaves of the same tobacco plant; (C) GUS staining on the one tobacco leaf; (D) GUS staining in transgenic *Populus alba* × *Populus glandulosa* seedings. Full-size ⬚ DOI: 10.7717/peerj.18292/fig-5

completely abolished GUS activity, whereas a mutation in element C did not affect GUS activity (Figs. 5B and 5C). To confirm these findings, we stably transformed the AB(C)-35Smini recombinant vector into *Populus alba × Populus glandulosa*. GUS staining was observed in transgenic *Populus alba × Populus glandulosa* seedlings (Fig. 5D). These results reveal that the MYB-TGACG *cis-elements*, but not MYB-TGACG-MYC, are the key elements determining gene expression.

DISCUSSION

Understanding the properties of promoters is crucial for studying the function of genes in specific tissues and organs (*Tian, Xie & Qu, 2013*). However, the mechanism of gene activity in plants remains largely unexplored (*Xun et al., 2021*). Previous studies have identified positive regulatory regions in the *SlDREBA4* promoter (−1,095 to −730 bp and −162 to −38 bp) responsible for the promoter heat response (*Yaschenko et al., 2022*). Similarly, three positive regulatory regions in response to infestation by *Nilaparvata lugens Stal* (BPH) were identified in the rice *OsAOS* promoter through truncation experiments (*Deng et al., 2020*). The *GmPRP2* promoter, cloned from soybean, exhibits root preference in soybean and transgenic *A. thaliana* and its core fragment for root preference expression is likely between −369 bp and +1 bp (*Li et al., 2022*).

The G-box (CACGTG) *cis-element* on the *SmCP* promoter of the *Solanum melongena* cysteine protease gene interacts specifically with a nuclear protein in leaves. Its strong binding activity to the G-box is consistent with the high expression of *SmCP* (*Roy et al., 2012*). The S2 site (ATACA), which was identified in the nuclear *RPL21* gene promoter encoding a plastid ribosomal protein, functions as a leaf-specific *cis-element*. The activation of the S2 site in leaves is achieved through its interaction with the leaf transcription factor S2F (*Xu et al., 2003*). In this study, we conducted the first characterization of the *PtoCP1* promoter of *P. tomentosa* and discovered a gene expression element, the MYB-TGACG-motif *cis-element*, within the region from −466 to −441 bp.

Transcription factor binding is regulated by various interactions, primarily involving *cis-element* binding (*Lagrange et al., 1997*). The gene expression regulation network mediated by transcription factors plays a vital role in plant growth, development, and stress response pathways. MYB transcription factors constitute one of the most widely distributed families of transcription factors in plants and are involved in development and stress response by binding to the MYB *cis-element* in target gene promoters (*Mehrotra et al., 2013*). Previous studies have demonstrated that the TGACG-motif serves as a recognition binding element for the TGA family (*Wang, Niu & Zheng, 2021*; *Idrovo Espín et al., 2012*). TGA transcription factors, belonging to the basic region leucine zipper (bZIP) family, serve as crucial regulatory factors in various cellular processes. TGA proteins bind to their target DNA sequences as dimers through a conserved bZIP domain, and their activity is influenced by different hormonal pathways, interacting proteins, and regulatory elements (*Qi et al., 2022*). Fourteen genes encoding TGA transcription factors have been identified in *Hevea brasiliensis*, with HbTGA1 showing the highest expression in latex (*Tomaž, Gruden & Coll, 2022*). *Wang et al. (2019)* have reported strong expression of TGA1 and TGA4 in mature organs of *A. thaliana*, including the leaf base, petioles, pedicel

axils, and the base of floral organs (*Guo et al., 2022*). Our findings indicate a close connection between MYB and TGACG elements, jointly determining gene expression. Neither the individual MYB nor TGACG elements alone can drive gene expression, suggesting potential interaction between MYB and TGA transcription factors. The MYB family can be categorized into 1R-, R2R3-, 3R-, and 4R-MYB proteins based on the number of MYB domains (*Wang et al., 2019*), whereas TGA transcription factors can be divided into multiple subfamilies. Therefore, further verification is required to determine the specific types of TGA and MYB proteins that bind to the MYB (CCGTTA) and TGACG *cis-elements*.

CONCLUSION

In this study, we identified the *PtoCP1* promoter for the first time and identified two core cis-elements (MYB-TGACG) that are associated with gene expression between −466 and −441 bp. Further analysis showed that MYB-TGACG was a positive regulatory element of gene expression. It is of great significance to further understand the gene expression of cysteine protease, and provides a valuable gene expression element for genetic engineering.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Yu Peng performed the experiments, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Xueqin Guo performed the experiments, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Yawei Fan conceived and designed the experiments, performed the experiments, prepared figures and/or tables, and approved the final draft.
- Han Liu performed the experiments, prepared figures and/or tables, and approved the final draft.
- Leiqian Sun performed the experiments, prepared figures and/or tables, and approved the final draft.

- Di Liu performed the experiments, prepared figures and/or tables, and approved the final draft.
- Hui Li performed the experiments, prepared figures and/or tables, and approved the final draft.
- Xin Wang conceived and designed the experiments, performed the experiments, prepared figures and/or tables, and approved the final draft.
- Hongli Guo performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Hai Lu conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data is available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.18292#supplemental-information.

REFERENCES

Berger N, Dubreucq B, Roudier F, Dubos C, Lepiniec L. 2011. Transcriptional regulation of *Arabidopsis* LEAFY COTYLEDON2 involves RLE, a cis-element that regulates trimethylation of histone H3 at lysine-27. *The Plant Cell* **23**:4065–4078 DOI 10.1105/tpc.111.087866.

Chen HJ, Su CT, Lin CH, Huang GJ, Lin YH. 2010. Expression of *sweet potato* cysteine protease SPCP2 altered developmental characteristics and stress responses in transgenic *Arabidopsis* plants. *Journal of Plant Physiology* **167**:838–847 DOI 10.1016/j.jplph.2010.01.005.

Cheng Z, Zhang J, Yin B, Liu Y, Wang B, Li H, Lu H. 2019. γVPE plays an important role in programmed cell death for xylem fiber cells by activating protease CEP1 maturation in *Arabidopsis thaliana*. *International Journal of Biological Macromolecules* **137**:703–711 DOI 10.1016/j.ijbiomac.2019.07.017.

Deng MH, Lv JH, Wang ZR, Zhu HS, Yang ZA, Yue YL, Zhao K. 2020. Two promoter regions confer heat-induced activation of *SlDREBA4* in *Solanum lycopersicum*. *Biochemistry Biophysics Research Communication* **524**:689–695 DOI 10.1016/j.bbrc.2020.01.153.

Du Q, Wang B, Wei Z, Zhang D, Li B. 2012. Genetic diversity and population structure of *Chinese White poplar (Populus tomentosa)* revealed by SSR markers. *Hered* **103**:853–862 DOI 10.1093/jhered/ess061.

Frank S, Hollmann J, Mulisch M, Matros A, Carrión CC, Mock HP, Hensel G, Krupinska K. 2019. Barley cysteine protease PAP14 plays a role in degradation of chloroplast proteins. *Journal of Experimental Botany* **70**:6057–6069 DOI 10.1093/jxb/erz356.

Gomez-Sanchez A, Gonzalez-Melendi P, Santamaria ME, Arbona V, Lopez-Gonzalvez A, Garcia A, Hensel G, Kumlehn J, Martinez M, Diaz I. 2019. Repression of drought-induced cysteine-protease genes alters barley leaf structure and responses to abiotic and biotic stresses. *Journal of Experimental Botany* **70**:2143–2155 DOI 10.1093/jxb/ery410.

Guo D, Li HL, Zhu JH, Wang Y, Peng SQ. 2022. *HbTGA1*, a TGA transcription factor from *Hevea brasiliensis*, regulates the expression of multiple natural rubber biosynthesis genes. *Frontiers in Plant Science* **13**:909098 DOI 10.3389/fpls.2022.909098.

Idrovo Espín FM, Peraza-Echeverria S, Fuentes G, Santamaría JM. 2012. In silico cloning and characterization of the TGA (TGACG MOTIF-BINDING FACTOR) transcription factors subfamily in *Carica papaya*. *Plant Physiology Biochemistry* **54**:113–122 DOI 10.1016/j.plaphy.2012.02.011.

Jai P, Sanjeeta K, Shruti K. 2015. Cathepsin B cysteine protease gene is upregulated during leaf senescence and exhibits differential expression behavior in response to phytohormones in *Picrorhiza kurrooa Royle ex Benth*. *Plant Gene* **3**:11–19 DOI 10.1016/j.plgene.2015.07.001.

James M, Poret M, Masclaux-Daubresse C, Marmagne A, Coquet L, Jouenne T, Chan P, Trouverie J, Etienne P. 2018. SAG12, a major cysteine protease involved in nitrogen allocation during senescence for seed production in *Arabidopsis thaliana*. *Plant Cell Physiology* **59**:2052–2063 DOI 10.1093/pcp/pcy125.

Kovács J, Poór P, Szepesi Á, Tari I. 2016. Salicylic acid induced cysteine protease activity during programmed cell death in *tomato* plants. *Acta Biologica Hungarica* **67**:148–158 DOI 10.1556/018.67.2016.2.3.

Lagrange T, Gauvin S, Yeo HJ, Mache R. 1997. S2F, a leaf-specific trans-acting factor, binds to a novel cis-acting element and differentially activates the *RPL21* gene. *The Plant Cell* **9**:1469–1479 DOI 10.1105/tpc.9.8.1469.

Li H, Xu X, Han K, Wang Z, Ma W, Lin Y, Hua H. 2022. Isolation and functional analysis of *OsAOS1* promoter for resistance to *Nilaparvata lugens Stål* infestation in *rice*. *Journal of Cell Physiology* **237**:1833–1844 DOI 10.1002/jcp.30653.

Liu X, Mo L, Guo X, Zhang Q, Li H, Liu D, Lu H. 2021. How cysteine protease gene *PtCP5* affects seed germination by mobilizing storage proteins in *Populus trichocarpa*. *International Journal of Molecular Sciences* **22**:12637 DOI 10.3390/ijms222312637.

Liu J, Sharma A, Niewiara MJ, Singh R, Ming R, Yu Q. 2018. Papain-like cysteine proteases in *Carica papaya*: lineage-specific gene duplication and expansion. *BMC Genomics* **9**:26 DOI 10.1186/s12864-017-4394-y.

Mehrotra R, Sethi S, Zutshi I, Bhalothia P, Mehrotra S. 2013. Patterns and evolution of ACGT repeat *cis*-element landscape across four plant genomes. *BMC Genomics* **14**:203 DOI 10.1186/1471-2164-14-203.

Qi P, Huang M, Hu X, Zhang Y, Wang Y, Li P, Chen S, Zhang D, Cao S, Zhu W, Xie J, Cheng J, Fu Y, Jiang D, Yu X, Li B. 2022. A Ralstonia solanacearum effector targets TGA transcription factors to subvert salicylic acid signaling. *The Plant Cell* **34**:1666–1683 DOI 10.1093/plcell/koac015.

Roy S, Choudhury SR, Singh SK, Das KP. 2012. Functional analysis of light-regulated promoter region of *AtPolλ* gene. *Planta* **235**:411–432 DOI 10.1007/s00425-011-1517-6.

Tian Xi, Xie J, Qu L. 2013. Growth and wood quality characteristics of 741 *poplar*. *Scientia Silvae Sinicae* **49**:129–135 DOI 10.11707/j.1001-7488.20130318.

Tomaž Š, Gruden K, Coll A. 2022. TGA transcription factors-structural characteristics as basis for functional variability. *Frontiers in Plant Science* **13**:935819 DOI 10.3389/fpls.2022.935819.

Wang X, Niu Y, Zheng Y. 2021. Multiple functions of MYB transcription factors in abiotic stress responses. *International Journal of Molecular Sciences* **22**:6125 DOI 10.3390/ijms22116125.

Wang Y, Salasini BC, Khan M, Devi B, Bush M, Subramaniam R, Hepworth SR. 2019. Clade I TGACG-motif binding basic leucine zipper transcription factors mediate BLADE-ON-PETIOLE-dependent regulation of development. *Plant Physiology* **180**:937–951 DOI 10.1104/pp.18.00805.

Xu ZF, Chye ML, Li HY, Xu FX, Yao KM. 2003. G-box binding coincides with increased *Solanum melongena* cysteine proteinase expression in senescent fruits and circadian-regulated leaves. *Plant Molecular Biology* **51**:9–19 DOI 10.1023/a:1020859518877.

Xun H, Zhang X, Yu J, Pang J, Wang S, Liu B, Dong Y, Jiang L, Guo D. 2021. Analysis of expression characteristics of *soybean* leaf and root tissue-specific promoters in *Arabidopsis* and soybean. *Transgenic Research* **30**:799–810 DOI 10.1007/s11248-021-00266-7.

Yan L, Yang Q, Han J. 2005. Research progress of cysteine proteases in plants. *Acta Prataculturae Sinica* **14**:9 DOI 10.3321/j.issn:1004-5759.2005.05.003.

Yaschenko AE, Fenech M, Mazzoni-Putman S, Alonso JM, Stepanova AN. 2022. Deciphering the molecular basis of tissue-specific gene expression in plants: can synthetic biology help? *Current Opinion in Plant Biology* **68**:102241 DOI 10.1016/j.pbi.2022.102241.

Zhang D, Liu D, Lv X, Wang Y, Xun Z, Liu Z, Li F, Lu H. 2014. The cysteine protease CEP1, a key executor involved in tapetal programmed cell death, regulates pollen development in *Arabidopsis*. *The Plant Cell* **26**:2939–2961 DOI 10.1105/tpc.114.127282.